

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
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 natural range of seasonal signals, variability
15 and temperature trends for 84 of these time series, the length, decadal trend
16 and data precision of each time series were systematically varied before fitting
17 generalized least squares (GLS) models to study the effect these variables has
18 on trend detection. We determined that the variables contributing most to ac-
19 curate trend detection in decreasing order are: time series length, the decadal
20 trend, variance, amount of missing data and the precision of the measure-
21 ments. Time series at least 25 years in length may be used tentatively for
22 climate change research, but time series >30 years in length are preferable.
23 The implication is that long-running thermometer time series in this dataset,
24 and others around the world, are more useful for decadal scale climate change
25 studies than the shorter, more precise UTR time series. It is important to note
26 that due to the nature of the dataset used in this study, instrument drift was
27 not able to be quantified. *Say something too about the decadal trend, vari-*
28 *ance, amount of missing data and the precision. RWS - Like what? That*
29 *greater decadal trends require fewer months of data to model accurately and*
30 *significantly? That variance has a large effect on the significance of the trend*
31 *detected, but not on the accuracy of the model? And that missing data er-*
32 *roneously increases the p-value of shallow trends, but decreases the p-value*
33 *of steep trends, adding to the risk of committing type I and type II errors?*
34 *That precision of 0.5C should be avoided unless the time series is quite long,*
35 *and that precision of 0.1C is just about as good as 0.001C? Easy to add, but*

38 1. Introduction

39 The roughly 3,000 km of South Africa's coastline is bordered by the Benguela and Agulhas
40 Currents (*e.g.* Roberts 2005; Hutchings et al. 2009), which, in combination with other nearshore
41 processes, affect the country's marine coastal ecosystems (Santos et al. 2012). A thorough under-
42 standing of these coastal processes is provided by several physical variables, of which temperature
43 is one of the main determinants (*e.g.* Blanchette et al. 2008; Tittensor et al. 2010; Couce et al.
44 2012). The statistical properties of *in situ* seawater temperature time series representing the whole
45 coastline – such as the annual mean, minimum and maximum temperature, and the thermal range
46 and variance characteristics – vary greatly among coastal sections due to the varying influence of
47 the Benguela and Agulhas Currents. Based on these thermal properties, the coastline has been
48 classified into a cool temperate west coast, a warm temperate south coast and tending towards
49 sub-tropical along the east coast (Smit et al. 2013; Mead et al. 2013). That the ocean temperature
50 of these regions is changing has been reported in recent years. For example, an increase of 0.55°C
51 to $0.7^{\circ}\text{C dec}^{-1}$ has been reported in the Agulhas Current (Rouault et al. 2009, 2010), while the
52 southern Benguela has decreased in $0.5^{\circ}\text{C dec}^{-1}$ during some parts of the year (Rouault et al.
53 2010).

54 The aforementioned climate change trends were derived from gridded sea surface temperature
55 (SST) records. Some sources warn of the pitfalls in using temperatures extracted from such grid-
56 ded products close to the shore (*insert references here*), and a study by Smit et al. (2013) showed
57 that SST data indeed have a warm bias as large as 6°C when compared to coastal *in situ* data.
58 Nevertheless, a widespread approach in coastal ecological research is to use satellite and/or model-
59 generated temperature data as a representation of the sea surface temperature (SST) along coast-
60 lines (*e.g.* Blanchette et al. 2008; Broitman et al. 2008; Tyberghein et al. 2012), because apparently

61 the dangers of applying gridded SSTs to the coast are not widely known or in many places in the
62 world there simply are no suitable *in situ* coastal temperature time series available. It is for this
63 reason that we strongly recommended the use of *in situ* data to support research conducted within
64 400 m from the shoreline.

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

80 This prompted us to examine the 129 *in situ* time series that comprise the SACTN. The range
81 of measurement precisions and statistical characteristics of this dataset was used to guide a series
82 of enquiry-driven analyses into the suitability of the time series to yield statistically significant
83 and accurate assessments of decadal temperature change. The length, decadal trend and data
84 precision of each time series were adjusted in a systematic manner, and forms the core of our

85 analyses. Furthermore, the natural variability of each of the time series, which differ more-or-less
86 predictably between coastlines variously affected by the Benguela and Agulhas Currents, was also
87 entered into the analysis. Our aim was to assess the effect that each of these variables has on the
88 ability of a model to produce a robust estimate of time series decadal trend. The effect gaps in the
89 time series may have on the fitting of models was also investigated as many of the time series used
90 here have some missing data scattered throughout, which is unavoidable for a 20+ year time series
91 that is sampled by hand by a single technician at each site.

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

94 **2. Methods**

95 *a. Data Sources*

96 Our study lies within the political borders of South Africa's coastline. The location of each point
97 of collection appears in Figure 1. Of these 129 time series, 43 are recorded with UTRs and the
98 other 86 with hand-held mercury thermometers. The oldest currently running time series began on
99 January 1st, 1972; there are 11 total time series that started in the 70s, 53 more started in the 80s,
100 34 began in the 90s, 18 in the 00s and 13 in the current decade.

101 The data are collected using two different methods and a variety of instruments. Hand-held
102 mercury thermometers (which are being phased out in favor of alcohol thermometers or electronic
103 instruments) are used in some instances at the shoreline, and represent seawater temperatures at
104 the surface. At other places, predominantly along the country's east coast, data are collected with
105 glass thermometers from small boats at the location of shark nets along the coast (Cliff et al. 1988).
106 Whereas both types of thermometers allow for a measurement precision of 0.1 °C, the recordings

are written down at a precision of 0.5 °C. Data at other localities are collected using delayed-mode instruments that are permanently moored shallower than 10 m, but generally very close to the surface below the low-water spring tide level.

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

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

b. Data Management

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

130 data points were changed to NA, meaning ‘not available’, before including them in the SACTN
131 dataset.

132 All analyses and data management performed in this paper were conducted with R version 3.3.1
133 (2016-06-21) (R Core Team 2013). The script and data used to conduct the analyses and create
134 the tables and figures in this paper may be found at https://github.com/schrob040/Trend_Analysis.

135 Any time series with a temporal precision greater than one day were averaged into daily values
136 before being aggregated into the SACTN. A series of additional checks were then performed (*e.g.*
137 removing long stretches in the time series without associated temperature recordings) and time
138 series shorter than five calendar years, collected at depths greater than 10 m or missing more than
139 15% of their monthly values were removed. At the time of this analysis, this useable daily dataset
140 consisted of 84 time series, consisting of 819,499 days of data; these data were then binned further
141 to the 26,924 monthly temperature values available for use in this study.

142 *c. Systematic Analysis of Time Series*

143 We used the 84 time series simply for their variance properties (comprised of seasonal, inter-
144 annual, decadal and noise components), which reflect that of the thermal environment naturally
145 present along the roughly 3,000 km of South African coastline. Linear trends that may have been
146 present in each time series were removed prior to the ensuing analysis by applying an ordinary
147 least squares regression and keeping the detrended residuals as anomaly time series. In doing so
148 we avoided the need to simulate a series of synthetic time series, whose variance components may
149 not have been fully representative of that naturally present in coastal waters. These detrended
150 anomaly time series (henceforth simply called ‘time series’) represent a range of time scales from
151 72 to 519 months in duration.

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

160 To gain further insight into the effect of time series length on trend detection, each time series
161 was first shortened to a minimum length of 5 years, starting in January so that the timing of the
162 seasonal signal for each time series would be equitable. After fitting the model (see *Time Series*
163 *Model*, below) to all 1,680 of the shortened time series, the next year of data for each time series
164 was added and the models fitted again. This process was iterated until the full length of each time
165 series was attained. For example, if a time series consisted of 12 full years of data, it would require
166 160 models (8 iterations of increasing length \times 5 decadal trends \times 4 levels of precision); similarly,
167 720 models would be applied to a 40 year time series. Considering the 84 time series available,
168 the total number of individual models required to capture each combination of variables quickly
169 increased to 36,220.

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

176 *d. Time Series Model*

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

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

193 where the lag-2 autocorrelated residuals are given by

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

194 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 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

216 added decadal trend increases. The p -value of the fitted models also vary in relation to time series
 217 duration and to the steepness of the added decadal trend (Figure 5). It is usually the longer time
 218 series equipped with steeper decadal trends that are able to produce model fits with estimated
 219 trends that are statistically significant. Note, however, that this p -value tests the null hypothesis
 220 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
 221 not different from the added trend. Taken together, these outcomes show that although our GLS
 222 model can very often result in trend estimates that *approach* the true trend, it is seldom that those
 223 estimates are statistically significant in the sense that the estimated trends differ statistically from
 224 $0\text{ }^{\circ}\text{C dec}^{-1}$.

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

232 The number of NAs permitted in any of our time series was limited to 15% per time series.
 233 Twenty-five of the 84 time series have fewer than 1% NA. An additional 45 time series have up
 234 to 5% NA, 10 have up to 10% NA and 4 have up to 15% NA. The mean number of NA for the data
 235 is 2.65%. The relationship between %NA and the p -value of the models is shown in Figure 7.
 236 At 2.5% or fewer NA their presence does not have any discernible effect on resultant p -values.
 237 Progressively increasing the number of NAs above 5%, however, leads to a drastic improvement
 238 of models fitted to series with no or gently increasing decadal trends (these generally have very
 239 large p -values indicative of very poor fits, perhaps due to the presence of a very weak signal), and

240 a significant deterioration of models fitted to data with steep decadal trends (for these data, the
241 model generally fits better at low numbers of NAs, as suggested by the greater number of p -values
242 that approach 0.05). In other words, the inclusion of missing values results in time series with no
243 added decadal trend to veer away from $0\text{ }^{\circ}\text{C dec}^{-1}$ towards a situation where they may erroneously
244 appear to display a trend. On the other hand, time series that do indeed have decadal trends tend
245 to produce fits that are not significantly different from $0\text{ }^{\circ}\text{C dec}^{-1}$.

246 Regarding the effect that the level of measurement precision has on the GLS models, we see in
247 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
248 differences in the modeled trends. The Root Mean Square Error (RMSE) between the slopes esti-
249 mated from $0.001\text{ }^{\circ}\text{C}$ and $0.01\text{ }^{\circ}\text{C}$ data is 0.001. The correspondence between the slopes estimated
250 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.

251 The effect of decreasing data measurement precision from $0.001\text{ }^{\circ}\text{C}$ to $0.5\text{ }^{\circ}\text{C}$ has almost no
252 appreciable effect on any of the measures of variance presented in this study. The effect of mea-
253 surement precision on the accuracy of the modeled slope, however, becomes very pronounced
254 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
255 trend of $0.05\text{ }^{\circ}\text{C dec}^{-1}$, the deviation from the true value of models fitted to data with a precision
256 of $0.1\text{ }^{\circ}\text{C}$ is negligible; however, the accuracy of the fitted model on data recorded at a precision
257 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
258 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 im-
259 proves to an average difference of 6.44% with a slope of $0.10\text{ }^{\circ}\text{C dec}^{-1}$, 2.24% at $0.15\text{ }^{\circ}\text{C dec}^{-1}$
260 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
261 less accurate modeled trends than at higher precisions; however, the current analysis did not high-
262 light one precision that consistently provides the most accurate estimate of the trends. This may

263 however become determinable in an analysis of synthetic data with variance structures that are
264 manipulated in a more consistent manner.

265 As the actual time series used to generate the data for this study are predominantly greater than
266 300 months in length and recorded at a data precision of 0.5°C , we would be remiss not to investi-
267 gate the interaction between the increase in accuracy provided by a lengthy time series, against the
268 decrease caused by a data precision of 0.5°C . In other words, at what point does a model fitted to a
269 longer time series, with less precise measurements (*e.g.* those taken by thermometers and reported
270 at a precision of 0.5°C), become as accurate as a time series with more precise measurements (*e.g.*
271 UTRs)? Figure 8 shows how varied the modeled trends become when a precision of 0.5°C is used,
272 and we see here that when these low resolution time series have a shallow slope of $0.05^{\circ}\text{C dec}^{-1}$,
273 a fitted model requires 24 months of additional data on average to have a comparable level of accu-
274 racy to a model fitted to data recorded at a precision of 0.1°C . This difference in length decreases
275 to 16 months when a larger slope $0.20^{\circ}\text{C dec}^{-1}$ is used.

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

281 4. Discussion

282 The strongest finding of this analysis is that the accurate detection of long-term trends in time
283 series primarily concerns the length of a dataset. But there is also a host of nuances resulting from
284 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 $0^{\circ}\text{C dec}^{-1}$ 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 data sets with steeper slopes in the temperature vs. time relationship. The selection of sites for long-term monitoring must therefore account for the location of study and necessitate adequate planning to collect a long enough time series.

The detection of long-term trends require long-term data, a fact that is already firmly established in climate change research (Ohring et al. 2005; IPCC 2013). The length of these time series is firmly under the control of the investigator with sufficient foresight and perseverance to plan the installation and management of new instrument networks that will yield usable results only after about three-quarters of a typical academic career has passed. Should such data already exist – and considering the scarcity of such long-term records that are already yielding benefits today – we

309 must ensure that these sources of data are managed and curated with great care and diligence as
310 they are practically irreplaceable. For this reason, it is essential that we understand the inherent
311 strengths and weaknesses of such existing sources of data so that we may fully maximize their
312 utility and extract from them the model coefficients needed to detect decadal temperature trends,
313 and know the accuracy of these estimates to the best of our ability. There are many time series
314 < 20 years in length that should be avoided, where possible, for trend analysis. These will mature
315 with time and their maintenance need to be ensured going forward.

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

326 We have demonstrated clearly that as the steepness of an expected decadal trend increases, the
327 ability for it to be modeled accurately increases, too. Our GLS model is generally not able to detect
328 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
329 rarely were we able to produce significant model fits at shallower slopes. Finding significant trends
330 at $< 0.05\text{ }^{\circ}\text{C dec}^{-1}$ was not possible. Based on the relationship between SD and the added decadal
331 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
332 trend of $0.10\text{ }^{\circ}\text{C dec}^{-1}$ should consist of roughly 640 months of data before our GLS model would

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

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

357 We have reflected on the importance of the accuracy of the models, and not only on the impor-
358 tance of their of significance. Indeed, the p -value given for the slope in a model does not show
359 how well the model detects the true trend in the data (known *a-priori* in this study); rather, it tells
360 us if the detected trend is significantly different from $0\text{ }^{\circ}\text{C dec}^{-1}$. *How many of the models in*
361 *the natural data produce accurate trend estimates but are not significant?* This is not particularly
362 useful for applying the results of climate change research more broadly to biotic interests. That
363 a long term trend does exist, may be accurately detected by a model and related to an observed
364 change in the natural world – such as range expansion/contraction of coastal biota (Bolton et al.
365 2012; Straub et al. 2016; Wernberg et al. 2016) – is more important than whether or not the model
366 can show if that trend is significantly different from $0\text{ }^{\circ}\text{C dec}^{-1}$ in a statistical sense.

367 We must mention also that much of the meta-data pertaining to the older temperature records
368 used here have over time been lost. As with the bulk of the International Comprehensive Ocean-
369 Atmosphere Data Set (ICOADS; Freeman et al. 2016), *in situ* coastal seawater temperature mon-
370 itoring that started in the 1970s in South Africa was not developed with climate change research
371 in mind, and comprehensive records that keep track of details of the instruments used, calibration,
372 their turnover, change in monitoring methods and locations and so forth are not always available
373 as per modern requirements (Aguilar et al. 2003). For studies of climate change *per se* this is a
374 serious limitation and it prevents us from knowing anything about the accuracy of the instruments
375 or potential issues of drift (stability) that may have occurred. We do know however that all time
376 series sampled with thermometers were sampled only with thermometers, and *vice versa* for the
377 UTR time series, ensuring that the precisions of the measured data used in this study are correct.
378 Moving forward with the further development of the SACTN and the establishment of a national
379 standard of data collection and instrument maintenance, we are able to record and archive all these
380 levels of pertinent meta-data, and allowing for the enforcement of SI traceability and the accurate

381 measurement of instrument drift (Jarraud 2008). Nevertheless, the detrended anomaly time series
382 used here were taken only for their variance properties, which we think accurately reflect that of
383 the various coastal sections around the coast. They provide a strong backbone for semi-artificial
384 time series, and we have shown how important insights about model fitting could be derived from
385 these data.

386 5. Conclusion

387 We draw several key conclusions:

- 388 1. There is a rapid increase in the accuracy and significance of modelled trends as time series
389 lengths extend from 10 to 20 years. This improvement slows from 20 to 30 years, and as time
390 series approach 40 years in length the accuracy of models becomes nearly exact. Modeled
391 trends from time series at or under 25 years in length should be interpreted with extreme
392 caution.
- 393 2. For our variable coastal seawater, a time series of 520 month in length is required to detect a
394 decadal trend in line with the global average (*i.e.* near $0.1\text{ }^{\circ}\text{C dec}^{-1}$) with perfect accuracy;
395 however, an additional 120 months of data is often required for the detected trend to be
396 considered significant ($p \leq 0.05$).
- 397 3. The length of a time series required to detect a decadal trend at $0.1\text{ }^{\circ}\text{C dec}^{-1}$ may rapidly
398 exceed 100 years when a large amount of variance is present.
- 399 4. The larger the decadal trend within a time series, the more accurately it will be modeled
400 regardless of the amount of variance in the time series.

5. There is a complicated relationship between the accuracy of a trend fitted to a time series and the %NA of that time series. As the %NA increases, so too does the change of committing type 1 (at gentle slopes) or type 2 errors (at steeper trends).
6. A precision greater than 0.5 °C is not required to confidently detect the long-term trend in a time series; however, precisions at or greater than 0.1 °C will reduce the length of time required to accurately detect a long term trend if one does exist.
7. Improving the precision of measurements to greater than 0.1 °C has almost no appreciable effect on a models ability to detect a long-term trend, provided that the reported effect size matches the level of precision by the instruments.

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

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

References

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

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

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

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

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

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

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

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

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

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

447 Hartmann, D., and Coauthors, 2013: Observations: Atmosphere and surface supplementary mate-
 448 rial. in: *Climate change 2013: The physical science basis*.

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

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

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

456 Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker, 2011: Reassessing biases
 457 and other uncertainties in sea-surface temperature observations measured *in situ* since 1850, Part
 458 1: measurement and sampling uncertainties. *Journal of Geophysical Research Atmospheres*,
 459 **116**.

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

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

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

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

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

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

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

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

480 Rouault, M., P. Penven, and B. Pohl, 2009: Warming in the Agulhas Current system since the
 481 1980’s. *Geophysical Research Letters*, **36 (12)**, 1–5.

482 Rouault, M., B. Pohl, and P. Penven, 2010: Coastal oceanic climate change and variability from
 483 1982 to 2009 around South Africa. *African Journal of Marine Science*, **32 (2)**, 237–246.

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

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

489 Scherrer, D., and C. Körner, 2010: Infra-red thermometry of alpine landscapes challenges climatic
 490 warming projections. *Global Change Biology*, **16 (9)**, 2602–2613, doi:10.1111/j.1365-2486.
 491 2009.02122.x.

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

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

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

502 Straub, S. C., M. S. Thomsen, and T. Wernberg, 2016: The dynamic biogeography of the an-
 503 thropocene: The speed of recent range shifts in seaweeds. *Seaweed Phylogeography*, Springer,
 504 63–93.

505 Tittensor, D. P., C. Mora, W. Jetz, H. K. Lotze, D. Ricard, E. V. Berghe, and B. Worm, 2010:
506 Global patterns and predictors of marine biodiversity across taxa. *Nature*, **466 (7310)**, 1098–
507 1101, doi:10.1038/nature09329.

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

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

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

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

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

517	LIST OF FIGURES	
518	Fig. 1. The location of the 129 time series available for use in this study	27
519	Fig. 2. Box and whisker plots summarizing the anomaly time series of the base 84 data sets	28
520	Fig. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend added to a	
521	time series	29
522	Fig. 4. The relationship between time series length and the standard error (SE) of the modelled trend	30
523	Fig. 5. The effect of the natural variation of a time series on the significance of the modelled trend	31
524	Fig. 6. The relationship between the effect of initial SD on the SE of a modelled trend	32
525	Fig. 7. The relationship between the percentage of missing values (%NA) and the significance of a fitted trend	33
526	Fig. 8. The minimal effect of rounding from	34

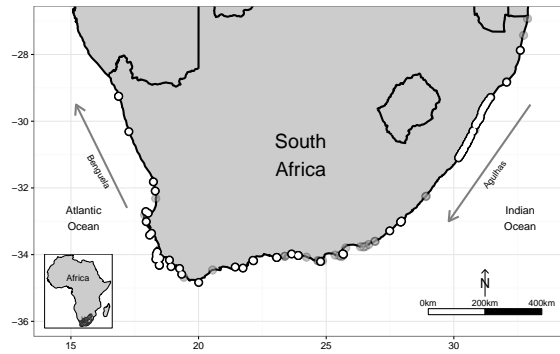
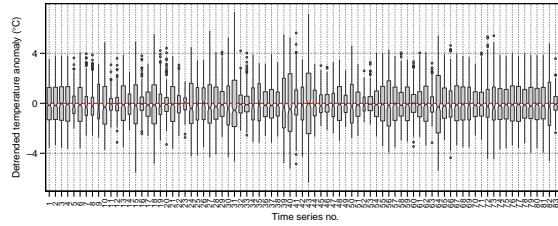


FIG. 1. The location of the 129 time series available for use in this study. The 84 time series actually used are shown as solid white circles and those not used are shown as opaque.



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

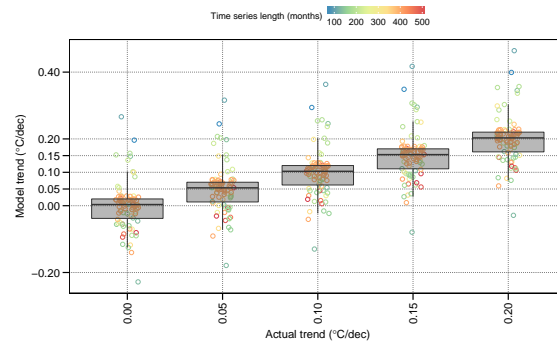


FIG. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend added to each time series. The box-and-whisker plots show the first and third quartile as the extremities of the boxes, the median is shown as the horizontal line within each box, and the minima and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

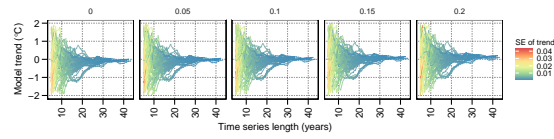


FIG. 4. The relationship between the length of a time series, the size of the modeled trend and its the standard error (SE). Each individual line shows the modeled trend for one of the 84 sites used in this analysis to which a model was fitted iteratively as the time series length was ‘grown’ from 5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

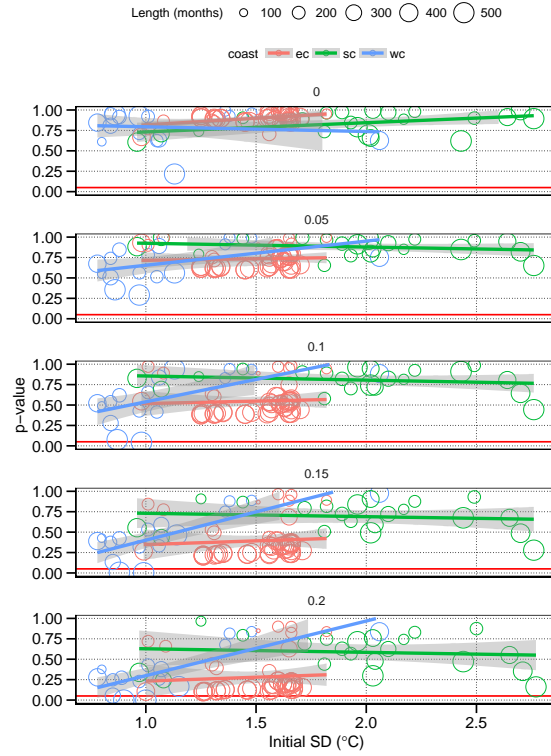


FIG. 5. The effect of the natural variation of a time series on the significance of the modelled trends estimated by the GLS. The size of the symbols are scaled proportionally to the time series length, with longer time series shown as larger circles.

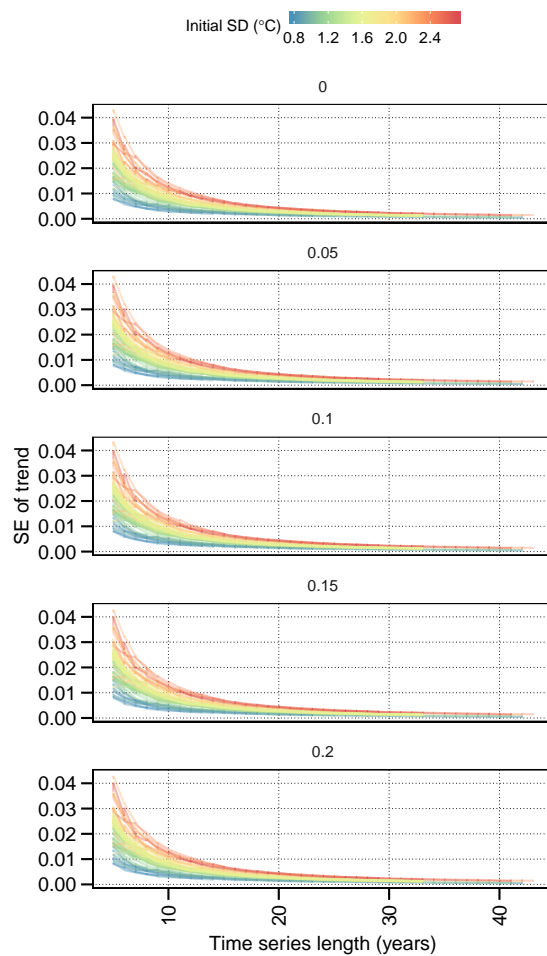


FIG. 6. The relationship between the effect of the initial SD of a time series on the SE of a modelled trend,
controlled for by the length of the time series.

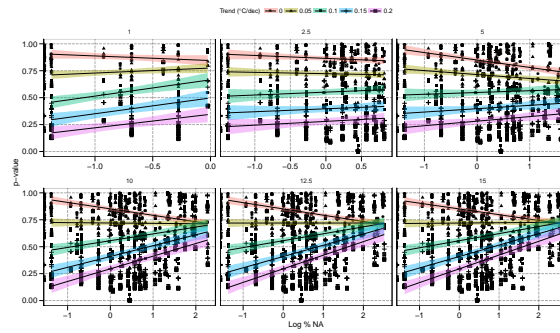


FIG. 7. The relationship between the percentage of missing values (%NA) and the significance of a fitted trend. Each panel shows the effect of an increasingly larger amount of missing values. The fitted lines and 95% confidence intervals represent each of the five decadal trends assessed.

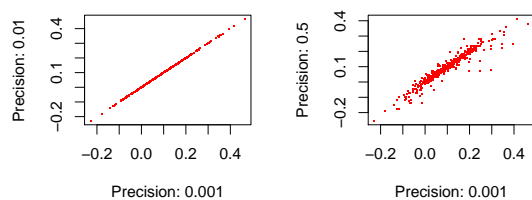


FIG. 8. The minimal effect of rounding from 0.001 °C to 0.01 °C may be seen in the panel on the right. The panel on the left shows that rounding from a precision of 0.001 °C to 0.5 °C has a more appreciable effect.