RECONSTRUCTING NORTHEASTERN PACIFIC CLIMATE VARIABILITY FROM THE ANNUAL GROWTH INCREMENTS OF PACIFIC GEODUCK

by

David C. Edge

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Copyright © David C. Edge 2022

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF GEOSCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2022



Table of Contents

[LIST OF FIGURES 6](#_Toc118026828)

[LIST OF TABLES 7](#_Toc118026829)

[ABSTRACT 8](#_Toc118026830)

[1. INTRODUCTION 9](#_Toc118026831)

[1.1. Motivation 9](#_Toc118026832)

[1.2. Background 10](#_Toc118026833)

[1.2.1 The Pacific Geoduck 10](#_Toc118026834)

[1.2.2 Northeast Pacific Climate 10](#_Toc118026835)

[1.2.3 Northeast Pacific Surface Currents and Radiocarbon 11](#_Toc118026836)

[2. PRESENT STUDY 13](#_Toc118026837)

[2.1 Pacific Geoduck Chronology Construction and SST Reconstruction 13](#_Toc118026838)

[2.2 14C Reconstruction and Water Mass Variability Assessment 13](#_Toc118026839)

[2.3 Evaluation of Methods Used to Quantify Reconstruction Uncertainties 14](#_Toc118026840)

[REFERENCES 15](#_Toc118026841)

[Appendix A 19](#_Toc118026842)

[A.1. Abstract 20](#_Toc118026843)

[A.2. Introduction 21](#_Toc118026844)

[A.3. Materials and Methods 22](#_Toc118026845)

[A.3.1 Sample Collection and Preparation 23](#_Toc118026846)

[A.3.2 Crossdating 25](#_Toc118026847)

[A.3.3 Radiocarbon Dating 26](#_Toc118026848)

[A.3.4 Chronology Construction 27](#_Toc118026849)

[A.3.5 Climate-Growth Relationships 29](#_Toc118026850)

[A.3.6 SST Reconstruction 30](#_Toc118026851)

[A.4 Results 31](#_Toc118026852)

[A.4.1 Sample Collection and Preparation 31](#_Toc118026853)

[A.4.2 Crossdating 33](#_Toc118026854)

[A.4.3 Radiocarbon dating 34](#_Toc118026855)

[A.4.4 Chronology construction 35](#_Toc118026856)

[A.4.5 Climate-Growth Relationships 36](#_Toc118026857)

[A.4.6 SST Reconstruction 36](#_Toc118026858)

[A.5 Discussion 39](#_Toc118026859)

[A.6 Acknowledgments, Samples, and Data 47](#_Toc118026860)

[A.7 References 48](#_Toc118026861)

[A.8. Supplemental Material 61](#_Toc118026862)

[Appendix B 70](#_Toc118026874)

[B.1. Abstract 71](#_Toc118026875)

[B.2. Introduction 72](#_Toc118026876)

[B.3. Methods and Background 74](#_Toc118026877)

[B.3.1. Oceanographic Setting 74](#_Toc118026878)

[B.3.2. Pre-bomb radiocarbon 76](#_Toc118026879)

[B.3.3. Radiocarbon and climate covariability 77](#_Toc118026880)

[B.3.4. Bomb-pulse radiocarbon data 78](#_Toc118026881)

[B.4. Results 79](#_Toc118026882)

[B.4.1. Pre-bomb radiocarbon 79](#_Toc118026883)

[B.4.2. Radiocarbon and climate covariability 80](#_Toc118026884)

[B.4.3. Bomb-pulse radiocarbon data 81](#_Toc118026885)

[B.5. Discussion 82](#_Toc118026886)

[B.6. Acknowledgements 89](#_Toc118026887)

[B.7. References 90](#_Toc118026888)

[B.8. Supplemental Material 96](#_Toc118026889)

[Appendix C 99](#_Toc118026890)

[C.1. Abstract 100](#_Toc118026891)

[C.2. Introduction 101](#_Toc118026892)

[C.3. Methods 102](#_Toc118026893)

[C.3.1. Chronologies 102](#_Toc118026894)

[C.3.2 Regression Assumptions 103](#_Toc118026895)

[C.3.3 Synthetic Chronologies 104](#_Toc118026896)

[C.3.4. Bootstrapping 106](#_Toc118026897)

[C.3.5. Confidence Interval Calculation 108](#_Toc118026898)

[C.3.6. Confidence Interval Testing 110](#_Toc118026899)

[C.4. Results 110](#_Toc118026900)

[C.4.1. Regression Assumptions 110](#_Toc118026901)

[C.4.2. Synthetic Chronologies 111](#_Toc118026902)

[C.4.3 Bootstrapping 113](#_Toc118026903)

[C.4.4 Confidence Interval Testing 114](#_Toc118026904)

[C.5. Discussion 115](#_Toc118026905)

[C.6. Conclusion 118](#_Toc118026906)

[C.7. References 119](#_Toc118026907)

# **List of Figures**

[Figure A.1. Site map and geoduck sample lmage 2](#_Toc116024270)4

[Figure A.2. Sample coverage and radiocarbon. 3](#_Toc116024271)2

[Figure A.3. Empirical comparison of detrending methods in the modern interval. 3](#_Toc116024270)4

[Figure A.4. Seasonality and stability of chronology-SST correlations 3](#_Toc116024270)6

[Figure A.5. SST reconstruction 3](#_Toc116024270)8

[Figure B.1. Study site 75](#_Toc116024270)

[Figure B.2. Pre-bomb radiocarbon](#_Toc116024270) 80

[Figure B.3. Tree Nob and volcanic proxy records 81](#_Toc116024270)

[Figure B.4. Bomb 14C correlations with monthly-averaged climate indicators 82](#_Toc116024270)

[Figure C.1. Synthetic chronology construction](#_Toc116024270) 105

[Figure C.2. Confidence Interval Testing: Intervals and Methods](#_Toc116024270) 109

[Figure C.3. Regression Assumptions Testing](#_Toc116024270) 111

[Figure C.4. Chronology Properties](#_Toc116024270) 112

[Figure C.5. Confidence Interval Performance](#_Toc116024270) 113

[Figure C.6. Correlations between Chronology Properties and Confidence Interval Capture](#_Toc116024270) 114

# LIST OF TABLES

[Table A.1. Crossdating statistics. 31](#_Toc111040046)

Table A.2. Radiocarbon dates and associated data. 33

Table B.1. Radiocarbon data. 79

Table C.1. MEboot. 107

# ABSTRACT

The long-term character and range of northeast Pacific climate variability is largely unknown due to the short period of instrumental record and poor agreement among existing reconstructions. To address this issue, a multi-centennial record of northeast Pacific climate is developed from a new archive, the Pacific geoduck, a long-lived marine bivalve known to form annual growth increments within its shell. The widths of these increments strongly covary with ambient water temperature, and calcium carbonate within these shells contain radiocarbon, precipitated from ambient seawater, and serves as an indicator of ocean circulation. This study describes the development of a multicentennial geoduck chronology, with some chronology segments extending 3000 years before present. The first portion of the study outlines the development of this chronology and the accompanying sea surface temperature (SST) reconstruction from growth-increment widths. The chronology is the first to date dead-collected material from the ocean floor and represents the longest chronology and associated annually resolved SST reconstruction yet developed from any marine organism in the region. In the second portion of the study, radiocarbon is sampled from the geoduck chronology at decadal resolution to quantify water mass variability and assess relationships with SST. This new decadal radiocarbon record is the only record of its kind in the northeast Pacific and describes a relatively stable state that can be interrupted by regimes of cold, radiocarbon-old water. In the final portion of this study, the uncertainty of paleoclimate reconstructions, including the geoduck SST reconstruction, are analyzed to determine the most skillful and robust method of defining uncertainties from crossdated paleoclimate proxies. The results lend support to the Maximum Entropy Bootstrap approach recently introduced to dendroclimatology but not yet widely adopted. Thus, this body of work demonstrates that Pacific geoduck can be utilized as both an SST and a water mass proxy over multiple centuries and that simulation experiments can serve to test novel methods across the growing diversity of crossdated paleoclimate proxies.

# 1. INTRODUCTION

## 1.1. Motivation

Decadal-scale variability in the northeast Pacific Ocean has been linked to several key fisheries, North American hydroclimate, and multiple ecological indicators (Clark et al., 1999; Hare et al., 1999; Chavez et al., 2003; Berkelhammer, 2019). A series of persistent marine heatwaves beginning in 2013, termed “The Blob”, have been linked to unprecedented losses to fisheries and ecosystem disruption (Bond et al., 2015; Di Lorenzo & Mantua, 2016; Amaya et al., 2020). Parameterization of the potential variability in this system is crucial to sound management of fisheries, water, and forest resources, but the capacity to make sound predictions is precluded by the short length of instrumental climate records, which are sparse prior to 1920 in the northeast Pacific, leading to an incomplete understanding of the range of natural variability.

Paleoclimate proxies have been used to reconstruct SST variability in the region, with high resolution proxies dominated by tree rings. Annual growth increments of trees provide excellent dating and sufficient temporal resolution to investigate climate relationships, however, these proxies are impacted only secondarily by SST, and thus may be unreliable proxies. Although many reconstructions of PDO have been undertaken based on tree rings (Biondi et al., 2001; D’Arrigo, 2001; Gedalof & Smith, 2001; MacDonald & Case, 2005; D’Arrigo & Wilson, 2006), there is significant disagreement among the reconstructions and therefore significant uncertainty in the character and coherence of this climate indicator prior to 1920 (Kipfmueller et al., 2012; Newman et al., 2016; Henley, 2017).

An organism residing directly within the environment of interest may yield a better proxy. Several marine bivalves deposit annual growth rings in their shells, and the methods developed for annual tree rings have recently been applied to these archives (eg. Strom et al., 2004; Butler et al., 2010). The exact calendar year of formation of these growth rings can be assigned based on crossdating, just as with trees (Black et al., 2019). Although the climate reconstructions produced from marine bivalves in the northeast Pacific have been relatively short (Strom et al., 2004; Black et al., 2009), the potential for longer reconstructions alongside the void of northeast Pacific paleoclimate knowledge provides impetus for this investigation.

## 1.2. Background

### 1.2.1 The Pacific Geoduck

A commercial fisher for Pacific geoduck exists along the coasts of Washington and Alaska in the United States and British Columbia in Canada. The relationship between geoduck growth increments and ambient marine conditions was first described in 1992 (Noakes and Campbell, 1992). Geoduck have been shown to produce annual growth increments (Strom et al., 2004; Black et al., 2008; Helser et al., 2012). The growth increments of multiple individuals collected within a defined region covary such that wide and narrow increments can be assigned to calendar years by crossdating (Glock & Pearson, 1937; Douglass, 1941; Fritts, 1971). Several recent publications describe chronologies derived from crossdated geoduck in the northeast Pacific with varied relationships to SST (Strom et al., 2004, Strom et al., 2005; Black et al., 2009). Unfortunately, these reconstructions comprise only live-collected geoduck and therefore extend at most to only the late 19th century.

### 1.2.2 Northeast Pacific Climate

SST anomalies along the Pacific coast of North America tend to covary over broad geographic extents, spanning the coast of California through the coast of Alaska. The PDO is the dominant mode of climate variability in the region, defined as the leading Empirical Orthogonal Function of gridded SST in the North Pacific north of 20°N latitude after the global mean anomaly has been subtracted to account for global warming (Mantua et al., 1997). The PDO is the result of interactions among multiple atmospheric and oceanographic drivers, the most influential of which is the Aleutian Low (Newman et al., 2016), which is a semi-permanent low-pressure system centered on the Aleutian Islands that peaks in intensity in the cool-season months. When the low intensifies, southerly winds intensify along the North American coast. This brings warm air masses northward, leading to anomalously warm coastal SSTs. Southerly winds in combination with the earth’s rotation also cause vertical water movement by pushing surface waters toward the coast and forcing them to depth in the process of downwelling, which affects marine biological productivity. Farther offshore, winds driven by the Aleutian Low mix the water column to depths of 100 m or more, (Crawford et al., 1988), which leads to anomalously cool SST anomalies. Thus, periods in which the Aleutian Low is strong, corresponding to positive “warm” phases of the PDO, SSTs are anomalously warm in the coastal northeastern Pacific and anomalously cool in the central northeastern Pacific. The opposite pattern in SST anomalies dominates when the Aleutian Low is weak and the PDO is in a negative “cool” phase.

Much of the paleoclimate work in this region has been driven by interest in the PDO. Although the PDO has been the target of many climate reconstructions (Biondi et al., 2001; D’Arrigo, 2001; Gedalof & Smith, 2001; MacDonald & Case, 2005; D’Arrigo & Wilson, 2006), these reconstructions correlate poorly prior to the twentieth century (Kipfmueller et al., 2012; Newman et al., 2016; Henley, 2017). The poor agreement among reconstructions is likely the result of the variability of teleconnections, the covariability of distant climate phenomena, the inconsistency of the PDO itself as a spatially coherent pattern, or the inconsistency of proxy-climate relationships (Gallant et al., 2013, Batehup et al., 2015; Franke, 2013).

### 1.2.3 Northeast Pacific Surface Currents and Radiocarbon

Water temperature and radiocarbon (14C) in the mixed layer of the Alaska Coastal Current (ACC), the northward flowing surface current extending from Vancouver Island to Anchorage, Alaska, have different sources of variability. SST is largely impacted by atmospheric dynamics, based on wind-driven mixing and interactions with air masses of varying temperature (Newman et al., 2016). The drivers of 14C variability are less well understood, but some candidates may include interactions with the atmospheric 14C reservoir, freshwater inputs, relative abundances of parent water masses, and upwelling (Guilderson et al., 2006; Hristova et al., 2019; Hutchinson, 2020). Some of these sources of 14C water mass variability may also impact SST.

A “water mass” is a volume of water with defined properties such as temperature and salinity imprinted at formation. Both temperature and Carbon-14 (14C) are considered water mass properties, that is, they are generally immutable characteristics. Because the only source of 14C is the atmosphere, its ultimate source, ocean water masses become depleted in 14C (relative to the atmosphere) when they move away from the surface (Stuiver et al., 1986). 14C is commonly utilized as a water mass tracer property, a property that is stable and can be used to identify the movement of a water mass through time. This principal has recently been utilized in the increments of marine bivalves, which offer a record of past marine radiocarbon (eg. Lower-Spies et al., 2020).

The growth increments of the geoduck are comprised of calcium carbonate (CaCO3), and work with other bivalves suggests the carbon therein is derived primarily from the dissolved inorganic caron of the ambient seawater (Adkins et al., 2002; Beirne et al., 2012). The geoduck growth increments could therefore be used as an archive of both water temperature and radiocarbon (14C). Sufficiently long records could improve understanding of the climate, and the collocated temperature/14C records could also be used to better understand the distinct drivers of the water mass properties. Furthermore, a 14C record with precise, independent dating could be used to improve radiocarbon dating of marine samples in the region.

# 2. PRESENT STUDY

The present study focuses on a Pacific geoduck chronology based on both live- and dead-collected samples with growth increments sensitive to water temperature and exactly dated 14C samples. This research comprises three distinct manuscripts. I first built a crossdated chronology from Pacific geoduck samples and used the ring widths to reconstruct SST. I then collected calcium carbonate samples from the dated growth increments to reconstruct 14C variability at the same site. After evaluating the SST reconstruction methods from the first manuscript, I subsequently tested several methods of defining uncertainty in this and similar reconstructions.

## 2.1 Pacific Geoduck Chronology Construction and SST Reconstruction

An existing geoduck chronology at Tree Nob, based on live-collected specimens, was used as the basis for a long chronology, which I extended with new subfossil collections. We chose Tree Nob, Coastal British Columbia, Canada, for its proximity to a long instrumental SST measurement station and the correlation between the growth increment chronology and regional SST (54°13'2.53"N, 130°47'21.59"W, Black et al., 2009). Due to the short lifespan of the bivalves relative to their persistence on the ocean floor, I coauthored a computer program and associated manuscript for identifying possible crossdates (Reynolds et al., 2020). I successfully crossdated 115 dead-collected specimens, spanning 3000 years, and verified the crossdating with radiocarbon samples. I investigated the impacts of detrending on the chronology coherence and climate signal. I determined the growth increment width chronology to aggregate 17 months of SST signal and developed an SST reconstruction. The manuscript demonstrates the viability of Pacific geoduck as a multicentennial proxy of SST and shows the warming SSTs after 1980 are unprecedented in the 400+ years of the reconstruction. The manuscript was published in Paleoceanography and Paleoclimatology in 2021 and is included here with permission as Appendix A.

## 2.2 14C Reconstruction and Water Mass Variability Assessment

The continuous portion of the geoduck chronology at Tree Nob spans 1725-2008. The chronology offers the longest archive of absolutely dated carbonate in the Northeast Pacific, allowing for annual isotopic analysis of both carbon and oxygen over more than 200 years. I designed a sampling plan to extract 10-years intervals of shell carbonate to provide nearly continuous decadal coverage. Samples were measured for 14C content, which I compared with the expected values, given the known ages of the growth increments. The resulting offsets (measured - expected) show a remarkable stability in the radiocarbon reservoir at Tree Nob with a single older-than-expected outlier in the early nineteenth century. A comparison of the radiocarbon offsets to SST shows a weak relationship, with a distinct outlier in the same early nineteenth century period. This manuscript demonstrates the value of the Pacific geoduck as a marine radiocarbon reservoir proxy. The radiocarbon record is the only record of its kind in the northeast Pacific and can serve to increase the accuracy of radiocarbon dating for marine samples in the region. The synchronous and collocated reconstructions of different water mass properties allow for a deeper understanding of past variability. This manuscript is in press at the journal Radiocarbon, with an expected publication date of late 2022 and is included here with permission as Appendix B.

## 2.3 Evaluation of Methods Used to Quantify Reconstruction Uncertainties

The sea surface temperature reconstruction described in section 2.1 and Appendix A includes reconstruction uncertainties. The uncertainties of paleoclimate reconstructions derived from crossdated archives can be calculated in a number of ways, and there are a number of sources of uncertainty. In dendroclimatology the sources of uncertainty are sometimes described as resulting from three distinct areas including detrending, chronology, and calibration. To address the appropriate method for defining uncertainties derived from the chronology and calibration, I reconstructed 3 climate variables, each by 1 real and 100 synthetic chronologies. I used an independent interval to test the skill of the uncertainty of the predictions, in a process with ample redundancy for each chronology. The results lend support to the method supported by the time series character of crossdated climate proxies and introduced to the dendroclimatic community by Ed Cook and colleagues (2013). I intend to submit the manuscript to the journal Dendrochronologia for review in late 2022, and it is included here as Appendix C.

# REFERENCES

Adkins, J. F., Griffin, S., Kashgarian, M., Cheng, H., Druffel, E. R. M., Boyle, E. A., ... & Shen, C. C. (2002). Radiocarbon dating of deep-sea corals. *Radiocarbon*, *44*(2), 567-580. Beirne, E. C., Wanamaker Jr, A. D., & Feindel, S. C. (2012). Experimental validation of environmental controls on the δ13C of Arctica islandica (ocean quahog) shell carbonate. *Geochimica et Cosmochimica Acta*, *84*, 395-409.

Amaya, D. J., Miller, A. J., Xie, S.-P., & Kosaka, Y. (2020). Physical drivers of the summer 2019 North Pacific marine heatwave. *Nature communications*, *11*(1), 1-9.

Batehup, R., McGregor, S., & Gallant, A. (2015). The influence of non-stationary teleconnections on palaeoclimate reconstructions of ENSO variance using a pseudoproxy framework. *Climate of the Past*, *11*(12), 1733-1749.

Berkelhammer, M. (2019). Synchronous modes of terrestrial and marine productivity in the North Pacific. *Frontiers in Earth Science*, *7*, 73.

Biondi, F., Gershunov, A., & Cayan, D. R. (2001). North Pacific decadal climate variability since 1661. *Journal of Climate*, *14*(1), 5-10.

Black, B. A., Copenheaver, C. A., Frank, D. C., Stuckey, M. J., & Kormanyos, R. E. (2009). Multi-proxy reconstructions of northeastern Pacific sea surface temperature data from trees and Pacific geoduck. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *278*(1-4), 40-47.

Black, B. A., Andersson, C., Butler, P. G., Carroll, M. L., DeLong, K. L., Reynolds, D. J., ... & Witbaard, R. (2019). The revolution of crossdating in marine palaeoecology and palaeoclimatology. *Biology letters*, 15(1), 20180665.

Black, B. A., Gillespie, D. C., MacLellan, S. E., & Hand, C. M. (2008). Establishing highly accurate production-age data using the tree-ring technique of crossdating: a case study for Pacific geoduck (Panopea abrupta). *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(12), 2572-2578.

Bond, N. A., Cronin, M. F., Freeland, H., & Mantua, N. (2015). Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophysical Research Letters*, *42*(9), 3414-3420.

Butler, P. G., Richardson, C. A., Scourse, J. D., Wanamaker Jr, A. D., Shammon, T. M., & Bennell, J. D. (2010). Marine climate in the Irish Sea: analysis of a 489-year marine master chronology derived from growth increments in the shell of the clam Arctica islandica. *Quaternary Science Reviews*, *29*(13-14), 1614-1632.

Chavez, F. P., Ryan, J., Lluch-Cota, S. E., & Ñiquen, M. (2003). From anchovies to sardines and back: multidecadal change in the Pacific Ocean. *Science*, *299*(5604), 217-221.

Clark, W. G., Hare, S. R., Parma, A. M., Sullivan, P. J., & Trumble, R. J. (1999). Decadal changes in growth and recruitment of Pacific halibut (Hippoglossus stenolepis). Canadian Journal of Fisheries and Aquatic Sciences, 56(2), 242-252.

Cook, E. R., Palmer, J. G., Ahmed, M., Woodhouse, C. A., Fenwick, P., Zafar, M. U., ... & Khan, N. (2013). Five centuries of Upper Indus River flow from tree rings. *Journal of hydrology*, *486*, 365-375.

Crawford, W. R., Huggett, W. S., & Woodward, M. J. (1988). Water transport through Hecate Strait, British Columbia. *Atmosphere-ocean*, *26*(3), 301-320.

D'Arrigo, R., & Wilson, R. (2006). On the Asian expression of the PDO. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *26*(12), 1607-1617.

D'Arrigo, R., Villalba, R., & Wiles, G. (2001). Tree-ring estimates of Pacific decadal climate variability. *Climate Dynamics*, 18(3-4), 219-224.

Di Lorenzo, E., & Mantua, N. (2016). Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nature Climate Change*, *6*(11), 1042-1047.

Douglass, A. E. (1941). Crossdating in dendrochronology. *Journal of Forestry*, *39*(10), 825-831.

Franke, J., Frank, D., Raible, C. C., Esper, J., & Brönnimann, S. (2013). Spectral biases in tree-ring climate proxies. *Nature Climate Change*, *3*(4), 360-364.

Fritts, H. C. (1971). Dendroclimatology and dendroecology. *Quaternary Research*, *1*(4), 419-449.

Gallant, A. J., Phipps, S. J., Karoly, D. J., Mullan, A. B., & Lorrey, A. M. (2013). Nonstationary Australasian teleconnections and implications for paleoclimate reconstructions. *Journal of Climate*, *26*(22), 8827-8849.

Gedalof, Z. e., & Smith, D. J. (2001). Interdecadal climate variability and regime‐scale shifts in Pacific North America. *Geophysical Research Letters*, *28*(8), 1515-1518.

Glock, W. S., & Pearson, G. A. (1937). *Principles and methods of tree-ring analysis*. Carnegie institution of Washington Washington, DC.

Guilderson, T. P., Roark, E. B., Quay, P. D., Page, S. R. F., & Moy, C. (2006). Seawater radiocarbon evolution in the Gulf of Alaska: 2002 observations. *Radiocarbon*, *48*(1), 1-15.

Hare, S. R., Mantua, N. J., & Francis, R. C. (1999). Inverse production regimes: Alaska and west coast Pacific salmon. Fisheries, 24(1), 6-14.

Helser, T. E., Kastelle, C. R., & Lai, H.-l. (2014). Modeling environmental factors affecting assimilation of bomb-produced Δ14C in the North Pacific Ocean: Implications for age validation studies. *Ecological modelling*, *277*, 108-118.

Henley, B. J. (2017). Pacific decadal climate variability: Indices, patterns and tropical-extratropical interactions. *Global and Planetary Change*, *155*, 42-55.

Hristova, H. G., Ladd, C., & Stabeno, P. J. (2019). Variability and trends of the Alaska Gyre from Argo and Satellite Altimetry. *Journal of Geophysical Research: Oceans*, *124*(8), 5870-5887.

Hutchinson, I. (2020). Spatiotemporal variation in ΔR on the West Coast of North America in the late Holocene: implications for dating the shells of marine mollusks. *American Antiquity*, *85*(4), 676-693.

Kipfmueller, K. F., Larson, E. R., & St. George, S. (2012). Does proxy uncertainty affect the relations inferred between the Pacific Decadal Oscillation and wildfire activity in the western United States? *Geophysical Research Letters*, *39*(4).

Lower‐Spies, E. E., Whitney, N. M., Wanamaker, A. D., Griffin, S. M., Introne, D. S., & Kreutz, K. J. (2020). A 250‐year, decadally resolved, radiocarbon time history in the Gulf of Maine reveals a hydrographic regime shift at the end of the Little Ice Age. *Journal of Geophysical Research: Oceans*, 125(9), e2020JC016579.

MacDonald, G. M., & Case, R. A. (2005). Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters*, *32*(8).

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, *78*(6), 1069-1080.

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., Mantua, N. J., Miller, A. J., Minobe, S., & Nakamura, H. (2016). The Pacific decadal oscillation, revisited. *Journal of Climate*, *29*(12), 4399-4427.

Reynolds, D. J., Edge, D. C., & Black, B. A. (2021). RingdateR: A statistical and graphical tool for crossdating. *Dendrochronologia*, *65*, 125797.

Strom, A., Francis, R. C., Mantua, N. J., Miles, E. L., & Peterson, D. L. (2004). North Pacific climate recorded in growth rings of geoduck clams: a new tool for paleoenvironmental reconstruction. *Geophysical Research Letters*, *31*(6).

Strom, A., Francis, R. C., Mantua, N. J., Miles, E. L., & Peterson, D. L. (2005). Preserving low-frequency climate signals in growth records of geoduck clams (*Panopea abrupta*). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *228*(1-2), 167-178.

Stuiver, M., Pearson, G. W., & Braziunas, T. (1986). Radiocarbon age calibration of marine samples back to 9000 cal yr BP. *Radiocarbon*, *28*(2B), 980-1021.

# Appendix A

A MULTICENTENNIAL PROXY RECORD OF NORTHEAST PACIFIC SEA SURFACE TEMPERATURES FROM THE ANNUAL GROWTH INCREMENTS OF PANOPEA GENEROSA

Edge, D. C., Reynolds, D. J., Wanamaker, A. D., Griffin, D., Bureau, D., Outridge, C., Stevick B.C., Weng, R., & Black, B. A. (2021). A Multicentennial Proxy Record of Northeast Pacific Sea Surface Temperatures From the Annual Growth Increments of Panopea generosa. Paleoceanography and Paleoclimatology, 36(9), e2021PA004291. DOI: 10.1029/2021PA004291

*The Licensed Content Publisher, John Wiley and Sons, has granted the author permission to reproduce the original article, “A Multicentennial Proxy Record of Northeast Pacific Sea Surface Temperatures From the Annual Growth Increments of Panopea generosa”, in this dissertation. License Number: 5365501192995, License date: Aug 10, 2022.*

**A Multicentennial Proxy Record of Northeast Pacific Sea Surface Temperatures from the Annual Growth Increments of *Panopea generosa***

David C. Edge1, David J. Reynolds2, Alan D. Wanamaker3, Daniel Griffin4, Dominique Bureau5, Christine Outridge1, Bethany C. Stevick6, Richard Weng1, and Bryan A. Black1

1 Laboratory of Tree Ring Research, University of Arizona, Bryant Bannister Tree Ring Building, University of Arizona, 1215 E Lowell St, Tucson, AZ, 85721, USA

2 Centre for Geography and Environmental Science, College of Life and Environmental Science, University of Exeter, Penryn Campus, Treliever Road, Penryn, Cornwall, TR10 9FE, UK

3 Department of Geological and Atmospheric Sciences, Iowa State University, 2237 Osborn Dr, Ames, IA 50011, USA

4 Department of Geography, Environment & Society, University of Minnesota, 267 19th Ave S, Minneapolis, MN 55455, USA

5 Department of Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, British Columbia, Canada

6 Washington Department of Fish and Wildlife, 1111 Washington Street SE, Olympia, Washington, USA

## A.1. Abstract

Growth-increment widths of Pacific geoduck (*Panopea generosa*), a long-lived bivalve, are used to develop the first marine-based, multicentennial, annually resolved, and exactly dated archive of Northeast Pacific sea surface temperatures (SST). The chronology is sampled from the Tree Nob Islands, British Columbia, Canada, continuously covers 1725–2008, and also contains nine older radiocarbon-dated segments, which together span 58% of the last 1500 years. Age-related growth declines were removed by aligning all increments relative to age of increment formation and fitting with a single detrending curve to preserve low-frequency signals. The geoduck chronology was used to reconstruct local SST variability across the seasonal window of April through November. The chronology at both the concurrent (lag-0) and following (lag+1) year are both highly significant predictors of SST in a stepwise multiple linear regression, explaining 54% of the variance in the period of instrumental overlap (1940–2001), passing strict tests of calibration-verification. Reconstructed SSTs contained significant spectral power at periods from 3 to 64 years, suggesting that 20th century variability in these periodicities is not unusual in the longer-term context. The period of lowest growth coincided with the Dalton minimum, an episode of reduced solar irradiance from 1790–1830, as well as the 1809 Unknown eruption, suggesting that solar and volcanic signals are present in the SST history. The most conspicuous aspect of the reconstruction is the steady and unprecedented warming trend that began in the mid-1800s and continues through present. The post-1976 interval includes the two warmest decades of the reconstruction.

## A.2. Introduction

In terrestrial environments, tree-ring data are a critically important indicator of long-term climate and environmental variability, especially at mid latitudes. A distinguishing characteristic of tree-ring data amongst climate proxies is that they are well replicated, annually resolved, and absolutely dated through the process of crossdating, in which synchronous patterns induced by climate are matched among individuals of a given species and location (Glock & Pearson, 1937; Douglass, 1941; Fritts, 1971). By this method, growth irregularities such as false rings, micro-rings, or locally absent rings can be identified such that each increment in the dataset is assigned its correct year of formation (Fritts, 1976). For live-collected samples, the year of collection anchors the absolute dating of the chronology. Where available, dead-collected material of unknown antiquity can be crossdated with one another and the live collected record. In doing so, chronologies can be generated that are much longer than the average individual lifespan for the species, in some cases spanning multiple millennia (Pilcher et al., 1984; Ferguson & Graybill, 1983). Such exact dating facilitates seamless integration of chronologies with one another and instrumental records (Briffa et al., 1996; Mann & Jones, 2003).

Over the past two decades, this same crossdating technique has been applied to growth increments formed in the hard parts of marine organisms to reconstruct environmental variability prior to the start of instrumental records (Black et al., 2019). This approach has been especially successful with long-lived bivalves *Arctica islandica* and *Glycymeris glycymeris* in the North Atlantic, resulting in a network of continuous, multicentennial chronologies from the Gulf of Maine through northern Norway (Butler et al., 2009; Butler et al., 2010; Wanamaker et al., 2012, Reynolds et al., 2017, Wanamaker et al., 2019). A chronology near Iceland spans more than a thousand years and provides exceptionally long-term perspectives of North Atlantic marine dynamics (Butler et al., 2013; Reynolds et al., 2016). In the Northeast Pacific (NEP), a network of crossdated growth-increment width chronologies with strong sensitivity to sea surface temperature (SST) has been developed from the long-lived bivalve geoduck (*Panopea generosa*; Strom 2004; Black et al., 2009). However, these chronologies have involved only live-collected material and the temporal extent is therefore limited by the maximum longevity of the species, which is approximately 150 yrs.

Although live-collected geoduck often doubles the length of the observational record in the coastal NEP (Black et al., 2009), longer-term perspectives on SST are needed. NEP SSTs vary at low frequencies and are characterized by rapid regime shifts associated with the restructuring of marine ecosystem species composition, food web structure, and energy flows (Clark et al., 1999; Hare et al., 1999; Chavez et al., 2003) while also synchronous with drought, snowpack, and fire frequency in western North America (Mote, 2006; Kitzberger et al., 2007; Berkelhammer, 2019). On even broader spatial scales, NEP SSTs have linkages to variability in global-scale temperature patterns (Meehl et al., 2011; Meehl et al., 2013; England et al., 2014; Thompson et al., 2015; Yin et al., 2018). Given the importance of NEP SSTs, there have been numerous attempts to reconstruct its longer-term history, largely from tree-ring data, but these histories poorly agree with one another prior to the twentieth century, leaving considerable uncertainty regarding pre-industrial ranges of climate variability (Kipfmueller et al., 2012; Newman et al., 2016; Henley, 2017).

Here, we use growth-increment widths of live- and dead-collected Pacific geoduck (*Panopea generosa*) to construct the first marine-based, multicentennial, annually resolved, and exactly dated archive of SST in the NEP. The crossdated chronology continuously spans 1725–2008 and also contains well-replicated, radiocarbon-dated segments that in combination span 58% of the last 1500 years. We explore ways to preserve not only the high-frequency variability inherent in the crossdated chronologies, but also use the robust replication and temporal depth of the sample population to preserve low-frequency variability. In so doing, we develop a reconstruction of regional SSTs along the British Columbia, Canada coast to address pre-1900 SST variability and evaluate the magnitude of 19th and 20th century warming.

## A.3. Materials and Methods

### A.3.1 Sample Collection and Preparation

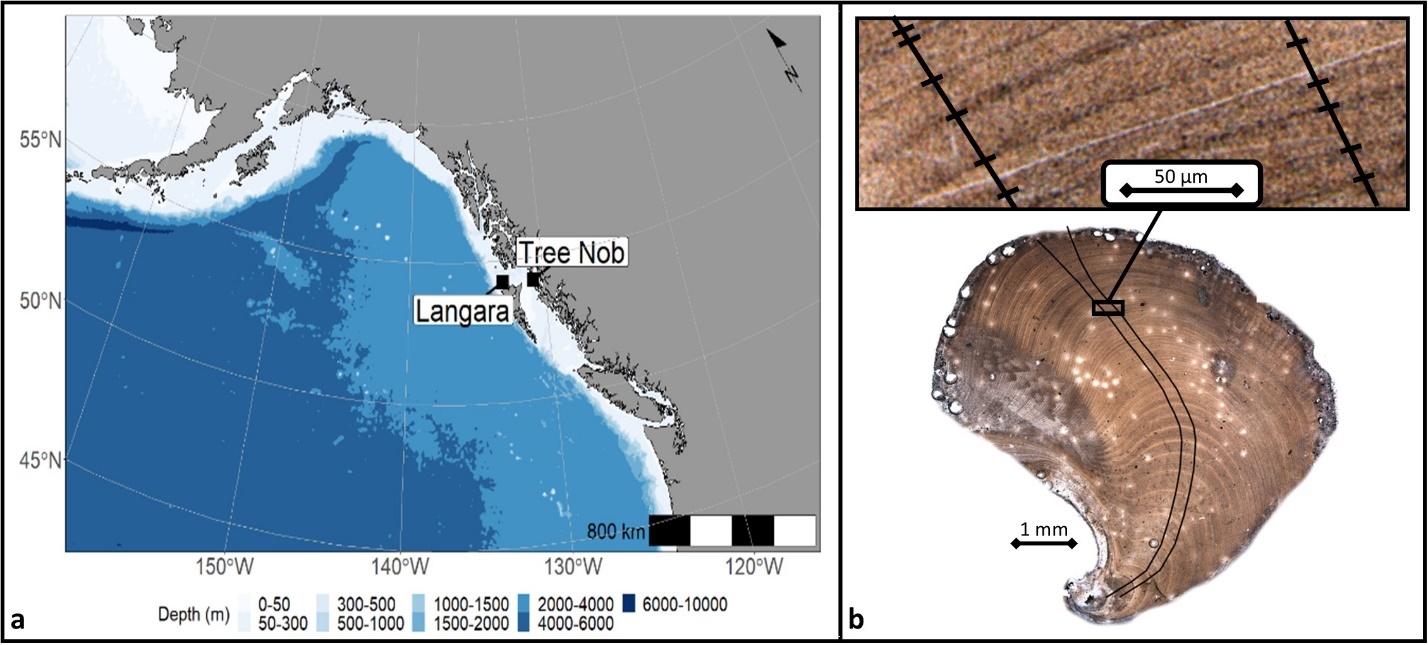
A Pacific geoduck growth-increment width chronology had been previously developed using live-collected samples from the Tree Nob Islands in northern British Columbia (Fig A.1; Black et al., 2009). Given the consistently great longevity of live-collected samples at this site, strong relationship with SST, and presence of dead shell on the ocean floor, Tree Nob was chosen as an ideal site for developing an extended chronology with dead-collected material. On June 26, 2018, professional divers from Trident Biologicals, Nanaimo, BC, Canada, excavated a pit approximately 2 m wide and 2 m deep at approximately 10 m depth at 54°13'2.53"N, 130°47'21.59"W (Fig A.1). A Venturi device was employed to move large quantities of sediment while geoduck shells were manually collected. Due to the instability of the bottom sediments, shells often slid from their original locations before they could be collected, complicating efforts to maintain stratigraphic order of the samples. Approximately 500 shells were gathered, occurring in various taphonomic states from highly preserved with a complete periostracum to heavily eroded with a relatively friable, “chalky” shell texture. In a small percentage (< 5%) valves remained attached by an intact ligament, but the vast majority were single valves assumed to be unique individuals.

The length (posterior to anterior), height (dorsal to ventral), width (shell margin to outer shell apex), and weight (mass in grams) of each valve was measured. We also used a subjective 5-level classification system for recording the general states of bioerosion, shell margin deterioration, and periostracum of each shell as potential proxy measures of shell antiquity (Butler et al., 2020).

The length (posterior to anterior), height (dorsal to ventral), width (shell margin to outer shell apex), and weight (mass in grams) of each valve was measured. We also used a subjective 5-level classification system for recording the general states of bioerosion, shell margin deterioration, and periostracum of each shell as potential proxy measures of shell antiquity (Butler et al., 2020).

The methods for shell preparation and acetate peel preparation were adapted from Richardson (2001). A 2 cm transverse section containing the midline was cut from the dorsal to ventral margins of the valve. The valve section was then embedded in two-part EpoxiCure from Buhler (Lake Bluff, IL) within a silicone mold. The embedded sample was cut along the midline of the umbo (Scourse et al., 2006, Fig A.2) with an IsoMet 5000 precision saw. One of the resulting halves was polished on an EcoMet 30 grinding wheel with 600-, 800-, 1000-, 1400- grit paper followed by 3-micron diamond paste. The polished sample was etched by bathing in 0.1 molar HCl for 135 seconds and immediately rinsed in tap water. Once completely dry, the minimum amount of acetone required to coat the surface was applied. A 125 µm acetate sheet was immediately placed over the acetone. After 25 minutes, the acetate sheet, partially dissolved by the acetone into the etched shell surface, was peeled away and placed between two glass microscope slides (Fig A.1b, additional details in Supplemental Material A.S.1.).

Figure A.1. Site map and geoduck sample lmage. a: Locations of Tree Nob geoduck collection site and a nearby SST and SSS instrumental station at Langara Island. b: Image of geoduck hinge with two measurement axes overlaid



Samples were imaged at 80-200x using a Leica M205 C stereo microscope and a Leica DMC 5400 20-megapixel camera. Increment widths were measured using the caliper tool in Image Pro Premier 9.3 (Media Cybernetics, Rockville, MD). A polyline was drawn along the curved axis of maximum growth, perpendicular to each increment. Increment boundaries were manually marked at the outer edge of the winter line (dark bands in Fig A.1b). Each individual was measured twice along separate axes of measurement. Although these axes were by necessity in close proximity to one another, the repeated measurements helped reduce error due to placement of increment boundaries on the winter line, which could occasionally be diffuse, especially in very narrow increments. All measurements were exported with micron precision.

### A.3.2 Crossdating

The existing absolutely dated Tree Nob chronology was used as a starting point for our dating (Black et al., 2008). Dead collected samples were crossdated against the live-collected chronology and then other dead-collected samples using a general ‘skeleton plotting’ approach. In skeleton plotting, the high-frequency (year-to-year) growth pattern in the sample is transferred to graph paper. Then skeleton plots are manually slid past one another to search for matches (Stokes and Smiley, 1968). Given the many pairwise comparisons possible, this can be a very time intensive procedure. To greatly expedite this process, a cross-correlation procedure analogous to skeleton plotting was performed in RingdateR (Supplemental Material A.S.2.; Reynolds et al., 2021), which extracts high-frequency (year-to-year) variability from each set of measurements and then explores all pairwise combinations and lags for possible matches. Because RingdateR automatically recommends matches from datasets involving many measurement time series, the time spent crossdating is primarily in confirming matches and revising missed/false rings. Time series plots and heat maps of running correlations can be used to help identify any missing or falsely added rings that can be visually confirmed by re-inspecting the sample (see Reynolds et al., 2021, Fig A.2b). Ultimately, we employed no statistical threshold for crossdating, rather, the suggested crossdating was confirmed by meeting three criteria: 1) one and only one lag produced a correlation far greater than all other possible lags, 2) running correlations were highly positive and stable throughout the period of overlap, and 3) visual inspection of the sample identified no errors in growth-increment interpretation or major distortions in growth.

Samples were initially added to the live-collected Tree Nob chronology or compared against one another to form chronologies “floating” in time. Samples that remained undated were regularly reexamined with the addition of new series. Crossdated series were checked for errors in RingdateR, and crossdating was additionally verified in Cofecha (Holmes, 1983). The default spline length of 32 years was reduced to 22 years to improve the common signal amongst the standardized series (Black et al., 2008). Mean sensitivity, a coarse indicator of year-to-year variability (Strackee and Jansma, 1992; Bunn et al., 2013), as well as series intercorrelation was calculated for the absolutely dated and floating chronologies using the Dendrochronology Program Library in R (dplR; Bunn, 2008).

### A.3.3 Radiocarbon Dating

Radiocarbon sampling and analysis was performed to date floating chronologies with replication greater than two individuals. Fourteen carbonate samples were collected for radiocarbon dating. Samples containing approximately ten growth years were obtained from the ventral margin (area of newest growth) using a diamond-grit cutting wheel, variable speed rotary tool. Two samples were taken from the absolutely dated (Modern) chronology, two each from Floating Chronologies 1–3 (the floating chronologies with replication >3 individuals), and one each from the remaining floating chronologies. Radiocarbon (14C) was analyzed at the National Ocean Sciences Accelerator Mass Spectrometry (NOSAMS) facility located in Woods Hole, MA, USA. All radiocarbon determinations were calibrated using the Marine20 radiocarbon curve (Heaton et al., 2020) using OxCal Version 4.3 (Ramsey, 1995).

In the paired samples taken from Floating Chronologies 1–3, one was taken near the beginning and one near the end of each chronology. The offset in the 14C dates between the paired samples was checked for consistency with the number of intervening years along the chronology, which provided a coarse test of crossdating accuracy. Subsequently, a Bayesian ‘wiggle matching’ approach was implemented in OxCal to further refine the 14C age estimates and uncertainties using the paired radiocarbon probability distributions and their respective known sclerochronological offsets (Fig A.4 inset; de Vries, 1958; Ramsey et al., 2001). The two samples from the Modern chronology (TND031 and TND045) were used to calculate the local reservoir age correction (DR) following methods outlined by Lower‐Spies et al. (2020), which was applied to all floating segments to improve dating accuracy.

### A.3.4 Chronology Construction

Once verification of crossdating was complete, measurement time series were standardized and averaged into population-level growth chronologies. Similar to trees, the radial growth increments of geoduck exhibit exponential decay, and these ontogenetic trends must be removed to isolate environmental signals. We explored two detrending techniques, the first of which was negative exponential (EXP) wherein a separate modified negative exponential function was fit to each measurement time series in the dataset after which observed values were divided by those predicted (Black et al., 2009; Butler et al., 2010; Butler et al., 2013). The other approach was regional curve standardization (RCS) in which all measurement data were aligned with respect to biological age and fit with a single age-varying spline (called the regional curve, RC; additional details in Supplemental Material A.S.4.; Melvin et al., 2007) prior to dividing observed values by those predicted (Mitchell 1967; Butler et al., 2010). Biases, however, could be introduced into an RCS chronology if long-lived individuals tend to grow more slowly than their shorter-lived counterparts, especially if growth from long-lived individuals dominates the early portion of the dataset (Schulman, 1954; Black et al., 2008). To identify any relationships between growth and lifespan, measurement series were subdivided into groups with respect to longevity among which age-specific increment width was compared (Esper et al., 2003). Bias may also be introduced if a climatic trend is imprinted on the RC, as could happen in the case of contemporaneous birth of a significant subset of samples (Melvin & Briffa, 2008). To address this issue, Melvin and Briffa developed ‘signal-free detrending,’ an iterative process in which each measurement time series is first divided by the mean chronology to remove the common, population-wide “signal”. This “signal-free” measurement series is then detrended, and the process is repeated until there is minimal difference between the current and prior detrending curve.

Geoduck growth is characterized by rapid decline during its first ~15 years followed by many decades of extremely slow growth. In trees, long periods of very slow growth can affect the fit of negative exponential detrending curves, inducing artifacts in variance and mean trends, especially near the beginning and end of the chronology (Cook & Peters 1997). A powerful and common solution is to apply an adaptive power transformation (EXPapt, RCSapt) to each measurement time series and then calculate residuals (subtraction) from the best-fit curve rather than indices (division; Cook & Peters 1997). A single chronology was developed for each of these approaches by averaging all standardized time series with respect to calendar year. The robust bi-weight mean was used to reduce the impact of outliers (Cook & Kairiukstis, 1990). All detrending and chronology construction was performed in R (R Core Team, 2013) using many tools of dplR (Bunn, 2008).

We reviewed the results of EXP and RCS detrending (with/without APT) on the detrended series and resultant chronologies to determine which approach best suited the data. This was accomplished by examining the detrended time series for evidence of artifacts such as ontogenetic trends in variance or clearly spurious trends in mean. The strength of the patterns shared among samples was quantified by 30-year-windowed (15-year overlap) mean interseries correlation () and expressed population signal (EPS; Wigley et al., 1984; Equation 3.44 in Cook & Kairiukstis, 1990). EPS is calculated as  
where N is the total sample size. An EPS exceeding 0.85 has been used canonically as a minimum threshold for dendroclimatic reconstruction, though this value is arbitrary and must be interpreted with caution given that non-climatic processes could also give rise to common variance (Buras, 2017). In this study, EPS is also used as a metric of shared variance to interpret the relative efficacy of various detrending techniques. In an important limitation, the EPS calculation does not account for the difference in mean between individual series resulting from RCS detrending (Jones et al., 2009). Thus, a more rigorous standard was applied in which the EPS values of RCS chronologies were adjusted based on the ratio of variance in the indices to the variance in the corresponding 21-year spline-detrended indices (EPSadj; Edge, 2021; Melvin & Briffa, 2014). We employed the more flexible 21-year spline (which tends to further reduce the EPSadj values) rather than the recommended 50-year spline to account for the shorter longevity and steep ontogenetic trend in geoduck growth compared with most tree species. Buras (2017) also suggests the use of subsample signal strength rather than EPS, however, we opt for a minimum sample size cutoff based on bootstrapped reconstruction residuals (detailed in section 2.6). The final chronologies were judged based on measures of chronology variance and by analysis of a smoothed periodogram with variable taper.

### A.3.5 Climate-Growth Relationships

Both temperature and sea surface salinity (SSS) have been connected to bivalve shell growth (Noakes & Campbell, 1992; Strom, 2004; Hiebenthal et al., 2012; Pourmozaffar, et al., 2020), and SST is known to covary with Pacific geoduck shell growth at Tree Nob (Black et al., 2009). Lewis and Cerrato (1997) hypothesized that shell growth in many bivalves is related to temperature through metabolic rate based on laboratory manipulations of temperature and food, wherein shell growth was found to continue despite loss of tissue mass in *Mya arenaria*. And Storr et al. (1982) found shell growth to be directly related to temperature up to 13 °C in another subtropical bivalve, *Mercenaria mercenaria*. We examined instrumental records of both SST and SSS for relationships to the Tree Nob chronology. SST and SSS have been recorded daily at Langara Lighthouse, approximately 145 km northwest of Tree Nob, since 1936 (Fig A.1, http://dfo-mpo.gc.ca). The period of 1940-2001 was selected for the climate-growth analysis based on replication in the chronology (n≥7) and data continuity at Langara. Daily data were aggregated to monthly means, and years with greater than three consecutive missing monthly values were discarded. Missing monthly SST and SSS values were filled using an interpolation algorithm in the forecast package in R that models seasonal patterns (Hyndman & Khandakar, 2007). We calculated correlations between the Tree Nob chronology and monthly-averaged SSS and SST over a 24-month period spanning January of the year prior to ring formation through December of the concurrent year. Significance of individual correlations was tested (α = 0.05) using the ‘exact’ simulation method to account for autocorrelation, wherein nonparametric estimates of series spectra were used to generate surrogates for a bootstrapping technique (Percival & Constantine, 2006; Meko et al., 2011). Those months that correlate most strongly with a chronology are generally sequential. Thus, we averaged across those months with peak correlations within SST and then within SSS. The SST mean series and the SSS mean series were entered as predictors of the Tree Nob chronology in a multiple linear regression using the MASS package in R with a p < 0.01 level to enter (Ripley et al., 2013). The parsimony of the model was further tested using Akaike Information Criterion (Sakamoto et al., 1986). This regression procedure allowed us to identify whether SST, SSS, or their combination were significant predictors of the Tree Nob chronology.

### A.3.6 SST Reconstruction

We used linear regression over the 1940–2001 interval to generate the reconstruction (note that the reconstruction ends in 2000 due to the use of forward-lagged chronology values [lag+1]). First, however, we log transformed the Tree Nob chronology to normalize a right-skewed distribution characterized by outlier years of positive growth (Fritts, 1976; Menesguen & Dreves, 1987). Regression coefficients were used to reconstruct SST at Langara Lighthouse over all chronology intervals with sufficiently high EPS values, including floating segments. We calculated three skill metrics wherein both the early (1940–1970) and late (1971–2000) periods were used alternatively as independent calibration and verification intervals (Mann & Rutherford, 2002). The first skill metric is mean squared error (MSE), which is a measure of the difference between the predicted and observed values (Gauss, 1821; Lehmann & Casella, 2006). The second is the reduction of error (RE) statistic, which is a measure of the fit of the reconstruction in the calibration interval relative to the fit of the mean of the target data (Fritts, 1976). Finally, the coefficient of efficiency (CE) is calculated similarly to RE but in the verification interval (Cook et al., 1994). A value > 0 for the RE and CE statistics indicates the reconstruction is a better predictor than the mean of the target, though the target mean is not an appropriate threshold to test skill when the target time series contains significant autocorrelation (Macias-Fauria et al., 2012). Therefore, we used a Monte Carlo phase randomization simulation in which a random time series of the same length and spectral properties as SST (Ebisuzaki, 1997) was generated using the Astrochron package in R (Meyers, 2014). The RE, CE, and MSE were calculated from this simulated SST series and the true SST series. This process was repeated 10,000 times, and the 99th percentile values were retained as significance thresholds (α = 0.01) (Macias-Fauria et al., 2012; Edge, 2021). Also, the empirical prediction intervals of the reconstruction were calculated as the median and 90-percentile residuals obtained between the reconstruction and the instrumental SST in the verification interval (Lee and Scholtes, 2014).

The reconstruction preiod was limited to years with a sample size (number of individuals representing a single calendar year; Fig A.2) sufficient to represent the population. To find the minimum necessary sample size, we used 1000 bootstrapped simulations at each possible sample size to determine at what sample size the MSE nears an asymptotic minimum. The simulated chronologies were created by sampling with replacement at each year to build chronologies of uniform sample size. The MSE was calculated for each simulated chronology over the full period of instrumental overlap. The median MSE of the 1000 simulations at each sample size was used to represent the error presesent in a chronology with the given sample size. A time series of these representative errors was used to find the minimum sample size required to adequately represent the population.

Wavelet analysis was performed on each segment of the reconstruction to assess frequency characteristics through time (Torrence & Compo, 1998) using the WaveletComp package in R (Roesch & Schmidbauer, 2018). Significance (p-value) was calculated by comparing the wavelet results against the wavelets of 10,000 simulated time series with the same autocorrelation and moving average structure (ARMA) as the reconstruction (Hyndman & Khandakar, 2007).

## A.4 Results

### A.4.1 Sample Collection and Preparation

Table **Error! No text of specified style in document.**.1. Note: Only chronology segments with sample size greater than 3 included in table.

In total, 262 shells were processed. After processing the first 50 randomly selected shells we determined whether biometric measurements could be used to broadly estimate longevity, defined as total number of growth increments in the sample. To provide precise ages, only samples that could be crossdated and measured through the terminal growth year were used. Damaged shells were also discarded from this calculation so that only samples with accurate masses were used. We found a significant relationship between sample mass and longevity (R2 = 0.383, n = 15, p = 0.014), so going forward we prioritized the processing of those shells over 70g

Figure **Error! No text of specified style in document.**.1

to eliminate most individuals less than ≈ 50 years in age. We found that, in general, measurement time series of at least 50 yr were necessary to extend the chronologies.



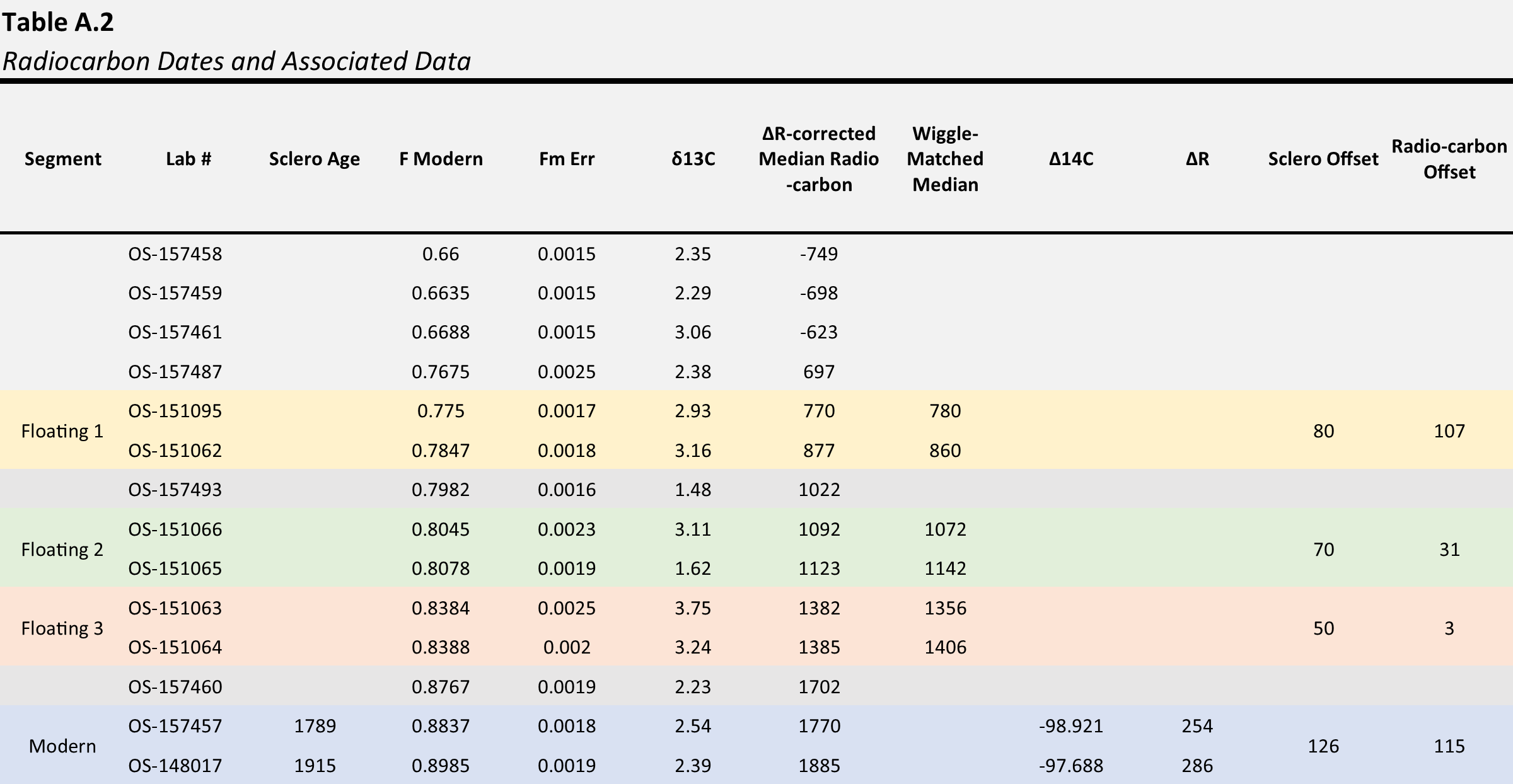
Figure A.2. Sample Coverage and Radiocarbon. Main panel: Segment Length Plot. Each line segment represents a sample with coverage of a precise number of calendar years. Each color represents a crossdated ‘floating’ chronology. Black asterisks denote radiocarbon samples taken from absolutely dated individuals to make local reservoir correction. Black triangles denote radiocarbon samples taken from floating chronologies for approximate dating. Inset: Calculating local reservoir correction. Blue ribbon – Marine radiocarbon curve given by Marine20. Red and green shaded regions show probability distributions of sample ages given by radiocarbon analysis. Black asterisks mark the true ages of samples given by crossdating. Offsets between median radiocarbon dates and true crossdate given. The local radiocarbon reservoir correction is given by the average of these two values 290 ± 45 years.

### A.4.2 Crossdating

Table 1. Crossdating statistics from the absolutely dated (‘Modern’) chronology and the three floating chronologies with greatest replication (Floating 1-3).

Of the 262 samples processed, 115 were successfully crossdated into 10 chronologies spanning more than 1200 years (Fig A.2). The median segment length of the crossdated individuals is 55.5 years, while the maximum longevity is 179 years. This is the longest-lived geoduck known, surpassing a 168-year-old individual (Bureau, 2002). Indices of bioerosion, shell margin deterioration, and periostracum preservation were compared to date of death, as established via crossdating and radiometric techniques. However, the relationships were not statistically significant. The series intercorrelations (the mean of correlation between each individual and the average of all others) of the four well-replicated chronologies averages approximately 0.8 (Table 1). The Modern chronology spans the interval from 1725 to 2008.

**Segment**: chronology interval location of sample, yellow, green, red, and blue coloring shows membership to a named, well-replicated chronology segment while gray indicates membership to one of the unnamed, poorly -replicated segments. **Lab#**: Sample ID given by NOSAMS. **Sclero Age**: Sample calendar age based on crossdating. **F Modern**: Sample’s radiocarbon content relative to modern standard as reported by NOSAMS. **Fm Err**: NOSAMS reported 1σ error of Fraction modern. **δ13C**: ratio of 13C:12C, reported in parts per thousand. **ΔR-corrected Median Radio-carbon**: median date based on F Modern after ΔR correction in OxCal. **Wiggle-Matched Median**: median date based on F Modern of paired samples using Sclero Offset after ΔR correction in OxCal. **Δ14C**: radiocarbon age based on the equation Δ14C = (Fm \* eλ(y – x) – 1) \* 1000. **ΔR**: local reservoir correction, relative to Marine20 (Heaton et al., 2020). **Sclero Offset**: Span in years between the innermost and outermost radiocarbon sampled in a series. Values are rounded to nearest 5 when the small, outer rings of sample could not be precisely counted. **Radiocarbon Offset**: Span in years between ΔR-corrected Median Radiocarbon dates of the sample pair.



### A.4.3 Radiocarbon dating

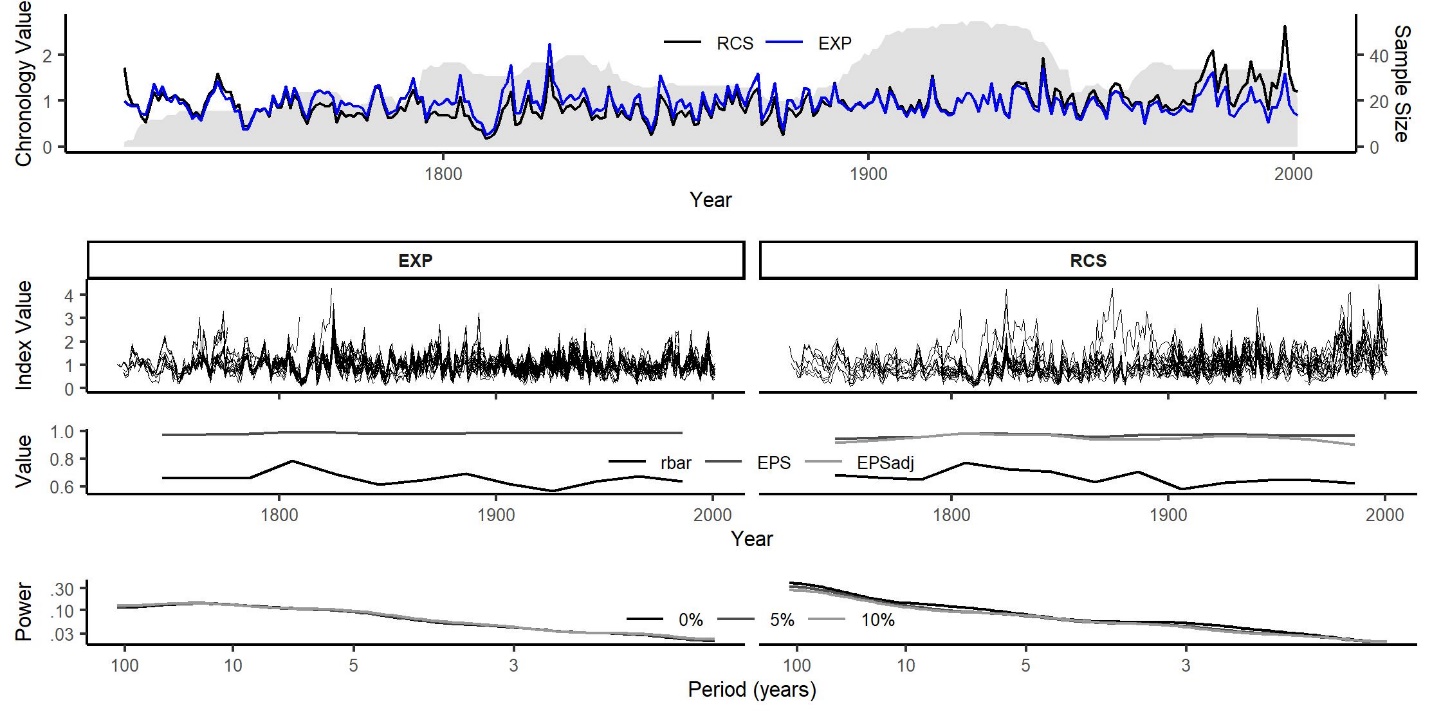


Figure A.3. Empirical comparison of detrending methods in the modern interval. **a**) Black line: RCS-detrended average ring width chronology, Blue line: Negative exponential chronology, Gray shading: Sample count **b**) Individual detrended ring width indices, negative exponential detrending on left, RCS on right **c**) rbar, EPS, and EPSadj statistics calculated over 30-year windows, 10-year overlap **d**) Smoothed periodogram with 0, 5, and 10% taper.

In Floating Chronologies 1–3, the number of years between the early and late increments estimated from the radiometric analysis generally matched the number of intervening years established by crossdating (Table 2). All discrepancies between crossdated/radiocarbon offset for paired samples are within the 1σ error of the radiocarbon ages. When accounting for the local DR, Floating Chronologies 1, 2, and 3 had start dates of 699, 1010, and 1248 CE, respectively. The remaining floating chronologies were dated to various intervals during the last three millenia, but with particulalrly strong coverage from approximately 500CE to present (Table 2; complete radiocarbon probability distributions in Supplemental Material A.S.3.). The crossdated/Marine20 radiocarbon ages (BP) of the samples used for the DR calculation are 860/606 (TND031) and 995/709 (TND045) for a regional offset of 270 years (Fig A.2 Inset; σ= 22).

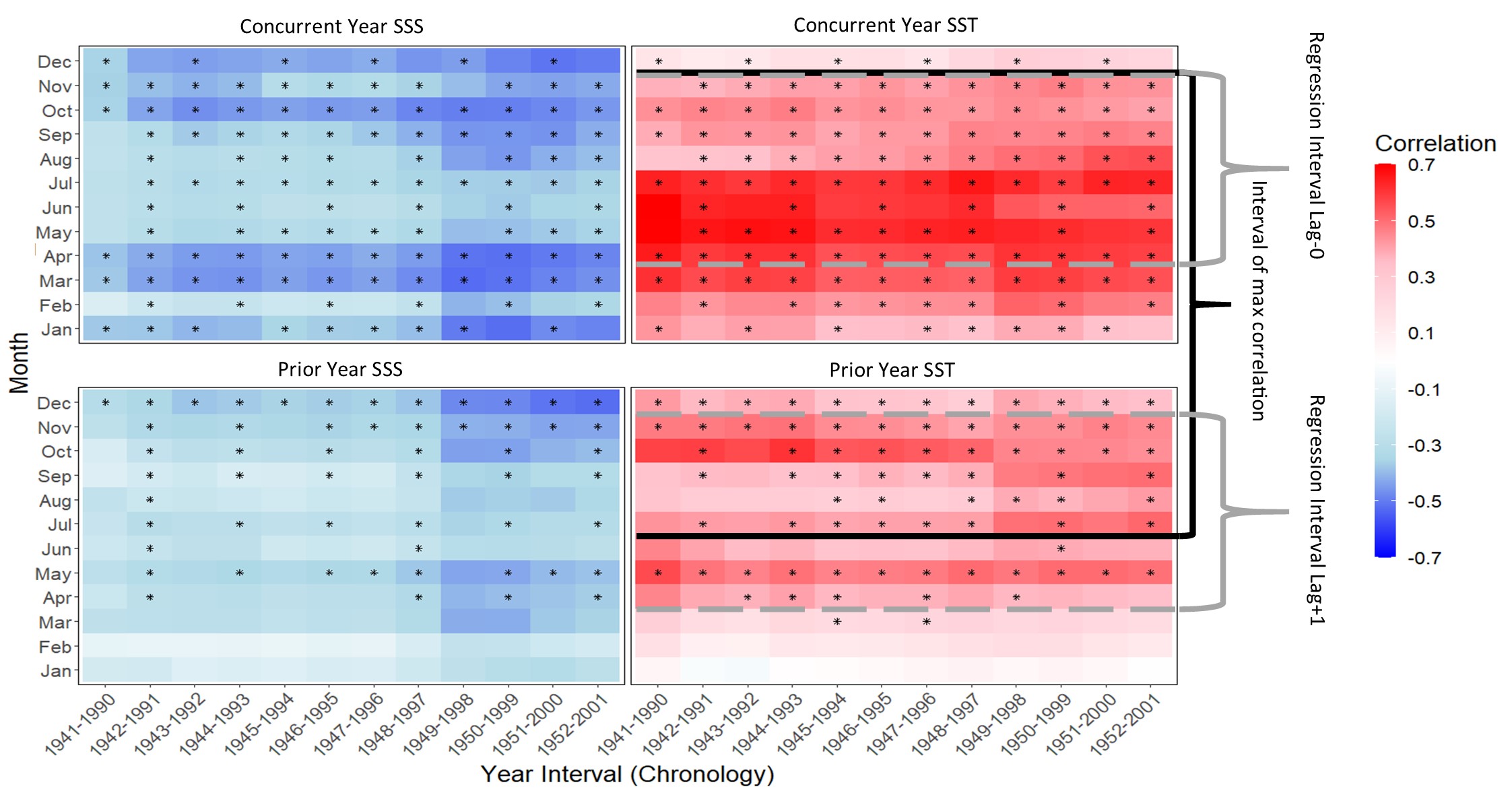
### A.4.4 Chronology construction

The RCS and EXP detrending methods (without adaptive power transform) produced chronologies with similar patterns in the year-to-year, high frequency domain (Fig A.3a). The indices resulting from RCS detrending are more variable with greater spread among individuals relative to indices resulting from EXP detrending (Fig A.3b). The EXP/RCS methods produced high average rbar (0.57/0.58, the average of correlation coefficients from all possible sample pairings), EPS (0.97/0.94), and EPSadj (0.91) statistics (Fig A.3c), though the RCS chronology contains much more spectral power at lower frequencies relative to the EXP chronology, especially in the 100-year domain (Fig A.3d; see also Supplemental Material A.S.5.). The retention of low-frequency variability is also apparent in a direct comparison of the two chronologies, especially the long-term 20th century increase that is captured by the RCS chronology but not the EXP chronology (Fig A.3a). There were no apparent age-specific differences in growth rate among cohorts of different longevities (Supplemental Material A.S.6.), suggesting that fast-growing individuals are not more likely to be short-lived relative to slow-growing individuals. This allowed detrending of all series by a single ontogenetic RCS growth curve. Note that this single curve was developed using all crossdated time series (Fig A.2), including live- and dead-collected individuals. This one RCS curve was used to produce a single chronology discontinuous in time, but such that segments were scaled relative to one another. We also produced a ‘signal-free’ RC that proved to be very similar to the standard RC. We did not use signal-free detrending in our final chronology due to the complexity of implementation with a discontinuous chronology and its similarity with the standard RC (Supplemental Material A.S.6.). Combining RCS detrending with APT altered low-frequency signals in the chronology by offsetting the positions of measurement time series relative to one another (Supplemental Material A.S.7.). We developed a method to return all series to their correct positions to confirm the cause of this artifact (Edge, 2021). Thus, we utilized the RCS method without APT for climate-growth relationships and reconstruction given its exceptional ability to retain low-frequency variability while retaining robust chronology statistics.

### A.4.5 Climate-Growth Relationships

Significant climate-growth relationships were found across nearly all months of the concurrent and prior year for both SSS and SST (Fig A.4). We found that the average of prior July – concurrent November provided the highest correlations with geoduck growth for both SSS and SST. A stepwise multiple linear regression showed that the average of prior July – concurrent November SST was the only significant predictor (p = 1.1e-10, R2 = 0.61) and that the average of prior July – concurrent November SSS did not contribute any additional explained variance (p = 0.427). The correlation with SST was significant after accounting for autocorrelation using Ebisuzaki surrogates (1997) in a 10000-iteration simulation (r = 0.78, α = 0.01; Meyers, 2014).

Figure A.4. Seasonality and stability of chronology-SST correlations. Left: Sea surface salinity (SSS) at Langara lighthouse correlated against the Tree Nob chronology in 50-year intervals. The leftmost column shows correlations between the chronology (1941-1990) and monthly SSS for the concurrent (upper panel) and prior (lower panel) years. Right: Sea surface temperature correlations with chronology. Asterisk indicates significance at α=.05, ‘exact’ simulation Monte Carlo (Percival & Constantine, 2006; Meko et al., 2011). Black box indicates SST interval of maximum correlation with concurrent year of chronology growth. The two gray boxes indicate the instrumental data interval used for the regression, signof lag with respect to SST.

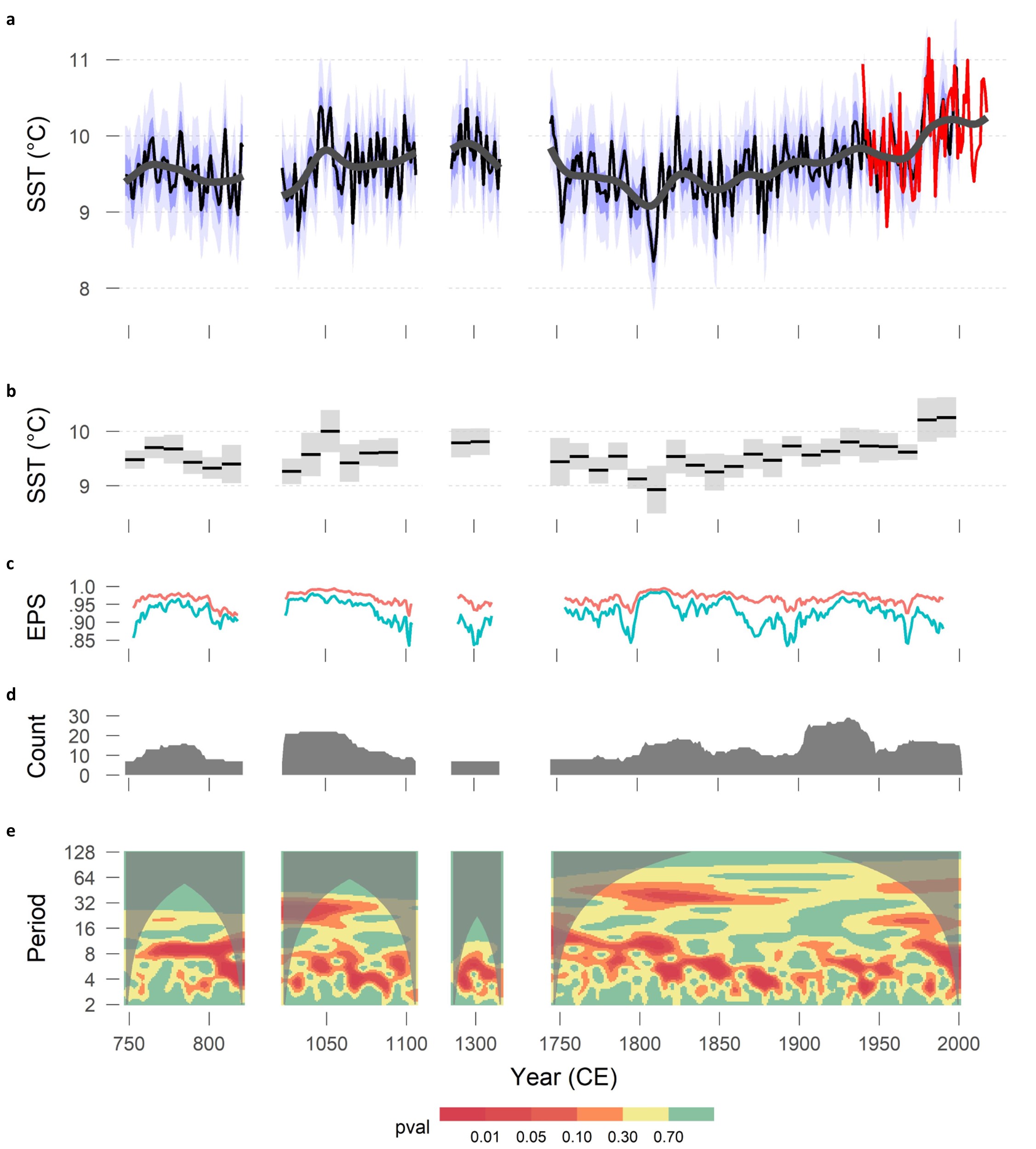


### A.4.6 SST Reconstruction

Although the average of prior July – concurrent November SST provided the highest correlations with the Tree Nob chronology, we chose a narrower window not exceeding a year in width for greatest utility and ease of interpretation. However, when the seasonal window is narrowed, the Tree Nob chronology lagged by one year also becomes significant in reconstruction models. The need for this lag is consistent with the significant correlations observed for prior year SST (Fig 4) and that a prior July – concurrent November (seventeen-month window of SST) generates the highest single-variable correlation with the Tree Nob chronology. SST averaged from April through November provides the greatest variance explained when including concurrent and lagged (lag+1) Tree Nob chronology in a model and was thus chosen as the target for the reconstruction. In a stepwise multiple linear regression, both concurrent and lagged Tree Nob chronology proved significant (p = 6.51e-06 and 0.00262 respectively), explaining 47% and 8% of variance, respectively. Regression residuals show no significant linear trend (p = 0.496), lag-1 autocorrelation (p = 0.11), or deviation from normality (p = 0.06). The split period calibration-verification resulted in significant RE and CE statistics for both the early (RE=0.54 > RE01=0.13; CE=0.49 > CE01=0.04) and late (RE=0.57 > RE01=0.14; CE=0.53 > CE01=0.05) calibration intervals. The adjusted R2 (Wherry Formula 1 from Yin & Fan, 2001) for the full period of overlap is 0.54 (n = 62). The 50- and 90-percentile prediction intervals are ±0.25 and ±0.66, respectively.

Reconstructed seasonal temperatures fluctuate at interannual to centennial scales within a narrow band of approximately 9-10 ºC until late in the 20th Century when warming consistently exceeds that long-term envelope (Fig A.5a). Warming trends are apparent in both the Floating 2 and Modern segments, though the trend in the Modern interval is longer, beginning in the mid-1800s and spanning to present, while also of greater magnitude. The Modern segment contains both the coldest and warmest reconstructed temperatures in 1810 and 1998, respectively. The 1976-77 Pacific regime shift (Miller et al., 1994) is captured by the reconstruction, and its magnitude/persistence is the most extreme in the record (Fig A.5a). The maximum and minimum temperatures of the Modern interval fall within the 90% prediction interval of some years in prior centuries, so there is some possibility that comparable extremes occurred prior to modern times. The final and penultimate 12-year median SSTs fall outside the 1σ range of all prior intervals (Fig A.5b). The EPS and EPS adjusted are high throughout, and those intervals with a minimum sample size of seven individuals are shown (Fig A.5b, c). Wavelet analysis shows significant (p < 0.1) and consistent power at periods of 4-8 years across all reconstruction intervals (Fig A.5d). Power at periods of 8- to 16-years and also 16- to 64-years are also prominent, though the significance is intermittent and the band variable, throughout the reconstruction.

Figure A.5. SST reconstruction. **a**) Seasonal (Apr-Nov) SST at Langara Lighthouse reconstructed from Tree Nob geoduck. *Red*: Instrumental SST (Apr-Nov average) as measured at Langara Lighthouse. *Black*: Reconstructed SST from Tree Nob chronology, time interval truncated to sample size ≥ 7. *Grey*: 40-year, 50% frequency cutoff cubic smoothing spline of reconstruction (and instrumental data after 2000). *Dark and light blue bands*: 50% and 90% prediction intervals based on validation interval error. The first three chronology intervals are dated from radiocarbon samples, calendar dates shown for these intervals are approximate. See Supplemental Material for age probability distributions. **b**) Black line segments show median reconstructed SST over 12-year window with no overlap. Gray shading shows 1σ range. **c)** EPS (orange) and adjusted EPS (blue) calculated over a running 20-year window. **d**) Chronology sample count (discrete shells). **e**) Wavelet of reconstructed SST. Coloring based on Monte Carlo significance (p-value).



## A.5 Discussion

In the marine realm, well-replicated, crossdated chronologies are expanding in spatial coverage and species diversity to address impacts of environmental variability on growth, environmental history, and interactions among species and trophic levels (Black et al., 2019). However, absolutely dated chronologies extending prior to 1800 CE remain less common and currently number less than a dozen (Black et al., 2019). Although six geoduck chronologies have been published, all were developed from live-collected material and thus extend, at most, into the mid-1800s CE (Strom et al., 2004; Black et al., 2009). The Tree Nob Chronology (TNC) is thus the first multicentennial, crossdated chronology of any species in the NEP and was made possible by combining live-collected individuals with overlapping dead-collected shells. Growth synchrony, or covariance, among samples was quite strong, facilitating crossdating and greatly reducing the chances of spurious matches among dead-collected individuals. Indeed, dead-collected individuals had only one very conspicuous placement in time according to cross-correlation analyses, and multiple radiometric dates independently verify crossdating accuracy. The mean correlation between each individual and the average of all others was consistently above 0.8, which is as high as is observed in other bivalve datasets including *A. islandica* or *G. glycymeris* (eg. Butler et al., 2009) and is among the highest values recorded for tree-ring data (eg. Stahle et al., 2013). RingdateR greatly improved the pace of extending the chronology, which ultimately yielded a temporal coverage of approximately 1200 of the last 3,000 years.

Exact crossdating ensures that high-frequency, year-to-year variability is fully expressed in the final chronology, but there may still be issues retaining low-frequency variability, especially at multidecadal and longer timescales. When fitting a separate function to each set of measurements to remove age-related growth declines (Cook & Kairiukstis, 1990) any trends longer than the measurement series are also removed. Therefore, low-frequency variability in the resulting chronology is limited to timescales that are less than the average series length (Cook et al., 1995). Applying RCS allowed us to avoid this ‘segment length curse’ by comparing each set of measurements to a single ‘regional curve’ of age-related growth (Mitchell, 1967). Given that the analysis involved living and dead samples, this single curve provided a universal benchmark of average growth that spanned many cohorts and environmental regimes (Briffa et al., 1992). Individuals that lived through poor (favorable) environmental regimes would have relatively slower (faster) growth compared to the curve, and this information is retained in the detrending process and incorporated into the chronology.

Although RCS, under the correct circumstances, provides a solution to the segment-length curse, it introduces several new assumptions and complications (Briffa & Melvin, 2011). Biases arise if longer-lived individuals inherently grow more slowly than short-lived counterparts, especially if those slower-growing individuals are over-represented early in the chronology (Schulman, 1954; Esper et al., 2003; Black et al., 2008). RCS also requires individuals that represent a range of environmental regimes and thus relatively evenly spread across a long timespan. Over-representation of samples in a narrow or fixed time window could result in a single climatic pattern being heavily imprinted on the regional curve, thereby distorting long-term trends (Melvin & Briffa, 2008). Finally, RCS tends to introduce relatively high levels of variance among indices and thus requires large sample counts to maintain a given level of signal strength in the chronology (Fritts, 1976; Esper et al., 2003; Melvin & Briffa, 2014).

The Tree Nob dataset appears to meet the underlying assumptions necessary for an RCS analysis. Ontogenetic growth declines are highly geometric, and geoduck lack the sustained growth pulses or suppressions common in trees following disturbance, allowing the development of a robust regional curve. Also, our analyses indicate that there are no strong relationships between growth rate and longevity. Sample counts are large, and shells were collected from a geographically focused area that should be climatologically homogenous. Moreover, samples represent intervals of time that span the majority of the past 1,500 years and thus capture the necessary diversity of environmental regimes. The RCS method has been previously used on geoduck (Strom et al., 2005), but with all live-collected individuals, which resulted in a chronology that did not substantially differ from negative exponential detrending. However, the approach has been effective at resolving low-frequency variability in *A. islandica* chronologies developed from live- and dead-collected samples (Butler et al., 2010; Butler et al., 2013). In the TNC, evidence for preservation of low frequency is apparent in the chronology periodogram (Fig 3D) as well as the steep increase over the most recent century (Fig A.3a). The spectral character of the RCS chronology is also a better approximation of the Langara SST record than the EXP chronology, particularly in the low-frequency domain (Supplemental Material A.S.5.). Poor fits of the regional curve functions tend to be most pronounced at the ends of the time series and can be amplified if low-frequency variability causes these end data points to exert leverage in the curve-fitting regression. However, signal-free detrending techniques did not make a difference in the Tree Nob RCS analysis, suggesting that there was no climate signal in the ontogenetically aligned data that could have biased the regional curve fit.

The adaptive power transformation has been commonly employed in tree rings (Esper et al., 2003; Fan et al., 2008, Panthi et al., 2017) and bivalves (Butler et al., 2010; Butler et al., 2013; Marali & Shon̈e, 2015) for the stabilization of variance. This tool is particularly useful when paired with exponential detrending, especially in situations where absolute growth rate drops to very low levels and remains in that state for prolonged periods (Cook & Peters 1997), as is common in bivalve datasets. APT has been paired with RCS in tree-ring research (Esper et al., 2003; Büntgen et al., 2005), but this detrending combination introduced low-frequency artifacts in the Tree Nob data because transformation of individual series altered their positions relative to the regional curve. This vertical offset of individual series radically altered low-frequency signals. Thus, care must be taken in implementing APT in combination with RCS when analyzing bivalve growth-increment data.

The TNC correlates very strongly with SST, especially over the seasonal window of April through November, which is consistent with earlier findings at Tree Nob (Black et al., 2008) and generally matches with the warm-season correlations of March-October in the Protection Island chronology in the Strait of Juan de Fuca (Strom et al., 2004). A key difference, however, is that in our reconstruction, the TNC lagged by one year was also a significant predictor of SST in a stepwise regression. This suggests there is ‘biological memory’ of the prior growing season in geoduck growth increments at this site, although memory at a point lower in the food chain cannot be ruled out. To further confirm this memory, we correlated the geoduck chronology against a much wider seasonal window of prior July through the concurrent November. When this window was used as a reconstruction target, only the current year of geoduck growth was significant. Thus, to reconstruct the seasonal window of current April through November, subsequent-year geoduck growth significantly increased predictive skill. Biological memory has been well documented in tree-ring records and can arise from the multi-year persistence of leaves in non-deciduous species or use of stored energy (Schulman, 1956; Matalas, 1962; Esper et al., 2015). There may be an analogous situation in bivalves in which stored glycogen, for which peak reserves occur in July and August, and is carried over from prior growing seasons (Soudant et al., 1996; Feldman et al., 2004).

The SST reconstruction from geoduck is characterized by considerable interannual to interdecadal variability and is linked to broad-scale climate patterns in the North Pacific basin (Supplemental Material A.S.8.). The most energetic of these in the El Niño Southern Oscillation (ENSO), which is teleconnected to the Aleutian Low (Schneider & Cournuelle, 2005; Newman et al., 2016) and thus is expected to be related to the geoduck SST reconstruction. The Niño3 index from the Hadley Centre Sea Ice and Sea Surface Temperature data set (Rayner et al., 2003) significantly correlates with Langara instrumental (1940-2017, r = 0.45, p = 2.8e-5) and also significantly correlates with the reconstruction over the interval 1870-2000 (r = 0.31, p = 3.0e-4). Warm years in the reconstruction correspond to major El Niño events including 1941 and 1998 (Trenberth et al., 2001). This may also explain some of the power in the 3- to 7-year window of the wavelet analysis of the SST reconstruction (Fig A.5d). Lower-frequency periodicities in the reconstruction are also consistent with the Pacific Decadal Oscillation (PDO), defined as the leading empirical orthogonal function of gridded SST north of 20° latitude in the North Pacific after the global warming trend has been removed. Over the 20th century, the PDO varies at 16–32-year frequencies (Mantua et al., 1997), and fifty-year cycles have also been noted in the instrumental record of the North Pacific (Minobe 1999). These frequencies are reflected in the Langara reconstruction over the past 1500 years, suggesting energy in these domains are characteristic of North Pacific variability over the past several centuries; however, the low-frequency power is variable in peak periodicity, intermittent in time, and less significant on average than peaks in the 4–8-year band. The Langara instrumental record significantly correlates with PDO (r=0.71, n=78, p < 1.0e-5) as does the Langara reconstruction from geoduck (r=0.49, n=108, p < 1.0e-5). The PDO index, however, is detrended to remove the global warming trend and accentuate interdecadal variability (Mantua et al., 1997) while the geoduck chronology was generated to preserve these longer-term trends. Thus, a more appropriate comparison is with the “Arc” pattern, calculated as the dominant mode of SST variability in the northeastern Pacific (east of 180° longitude) without detrending. The Arc pattern is more closely focused on the northeastern Pacific and is dynamically linked to the atmosphere (Johnstone and Mantua 2014). Correlation between the Arc pattern and the Langara reconstruction is r = 0.62 (p < 1.0e-5), underscoring the strength of this broad-scale climate pattern in the geoduck increment widths that includes the 20th century warming trend.

Multiple studies have targeted the PDO or Gulf of Alaska sea surface temperatures for reconstruction. Analysis involving annually resolved archives are dominated by tree-ring datasets (Biondi et al., 2001; D’Arrigo, 2001; Gedalof & Smith, 2001; MacDonald & Case, 2005; D’Arrigo & Wilson, 2006), though long-term instrumental precipitation records from China have also been used (Shen et al., 2006). These reconstructions, however, poorly agree with one another before 1900 (Kipfmueller et al., 2012; Henley, 2017), and thus there is considerable uncertainty about the variability of the North Pacific prior to the 20th century (Cook, 2009; Newman et al., 2016). Although the geoduck-based reconstruction contains 20th century warming trends and other reconstructions may not, especially if the target is explicitly the PDO, we compared the Langara history to six annually resolved proxy records that are relevant to northeastern Pacific SST (Biondi et al., 2001; D’Arrigo, 2001; Gedalof & Smith, 2001; MacDonald & Case, 2005; D’Arrigo & Wilson, 2006; Shen et al., 2006). Among these only the Gedalof & Smith (2001) reconstruction significantly and stably correlates (r=0.39, n=250, p < 1.0e-5), with comparable relationships pre- and post- 1900 with the Langara reconstruction. The combination of the Langara reconstruction with that from Gedalof and Smith by simple average produces a series that correlates much more strongly with the PDO index (r=0.66, n=82, p < 1.0e-5) than either series individually. This result is consistent with earlier findings that a composite of multiple geoduck and SST-sensitive tree-ring chronologies could explain greater quantities of SST variance when combined (Black et al., 2009). The agreement between the Langara reconstruction and that of Gedalof and Smith (2001) suggests that trees and bivalves likely share common patterns, and that with careful selection of chronologies, could be combined to generate much longer multi-proxy histories. Previous SST-related reconstruction attempts may suffer from some combination of instability in the dominant modes of climate variability (Gedalof et al., 2002; Bond et al., 2003; Di Lorenzo et al., 2008), spatial limitations of proxy networks, inconsistencies of teleconnections (Gallant et al., 2013, Batehup et al., 2015), idiosyncrasies of reconstruction methods, and incompatibility of frequency characteristics (Franke et al., 2013). The geoduck chronology could provide a much longer, annually resolved, marine-based estimate of SST variability with which to screen tree-ring chronologies for stability over multicentennial timescales, which can be difficult to determine given the brevity of the instrumental record, the relatively low number of multidecadal cycles, and the possibility of changing climate-growth relationships with anthropogenic warming (Frank et al., 2007; Wilson et al., 2007; Esper et al., 2009).

In addition to broad-scale teleconnected climate processes, the geoduck-based reconstruction also appears to capture cooling associated with minima in solar activity and major volcanic eruptions. The coldest decade of the reconstruction centered around 1810 coincides with the Dalton Minimum as well as the coldest decade of the last 500 years reported in a review of Northern Hemisphere (NH) temperature reconstructions (Cole-Dai et al., 2009). Volcanism is also a likely contributor to the cold, and thus the slow growth that geoduck exhibit during this period (Wagner & Zorita, 2005). The two largest NH volcanic eruptions since 1735 occurred in 1809 (Unknown) and 1815 (Tambora; Gao et al., 2008). At least one of these appears to have a signature in the geoduck-based SST reconstruction, indicated by the three coldest years on record in 1809-1811. The year 1816, however, is relatively warm in the Langara reconstruction and was likely associated with an El Niño event (Li et al., 2013), though the reconstruction returns to conspicuously cool conditions in 1817 and 1818. The 1982 and 1991 eruptions of El Chicón and Pinatubo also foretell brief, sharp cooling events in the warmest reconstruction interval. Thus, volcanic activity may have an influence on the regional SST but may also be overridden by other climatic processes (Adams et al., 2003).

One of the most conspicuous features of the reconstruction is the warming trend that began around 1850 and continues through the end of the record. Indeed, the decades since the 1976 shift to a warm regime (Nitta & Yamada, 1989; Trenberth 1990; Latif & Barnett; 1996) are the warmest of the 50 decades in the reconstruction (Fig A.5b), consistent with NH surface temperature and global SST reconstructions (Mann et al., 2009). Although the reconstruction ends in 2000 and is specific to a nearshore location, it does underscore the long-term warming that has occurred in the northeastern Pacific over the past century or longer and thereby provides context for the recent and apparently unprecedented marine heatwaves known colloquially as “The Blob” (Bond et al., 2015, DiLorenzo & Mantua 2016). The first of these warm-water events occurred between 2014–2015 followed by another in 2019. These heatwaves were characterized by significant reductions in krill and forage fish, species redistributions, harmful algal blooms, seabird mortality, broad-scale marine mammal strandings, and the closure of multiple fisheries (Kintisch, 2015; Gewin, 2015; Amaya et al., 2020; Cornwall, 2019). Peak warming occurred primarily offshore and is thus not reflected well in the Langara instrumental record, but the warming trends that culminated in these apparently unprecedented heatwaves are clear in the geoduck-based reconstruction.

Ultimately, the TNC continuously extends between 1725 and 2008 while also providing multiple windows that span 58% of the past 1500 years. As such, this is the longest annually resolved marine archive currently available in the North Pacific, and with a high probability of being substantially extended. Indeed, with additional collections, there is the distinct possibility that gaps in the chronology could be filled to yield a continuous millennial to 1,500-year record. From the limited number of radiocarbon dates of randomly selected, undated shells, the minimum preservation is 3,000 years, but may be much greater considering that *A. islandica* that died more than 10,000 years BP have been retrieved from the North Atlantic seafloor (Butler et al., 2020). Given the abundance of geoduck along the NEP coast and typical average lifespan, chronologies of similar length could be developed from the Strait of Juan de Fuca to Kodiak, Alaska. To better constrain past climate variability, the exactly dated framework of the TNC could also be sampled for isotopic or possibly microchemical analysis. Oxygen isotope ratios have proven to be a robust indicator of past hydrography (i.e, combined effects of salinity and temperature changes), especially for *A. islandica* and may also be useful for geoduck (Wanamaker et al., 2008; Reynolds et al., 2013; Reynolds et al., 2016).

On broader spatial scales, NEP SST modulates global temperatures via fluxes of heat storage and release, especially on decadal timeframes (Meehl et al., 2011; Meehl et al., 2013; England et al., 2014; Thompson et al., 2015; Yin et al., 2018). Thus, better constraining the past may afford some predictive skill for the future. For example, uncertainty in short-term (<30 years) global warming projections is largely due to variability in global ocean heat content (Smith et al., 2007; Hawkins & Sutton 2009) while decadal climate prediction skill for North America, and associated climate adaptation planning, is also contingent on the predictability of NEP SST regime (Meehl et al., 2009). Moreover, a high-resolution reconstruction of basin-wide NEP SST, possibly generated through multi-proxy approaches, could also be used to test hypotheses regarding the origins of NEP SST decadal-scale variability. NEP SST regimes have been interpreted by some as oscillatory, with forcing mechanisms related to Rossby wave propagation time or gyre circulation (Minobe, 1999; Chao et al., 2000; Mantua & Hare, 2002). However, an increasing body of evidence suggests that NEP SST is better described as random noise resulting from higher-frequency signals of tropical (e.g. the El Niño Southern Oscillation) and extratropical (e.g. variability in the Aleutian Low) origin being buffered in the water column of the NEP (Pierce, 2001; Newman et al., 2003; Schneider & Cornuelle, 2005; Mestas-Nunez & Miller, 2006; Newman et al., 2016). A basin-scale NEP reconstruction could be compared to independent annually resolved ENSO reconstructions to test these relationships over multicentennial timescales. Also, residual variance could be compared to eastern Asian climate or that of the western Pacific to better identify extratropical influences on NEP SST. As such, geoduck represent a key archive that could be greatly expanded in temporal and spatial scope to address basic questions of past NEP variability and its drivers.

## A.6 Acknowledgments, Samples, and Data

Shell collection was performed by Trident Biological Consulting Ltd., and we thank Mike Atkins for his thoughtful and professional services that provided this unique collection. We also thank Hank Carson at the Washington Department of Fish and Wildlife for his support in the sample collection process. This work is funded by the National Science Foundation (AGS Award Number: 1855628 to BAB; Award Number: 1602751 to ADW; Award Number: 1602633 to DG). The growth increment mesaurements and related ‘pith offsets’, master chronology, and reconstruction data will be archived at NOAA Paleoclimatology (https://www.ncdc.noaa.gov/paleo/study/33312). Data are temporarily stored at: https://drive.google.com/drive/folders/1wyeuAe76vlL6bB6aibPuBD6yzE2FSocC?usp=sharing

## A.7 References

Adams, J. B., Mann, M. E., & Ammann, C. M. (2003). Proxy evidence for an El Nino-like response to volcanic forcing. *Nature*, *426*(6964), 274-278.

Amaya, D. J., Miller, A. J., Xie, S.-P., & Kosaka, Y. (2020). Physical drivers of the summer 2019 North Pacific marine heatwave. *Nature communications*, *11*(1), 1-9.

Batehup, R., McGregor, S., & Gallant, A. (2015). The influence of non-stationary teleconnections on palaeoclimate reconstructions of ENSO variance using a pseudoproxy framework. *Climate of the Past*, *11*(12), 1733-1749.

Berkelhammer, M. (2019). Synchronous modes of terrestrial and marine productivity in the North Pacific. *Frontiers in Earth Science*, *7*, 73.

Biondi, F., Gershunov, A., & Cayan, D. R. (2001). North Pacific decadal climate variability since 1661. *Journal of Climate*, *14*(1), 5-10.

Black, B. A., Andersson, C., Butler, P. G., Carroll, M. L., DeLong, K. L., Reynolds, D. J., Schöne, B. R., Scourse, J., van der Sleen, P., & Wanamaker, A. D. (2019). The revolution of crossdating in marine palaeoecology and palaeoclimatology. *Biology letters*, *15*(1), 20180665.

Black, B. A., Copenheaver, C. A., Frank, D. C., Stuckey, M. J., & Kormanyos, R. E. (2009). Multi-proxy reconstructions of northeastern Pacific sea surface temperature data from trees and Pacific geoduck. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *278*(1-4), 40-47.

Black, B. A., Gillespie, D. C., MacLellan, S. E., & Hand, C. M. (2008). Establishing highly accurate production-age data using the tree-ring technique of crossdating: a case study for Pacific geoduck (Panopea abrupta). *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(12), 2572-2578.

Bond, N., Overland, J., Spillane, M., & Stabeno, P. (2003). Recent shifts in the state of the North Pacific. *Geophysical Research Letters*, *30*(23).

Bond, N. A., Cronin, M. F., Freeland, H., & Mantua, N. (2015). Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophysical Research Letters*, *42*(9), 3414-3420.

Briffa, K. R., Jones, P. D., Bartholin, T. S., Eckstein, D., Schweingruber, F. H., Karlen, W., Zetterberg, P., & Eronen, M. (1992). Fennoscandian summers from AD 500: temperature changes on short and long timescales. *Climate Dynamics*, *7*(3), 111-119.

Briffa, K. R., Jones, P. D., Schweingruber, F. H., Karlén, W., & Shiyatov, S. G. (1996). Tree-ring variables as proxy-climate indicators: problems with low-frequency signals. In *Climatic variations and forcing mechanisms of the last 2000 years* (pp. 9-41). Springer.

Briffa, K. R., & Melvin, T. M. (2011). A closer look at regional curve standardization of tree-ring records: justification of the need, a warning of some pitfalls, and suggested improvements in its application. In *Dendroclimatology* (pp. 113-145). Springer.

Bunn, A. G. (2008). A dendrochronology program library in R (dplR). *Dendrochronologia*, *26*(2), 115-124.

Bunn, A. G., Jansma, E., Korpela, M., Westfall, R. D., & Baldwin, J. (2013). Using simulations and data to evaluate mean sensitivity (ζ) as a useful statistic in dendrochronology. *Dendrochronologia*, *31*(3), 250-254.

Büntgen, U., Esper, J., Frank, D. C., Nicolussi, K., & Schmidhalter, M. (2005). A 1052-year tree-ring proxy for Alpine summer temperatures. *Climate Dynamics*, *25*(2), 141-153.

Bureau, D. (2002). Age, size structure and growth parameters of geoducks (Panopea abrupta, Conrad 1849) from 34 locations in British Columbia sampled between 1993 and 2000. *Can Tech Rep Fish Aquat Sci*, *2413*, 1-84.

Butler, P. G., Fraser, N. M., Scourse, J. D., Richardson, C. A., Bryant, C., & Heinemeier, J. (2020). Is there a reliable taphonomic clock in the temperate North Atlantic? An example from a North Sea population of the mollusc Arctica islandica. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *560*, 109975.

Butler, P. G., Richardson, C. A., Scourse, J. D., Wanamaker Jr, A. D., Shammon, T. M., & Bennell, J. D. (2010). Marine climate in the Irish Sea: analysis of a 489-year marine master chronology derived from growth increments in the shell of the clam Arctica islandica. *Quaternary Science Reviews*, *29*(13-14), 1614-1632.

Butler, P. G., Scourse, J. D., Richardson, C. A., Wanamaker Jr, A. D., Bryant, C. L., & Bennell, J. D. (2009). Continuous marine radiocarbon reservoir calibration and the 13C Suess effect in the Irish Sea: Results from the first multi-centennial shell-based marine master chronology. *Earth and Planetary Science Letters*, *279*(3-4), 230-241.

Butler, P. G., Wanamaker Jr, A. D., Scourse, J. D., Richardson, C. A., & Reynolds, D. J. (2013). Variability of marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on growth increments in the bivalve Arctica islandica. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *373*, 141-151.

Chao, Y., Ghil, M., & McWilliams, J. C. (2000). Pacific interdecadal variability in this century's sea surface temperatures. Geophysical Research Letters, 27(15), 2261-2264.

Chavez, F. P., Ryan, J., Lluch-Cota, S. E., & Ñiquen, M. (2003). From anchovies to sardines and back: multidecadal change in the Pacific Ocean. *Science*, *299*(5604), 217-221.

Clark, W. G., Hare, S. R., Parma, A. M., Sullivan, P. J., & Trumble, R. J. (1999). Decadal changes in growth and recruitment of Pacific halibut (Hippoglossus stenolepis). *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(2), 242-252.

Cole‐Dai, J., Ferris, D., Lanciki, A., Savarino, J., Baroni, M., & Thiemens, M. H. (2009). Cold decade (AD 1810–1819) caused by Tambora (1815) and another (1809) stratospheric volcanic eruption. *Geophysical Research Letters*, *36*(22).

Cook, E. (2009). North Pacific climate variability over the past millennium from long tree-ring records. AGU Fall Meeting Abstracts.

Cook, E. R., Briffa, K. R., & Jones, P. D. (1994). Spatial regression methods in dendroclimatology: a review and comparison of two techniques. *International Journal of Climatology*, *14*(4), 379-402.

Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A., & Funkhouser, G. (1995). The'segment length curse'in long tree-ring chronology development for palaeoclimatic studies. *The Holocene*, *5*(2), 229-237.

Cook, E. R., & Kairiukstis, L. A. (1990). *Methods of dendrochronology: applications in the environmental sciences*. Springer Science & Business Media.

Cook, E. R., & Peters, K. (1997). Calculating unbiased tree-ring indices for the study of climatic and environmental change. *The Holocene*, *7*(3), 361-370.

Cornwall, W. (2019). A new ‘Blob’ menaces Pacific ecosystems. Science, 365(6459), 1233-1233.

D'Arrigo, R., & Wilson, R. (2006). On the Asian expression of the PDO. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *26*(12), 1607-1617.

D'Arrigo, R., Villalba, R., & Wiles, G. (2001). Tree-ring estimates of Pacific decadal climate variability. Climate Dynamics, 18(3-4), 219-224.

de Vries, H. (1958). Variation in concentration of radiocarbon with time and location on earth. *Proc. Koninkl. Nederl. Akad. Wetenschappen, B*, *61*, 1-9.

Di Lorenzo, E., & Mantua, N. (2016). Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nature Climate Change*, *6*(11), 1042-1047.

Di Lorenzo, E., Schneider, N., Cobb, K. M., Franks, P., Chhak, K., Miller, A. J., McWilliams, J. C., Bograd, S. J., Arango, H., & Curchitser, E. (2008). North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophysical Research Letters*, *35*(8).

Douglass, A. E. (1941). Crossdating in dendrochronology. *Journal of Forestry*, *39*(10), 825-831.

Ebisuzaki, W. (1997). A method to estimate the statistical significance of a correlation when the data are serially correlated. *Journal of Climate*, *10*(9), 2147-2153.

England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., Gupta, A. S., McPhaden, M. J., Purich, A., & Santoso, A. (2014). Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. Nature Climate Change, 4(3), 222-227.

Esper, J., Cook, E. R., Krusic, P. J., Peters, K., & Schweingruber, F. H. (2003). Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies.

Esper, J., & Frank, D. (2009). Divergence pitfalls in tree-ring research. *Climatic Change*, *94*(3), 261-266.

Esper, J., Schneider, L., Smerdon, J. E., Schöne, B. R., & Büntgen, U. (2015). Signals and memory in tree-ring width and density data. *Dendrochronologia*, *35*, 62-70.

Fan, Z. X., Bräuning, A., & Cao, K. F. (2008). Tree‐ring based drought reconstruction in the central Hengduan Mountains region (China) since AD 1655. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *28*(14), 1879-1887.

Feldman, K., Vadopalas, B., Armstrong, D., Friedman, C., Hilborn, R., Naish, K., Orensanz, J., Valero, J., Ruesink, J., & Suhrbier, A. (2004). Comprehensive literature review and synopsis of issues relating to geoduck (Panopea abrupta) ecology and aquaculture production. *Olympia, WA: Washington State Department of Natural Resources*.

Ferguson, C. W., & Graybill, D. (1983). Dendrochronology of bristlecone pine: a progress report. *Radiocarbon*, *25*(2), 287-288.

Frank, D., Büntgen, U., Böhm, R., Maugeri, M., & Esper, J. (2007). Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. *Quaternary Science Reviews*, *26*(25-28), 3298-3310.

Franke, J., Frank, D., Raible, C. C., Esper, J., & Brönnimann, S. (2013). Spectral biases in tree-ring climate proxies. *Nature Climate Change*, *3*(4), 360-364.

Fritts, H. C. (1971). Dendroclimatology and dendroecology. *Quaternary Research*, *1*(4), 419-449.

Fritts, H. C. (1976). Tree rings and climate. Academic Press. New York.

Gallant, A. J., Phipps, S. J., Karoly, D. J., Mullan, A. B., & Lorrey, A. M. (2013). Nonstationary Australasian teleconnections and implications for paleoclimate reconstructions. *Journal of Climate*, *26*(22), 8827-8849.

Gao, C., Robock, A., & Ammann, C. (2008). Volcanic forcing of climate over the past 1500 years: An improved ice core‐based index for climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D23).

Gauss, C. (1821). Theory of the combination of observations which leads to the smallest errors. *Gauss Werke*, *4*, 1-93.

Gedalof, Z. e., Mantua, N. J., & Peterson, D. L. (2002). A multi‐century perspective of variability in the Pacific Decadal Oscillation: New insights from tree rings and coral. *Geophysical Research Letters*, *29*(24), 57-51-57-54.

Gedalof, Z. e., & Smith, D. J. (2001). Interdecadal climate variability and regime‐scale shifts in Pacific North America. *Geophysical Research Letters*, *28*(8), 1515-1518.

Gewin, V. (2015). North Pacific ‘blob’stirs up fisheries management. *Nature News*, *524*(7566), 396.

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. Bulletin of the American Meteorological Society, 90(8), 1095-1108.

Richardson, C. A. (2001). Molluscs as archives of environmental change. *Oceanogr. Mar. Biol. Annu. Rev*, *39*, 103-164.

Glock, W. S., & Pearson, G. A. (1937). *Principles and methods of tree-ring analysis*. Carnegie institution of Washington Washington, DC.

Hare, S. R., Mantua, N. J., & Francis, R. C. (1999). Inverse production regimes: Alaska and west coast Pacific salmon. *Fisheries*, *24*(1), 6-14.

Heaton, T. J., Köhler, P., Butzin, M., Bard, E., Reimer, R. W., Austin, W. E., Ramsey, C. B., Grootes, P. M., Hughen, K. A., & Kromer, B. (2020). Marine20—the marine radiocarbon age calibration curve (0–55,000 cal BP). *Radiocarbon*, *62*(4), 779-820.

Henley, B. J. (2017). Pacific decadal climate variability: Indices, patterns and tropical-extratropical interactions. *Global and Planetary Change*, *155*, 42-55.

Hiebenthal, C., Philipp, E., Eisenhauer, A., & Wahl, M. (2012). Interactive effects of temperature and salinity on shell formation and general condition in Baltic Sea Mytilus edulis and Arctica islandica. *Aquatic Biology*, *14*(3), 289-298.

Holmes, R. L. (1983). Computer-assisted quality control in tree-ring dating and measurement.

Hyndman, R. J., & Khandakar, Y. (2007). *Automatic time series for forecasting: the forecast package for R*. Monash University, Department of Econometrics and Business Statistics ….

Johnstone, J. A., & Mantua, N. J. (2014). Atmospheric controls on northeast Pacific temperature variability and change, 1900–2012. *Proceedings of the National Academy of Sciences*, *111*(40), 14360-14365.

Jones, P. D., Briffa, K. R., Osborn, T., Lough, J. M., van Ommen, T. D., Vinther, B. M., Luterbacher, J., Wahl, E., Zwiers, F., & Mann, M. E. (2009). High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *The Holocene*, *19*(1), 3-49.

Kintisch, E. (2015). ‘The Blob’ invades Pacific, flummoxing climate experts. Science, 348(6230), 17-18.

Kipfmueller, K. F., Larson, E. R., & St. George, S. (2012). Does proxy uncertainty affect the relations inferred between the Pacific Decadal Oscillation and wildfire activity in the western United States? *Geophysical Research Letters*, *39*(4).

Kitzberger, T., Brown, P. M., Heyerdahl, E. K., Swetnam, T. W., & Veblen, T. T. (2007). Contingent Pacific–Atlantic Ocean influence on multicentury wildfire synchrony over western North America. *Proceedings of the National Academy of Sciences*, *104*(2), 543-548.

Latif, M., & Barnett, T. P. (1996). Decadal climate variability over the North Pacific and North America: Dynamics and predictability. *Journal of Climate*, *9*(10), 2407-2423.

Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.

Li, J., Xie, S.-P., Cook, E. R., Morales, M. S., Christie, D. A., Johnson, N. C., Chen, F., D’Arrigo, R., Fowler, A. M., & Gou, X. (2013). El Niño modulations over the past seven centuries. *Nature Climate Change*, *3*(9), 822-826.

Lower‐Spies, E. E., Whitney, N. M., Wanamaker, A. D., Griffin, S. M., Introne, D. S., & Kreutz, K. J. (2020). A 250‐Year, Decadally Resolved, Radiocarbon Time History in the Gulf of Maine Reveals a Hydrographic Regime Shift at the End of the Little Ice Age. *Journal of Geophysical Research: Oceans*, *125*(9), e2020JC016579.

MacDonald, G. M., & Case, R. A. (2005). Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters*, *32*(8).

Macias-Fauria, M., Grinsted, A., Helama, S., & Holopainen, J. (2012). Persistence matters: Estimation of the statistical significance of paleoclimatic reconstruction statistics from autocorrelated time series. *Dendrochronologia*, *30*(2), 179-187.

Mann, M. E., & Jones, P. D. (2003). Global surface temperatures over the past two millennia. *Geophysical Research Letters*, *30*(15).

Mann, M. E., & Rutherford, S. (2002). Climate reconstruction using ‘Pseudoproxies’. *Geophysical Research Letters*, *29*(10), 139-131-139-134.

Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., & Ni, F. (2009). Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science*, *326*(5957), 1256-1260.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, *78*(6), 1069-1080.

Mantua, N. J., & Hare, S. R. (2002). The Pacific decadal oscillation. Journal of oceanography, 58(1), 35-44.

Marali, S., & Schöne, B. R. (2015). Oceanographic control on shell growth of Arctica islandica (Bivalvia) in surface waters of Northeast Iceland—Implications for paleoclimate reconstructions. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *420*, 138-149.

Matalas, N. C. (1962). Statistical properties of tree ring data. *Hydrological Sciences Journal*, *7*(2), 39-47.

Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A., & Trenberth, K. E. (2011). Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. Nature Climate Change, 1(7), 360-364.

Meehl, G. A., Hu, A., Arblaster, J. M., Fasullo, J., & Trenberth, K. E. (2013). Externally forced and internally generated decadal climate variability associated with the Interdecadal Pacific Oscillation. Journal of Climate, 26(18), 7298-7310.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., & Hawkins, E. (2009). Decadal prediction: Can it be skillful? Bulletin of the American Meteorological Society, 90(10), 1467-1486.

Meko, D. M., Touchan, R., & Anchukaitis, K. J. (2011). Seascorr: A MATLAB program for identifying the seasonal climate signal in an annual tree-ring time series. *Computers & Geosciences*, *37*(9), 1234-1241.

Melvin, T. M., & Briffa, K. R. (2008). A “signal-free” approach to dendroclimatic standardisation. *Dendrochronologia*, *26*(2), 71-86.

Melvin, T. M., & Briffa, K. R. (2014). CRUST: Software for the implementation of Regional Chronology Standardisation: Part 2. Further RCS options and recommendations. *Dendrochronologia*, *32*(4), 343-356.

Melvin, T. M., Briffa, K. R., Nicolussi, K., & Grabner, M. (2007). Time-varying-response smoothing. Dendrochronologia, 25(1), 65-69.

Menesguen, A., & Dreves, L. (1987). Sea-temperature anomalies and population dynamics variations: effects on growth and density of three bivalves. *Mar. Ecol. Prog. Ser*, *36*, 11-21.

Mestas-Nunez, A. M., & Miller, A. J. (2006). Interdecadal variability and climate change in the eastern tropical Pacific: A review. Progress in Oceanography, 69(2-4), 267-284.

Meyers, S.R. (2014). Astrochron: An R Package for Astrochronology. https://cran.r-project.org/package=astrochron.

Miller, A. J., Cayan, D. R., Barnett, T. P., Graham, N. E., & Oberhuber, J. M. (1994). The 1976-77 climate shift of the Pacific Ocean. Oceanography, 7(1), 21-26.

Minobe, S. (1999). Resonance in bidecadal and pentadecadal climate oscillations over the North Pacific: Role in climatic regime shifts. *Geophysical Research Letters*, *26*(7), 855-858.

Mitchell, V. L. (1967). *An investigation of certain aspects of tree growth rates in relation to climate in the central Canadian boreal forest*.

Mote, P. W. (2006). Climate-driven variability and trends in mountain snowpack in western North America. *Journal of Climate*, *19*(23), 6209-6220.

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., Mantua, N. J., Miller, A. J., Minobe, S., & Nakamura, H. (2016). The Pacific decadal oscillation, revisited. *Journal of Climate*, *29*(12), 4399-4427.

Newman, M., Compo, G. P., & Alexander, M. A. (2003). ENSO-forced variability of the Pacific decadal oscillation. Journal of Climate, 16(23), 3853-3857.

Nitta, T., & Yamada, S. (1989). Recent warming of tropical sea surface temperature and its relationship to the Northern Hemisphere circulation. *Journal of the Meteorological Society of Japan. Ser. II*, *67*(3), 375-383.

Noakes, D., & Campbell, A. (1992). Use of geoduck clams to indicate changes in the marine environment of Ladysmith Harbour, British Columbia. *Environmetrics*, *3*(1), 81-97.

Panthi, S., Bräuning, A., Zhou, Z.-K., & Fan, Z.-X. (2017). Tree rings reveal recent intensified spring drought in the central Himalaya, Nepal. *Global and Planetary Change*, *157*, 26-34.

Percival, D. B., & Constantine, W. L. (2006). Exact simulation of Gaussian time series from nonparametric spectral estimates with application to bootstrapping. *Statistics and Computing*, *16*(1), 25-35.

Pierce, D. W. (2001). Distinguishing coupled ocean–atmosphere interactions from background noise in the North Pacific. Progress in Oceanography, 49(1-4), 331-352.

Pilcher, J. R., Baillie, M. G., Schmidt, B., & Becker, B. (1984). A 7,272-year tree-ring chronology for western Europe. *Nature*, *312*(5990), 150-152.

Pourmozaffar, S., Tamadoni Jahromi, S., Rameshi, H., Sadeghi, A., Bagheri, T., Behzadi, S., Gozari, M., Zahedi, M. R., & Abrari Lazarjani, S. (2020). The role of salinity in physiological responses of bivalves. *Reviews in Aquaculture*, *12*(3), 1548-1566.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramsey, C. B. (1995). Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon*, *37*(2), 425-430.

Ramsey, C. B., van der Plicht, J., & Weninger, B. (2001). ‘Wiggle matching’radiocarbon dates. *Radiocarbon*, *43*(2A), 381-389.

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., Kent, E., & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(D14).

Reynolds, D. J., Butler, P., Williams, S., Scourse, J., Richardson, C., Wanamaker Jr, A., Austin, W., Cage, A., & Sayer, M. (2013). A multiproxy reconstruction of Hebridean (NW Scotland) spring sea surface temperatures between AD 1805 and 2010. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *386*, 275-285.

Reynolds, D. J., Edge, D. C., & Black, B. A. (2021). RingdateR: A statistical and graphical tool for crossdating. *Dendrochronologia*, *65*, 125797.

Reynolds, D. J., Scourse, J., Halloran, P., Nederbragt, A., Wanamaker, A. D., Butler, P., Richardson, C., Heinemeier, J., Eiriksson, J., & Knudsen, K. (2016). Annually resolved North Atlantic marine climate over the last millennium. *Nature communications*, *7*(1), 1-11.

Reynolds, D. J., Richardson, C., Scourse, J., Butler, P., Hollyman, P., Roman-Gonzalez, A., & Hall, I. R. (2017). Reconstructing North Atlantic marine climate variability using an absolutely-dated sclerochronological network. Palaeogeography, Palaeoclimatology, Palaeoecology, 465, 333-34.

Roesch, A. and Schmidbauer, H. (2018). WaveletComp: Computational Wavelet Analysis. R package version 1.1. https://CRAN.R-project.org/package=WaveletComp.

Schneider, N., & Cornuelle, B. D. (2005). The forcing of the Pacific decadal oscillation. *Journal of Climate*, *18*(21), 4355-4373.

Schulman, E. (1954). Longevity under adversity in conifers. *Science*, *119*(3091), 396-399.

Schulman, E. (1956). Dendroclimatic changes in semiarid America. University of Arizona Press.

Scourse, J., Richardson, C., Forsythe, G., Harris, I., Heinemeier, J., Fraser, N., Briffa, K., & Jones, P. (2006). First cross-matched floating chronology from the marine fossil record: data from growth lines of the long-lived bivalve mollusc Arctica islandica. *The Holocene*, *16*(7), 967-974.

Shen, C., Wang, W. C., Gong, W., & Hao, Z. (2006). A Pacific Decadal Oscillation record since 1470 AD reconstructed from proxy data of summer rainfall over eastern China. *Geophysical Research Letters*, *33*(3).

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007). Improved surface temperature prediction for the coming decade from a global climate model. Science, 317(5839), 796-799.

Soudant, P., Marty, Y., Moal, J., Robert, R., Quéré, C., Le Coz, J. R., & Samain, J. F. (1996). Effect of food fatty acid and sterol quality on Pecten maximus gonad composition and reproduction process. *Aquaculture*, *143*(3-4), 361-378.

Stahle, D., Griffin, R., Meko, D., Therrell, M., Edmondson, J., Cleaveland, M., Stahle, L., Burnette, D., Abatzoglou, J., & Redmond, K. (2013). The ancient blue oak woodlands of California: Longevity and hydroclimatic history. *Earth Interactions*, *17*(12), 1-23.

Stokes, M. A., & Smiley, T. L. (1968). Introduction to tree-ring dating. University of Chicago.

Strackee, J., & Jansma, E. (1992). The statistical properties of mean sensitivity–a reappraisal. *Dendrochronologia*, *10*, 121-135.

Strom, A., Francis, R. C., Mantua, N. J., Miles, E. L., & Peterson, D. L. (2004). North Pacific climate recorded in growth rings of geoduck clams: a new tool for paleoenvironmental reconstruction. *Geophysical Research Letters*, *31*(6).

Strom, A., Francis, R. C., Mantua, N. J., Miles, E. L., & Peterson, D. L. (2005). Preserving low-frequency climate signals in growth records of geoduck clams (Panopea abrupta). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *228*(1-2), 167-178.

Team, R. C. (2013). R: A language and environment for statistical computing.

Thompson, D. M., Cole, J. E., Shen, G. T., Tudhope, A. W., & Meehl, G. A. (2015). Early twentieth-century warming linked to tropical Pacific wind strength. Nature Geoscience, 8(2), 117-121.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61-78.

Trenberth, K. E. (1990). Recent observed interdecadal climate changes in the Northern Hemisphere. *Bulletin of the American Meteorological Society*, *71*(7), 988-993.

Trenberth, K. E., & Stepaniak, D. P. (2001). Indices of el Niño evolution. *Journal of Climate*, *14*(8), 1697-1701.

Wagner, S., & Zorita, E. (2005). The influence of volcanic, solar and CO 2 forcing on the temperatures in the Dalton Minimum (1790–1830): a model study. *Climate Dynamics*, *25*(2), 205-218.

Wanamaker, A. D., Butler, P. G., Scourse, J. D., Heinemeier, J., Eiríksson, J., Knudsen, K. L., & Richardson, C. A. (2012). Surface changes in the North Atlantic meridional overturning circulation during the last millennium. *Nature communications*, *3*(1), 1-7.

Wanamaker, A. D., Griffin, S. M., Ummenhofer, C. C., Whitney, N. M., Black, B., Parfitt, R., Lower‐Spies, E. E., Introne, D., & Kreutz, K. J. (2019). Pacific climate influences on ocean conditions and extreme shell growth events in the Northwestern Atlantic (Gulf of Maine). *Climate Dynamics*, *52*(11), 6339-6356.

Wanamaker, A. D., Heinemeier, J., Scourse, J. D., Richardson, C. A., Butler, P. G., Eiríksson, J., & Knudsen, K. L. (2008). Very long-lived mollusks confirm 17th century AD tephra-based radiocarbon reservoir ages for North Icelandic shelf waters. Radiocarbon, 50(3), 399-412.

Wigley, T. M., Briffa, K. R., & Jones, P. D. (1984). On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. Journal of Applied Meteorology and Climatology, 23(2), 201-213.

Wilson, R., D'Arrigo, R., Buckley, B., Büntgen, U., Esper, J., Frank, D., Luckman, B., Payette, S., Vose, R., & Youngblut, D. (2007). A matter of divergence: tracking recent warming at hemispheric scales using tree ring data. *Journal of Geophysical Research: Atmospheres*, *112*(D17).

Yin, J., Overpeck, J., Peyser, C., & Stouffer, R. (2018). Big Jump of Record Warm Global Mean Surface Temperature in 2014–2016 Related to Unusually Large Oceanic Heat Releases. Geophysical Research Letters, 45(2), 1069-1078.

Yin, P., & Fan, X. (2001). Estimating R 2 shrinkage in multiple regression: A comparison of different analytical methods. The Journal of Experimental Education, 69(2), 203-224.

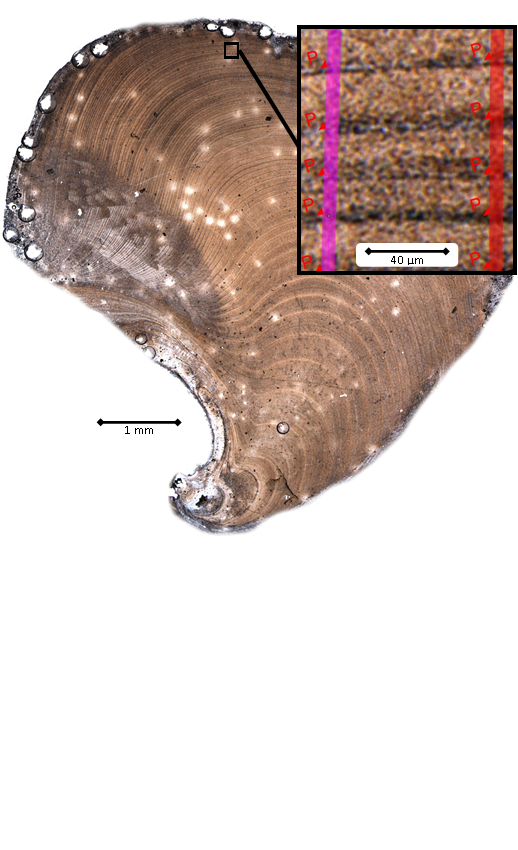
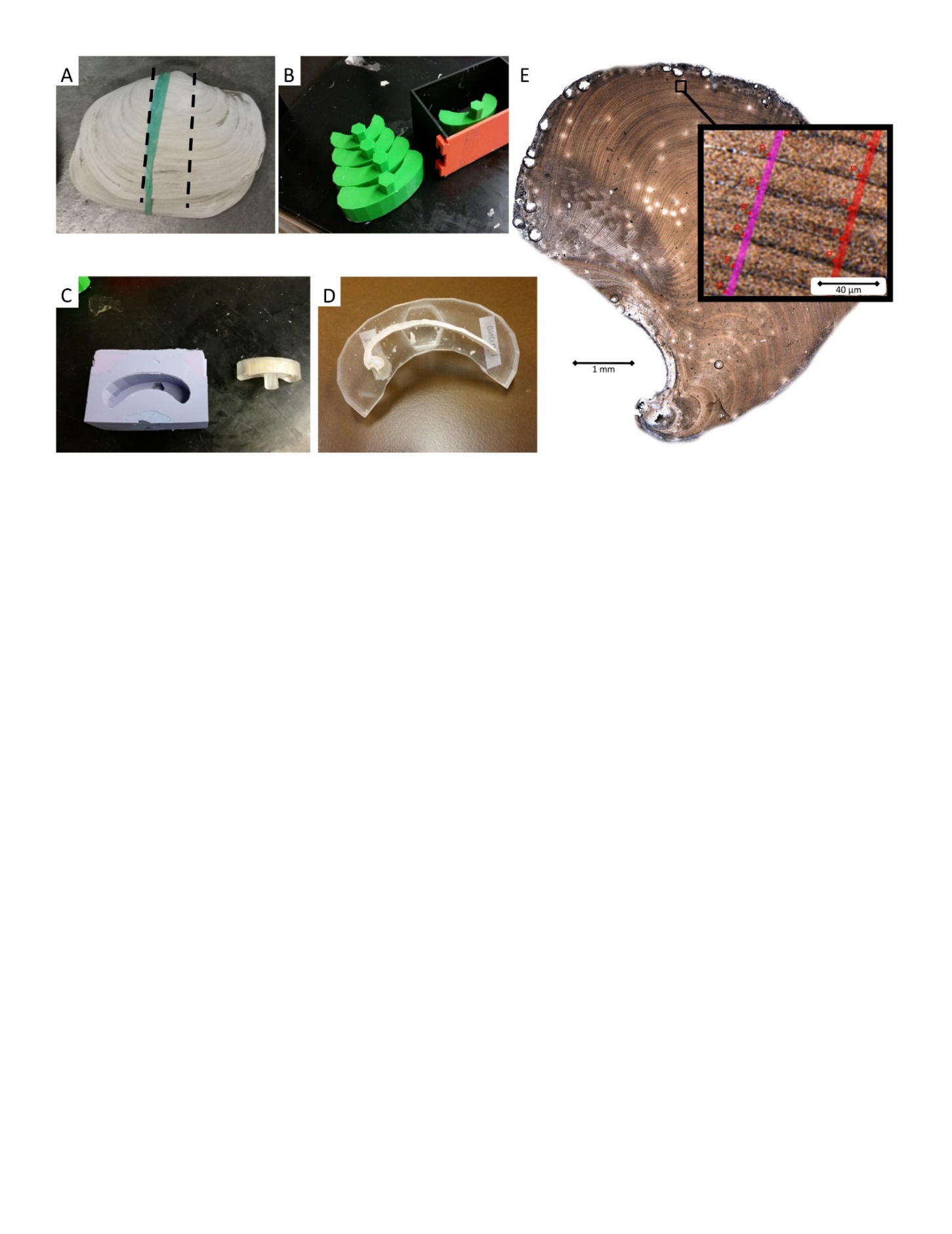
## A.8. Supplemental Material

**Contents**

Figures A.S.1. to A.S.8.

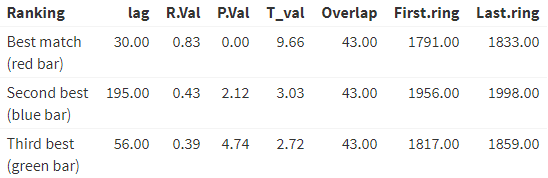
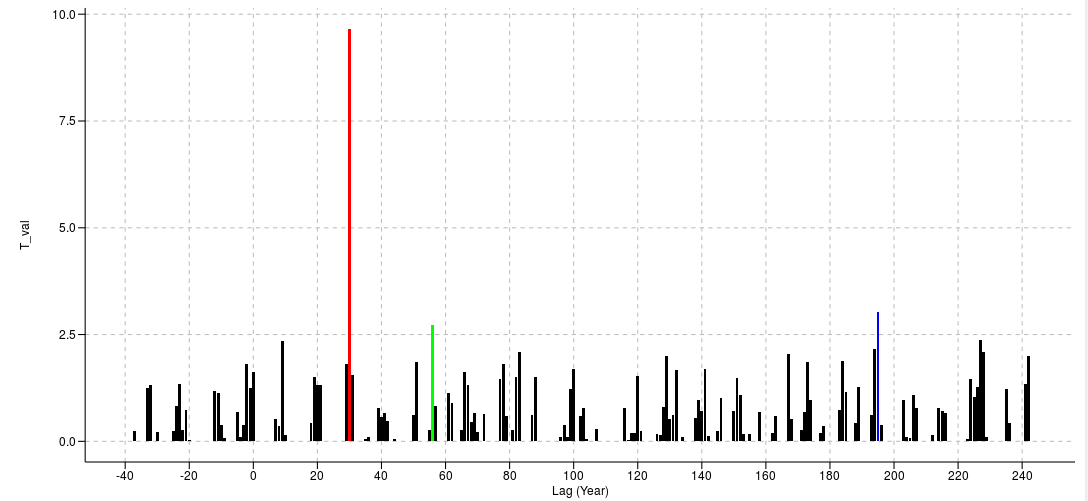
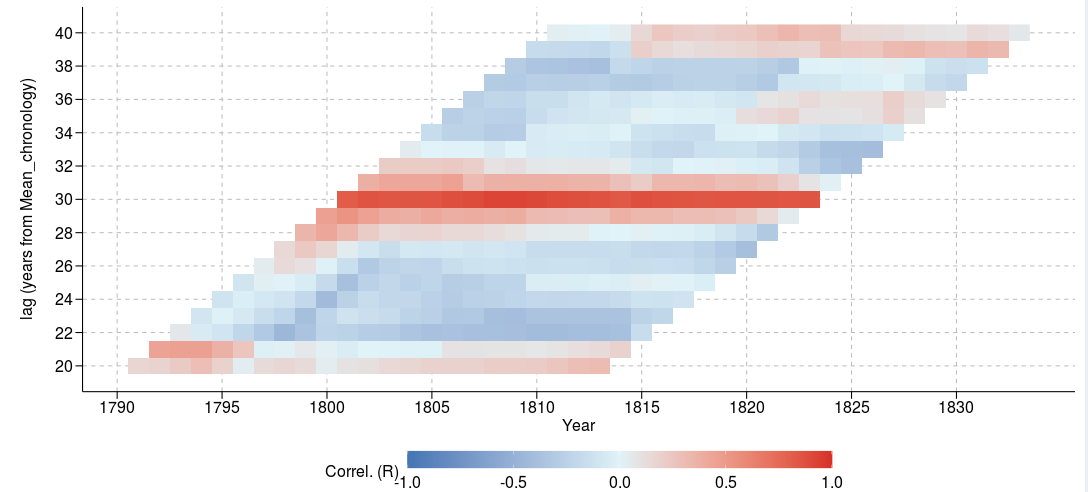
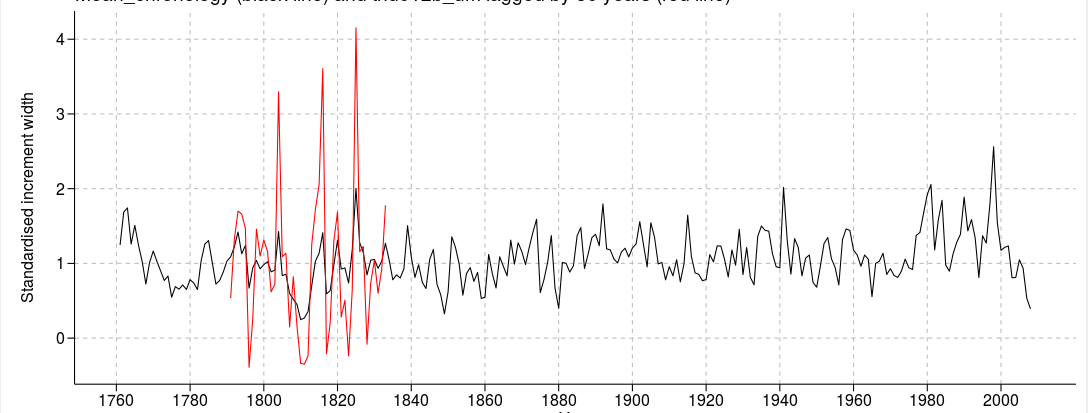
**Introduction**

The Supplemental Material consists of eight figures: A.S.1) detailed illustrations of sample embedding and processing procedure, A.S.2.) an example of crossdating in RingdateR, A.S.3.) full radiocarbon probability distributions for individual and paired samples, A.S.4.) Development of primary regional curve used for master chronology, A.S.5.) a comparison of RCS and EXP detrending by cross-wavelet with instrumental SST, A.S.6.) regional curves developed by various methods, A.S.7.) an illustration of issues observed when combing adaptive power transformation with regional curve standardization, and A.S.8.) Correlations between gridded, regional SST and the Langara reconstruction.



E

Figure A.S.1. Shell embedding and preparation. A) Some shells were received articulated, and only one valve was sectioned around the hinge, B) Molds are cast in 3D-printed casings of variable size, C) Silicone molds are used to embed shells in epoxy resin with hex key to aid in affixing to the saw, D) Embedded section is cut at the hinge apex and the hinge plate is polished, E) Magnified geoduck hinge plate with 1mm scale bar; Inset) Growth measurement axis and growth increment boundaries overlaid on image with 40 μm scale bar.



A

B

C

D

Figure A.S.2. Crossdating in RingdateR. A) Black line: detrended chronology, Red line: new detrended measurement time series at recommended lag B) Windowed correlations represented by red (positive) and blue (negative) – red line in middle shows consistently positive and high correlations at recommended lag. C) T-test values of correlation between chronology and time series at various lags – red, blue, and green bars show 1st, 2nd, and 3rd recommended lags respectively. D) Statistics of recommended crossdating positions.

Histogram

Description automatically generated

Figure A.S.3. Lower panel: Individual age probability distributions based on Marine20 curve and local reservoir correction at Tree Nob. Green distributions represent distinct chronologies of lesser replication (2 ≤ n ≤ 5) Orange, Pink, and Grey distributions represented the well-replicated chronologies. Upper panel: Age probability distributions of well-replicated floating chronologies used for SST reconstruction. These distributions are based on two samples, one near the beginning and one near the end, per chronology.

A picture containing graphical user interface

Description automatically generated

Figure A.S.4. Regional curve development. Gray lines: all crossdated increment width series. Black line: robust biweight mean of all increment widths. Blue line: Regional curve developed from time varying spline of average increment widths. Red, dashed line on left: lower cutoff at ontogenetic age 12 due to high variability of growth around mean. Red, dashed line on right: upper cutoff at ontogenetic age 100 after which minimum sample depth falls below 20.

A picture containing chart

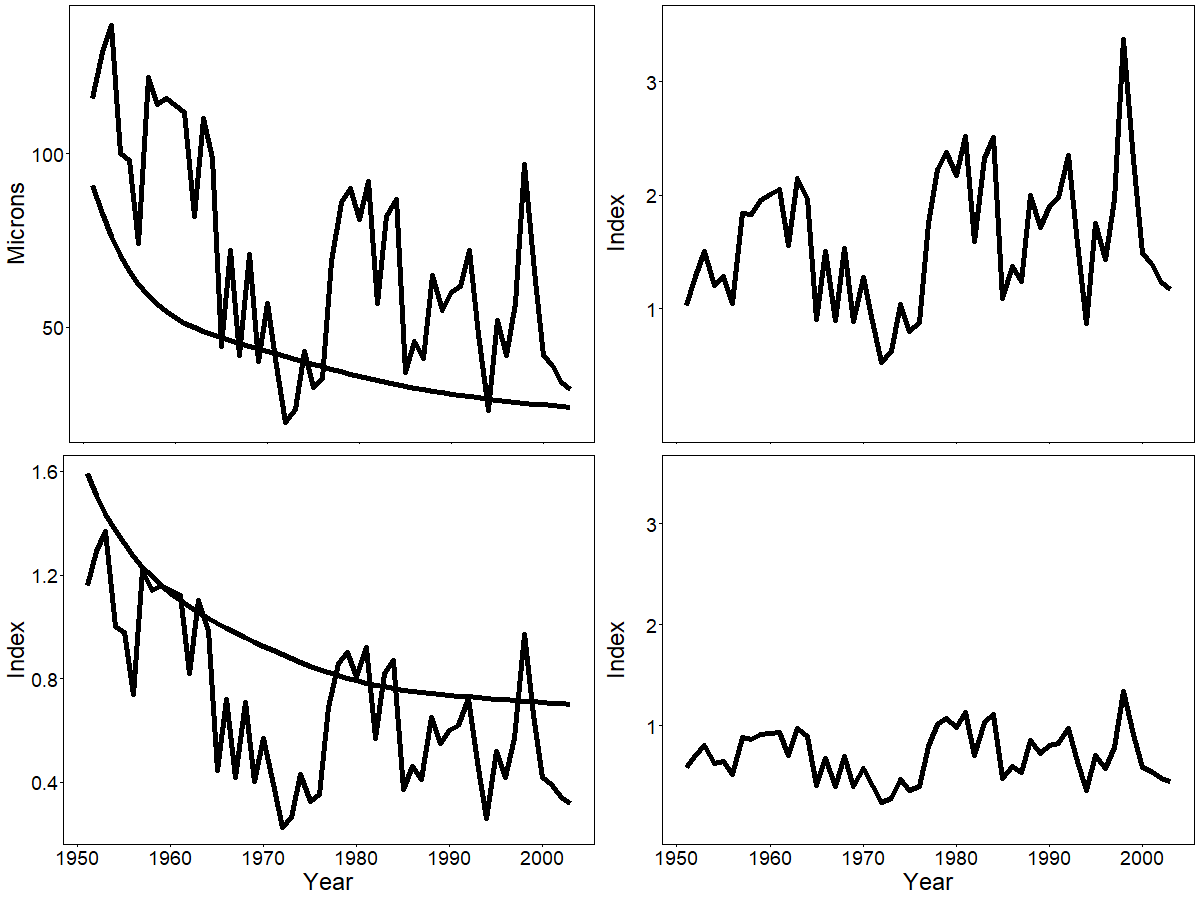
Description automatically generated

Figure A.S.5. Chronology and instrumental SST cross wavelet. A) RCS chronology – Langara SST cross-wavelet power at each year of overlap. Arrows indicate phase coherence, right = in phase, left = antiphase. B) RCS chronology – Langara SST cross-wavelet power averages by period. Red dots indicate significance, p < 0.05. C) EXP chronology – Langara SST cross-wavelet power at each year of overlap. D) EXP chronology – Langara SST cross-wavelet power averages by period.

Chart, line chart

Description automatically generated

Figure A.S.6. Ontogenetic growth and regional curves. Rcall: regional curve (RC) developed from all crossdated samples, used to develop master chronology. RCsigfree: RC developed using signal free detrending, based on only samples from the Modern interval. ModernGrowth: Average increment width for all samples in the Modern interval. ShortLived: Average increment width for samples in the shortest-lived tercile of all crossdated samples. LongLived: Average increment width for samples in the longest-lived tercile of all crossdated samples.



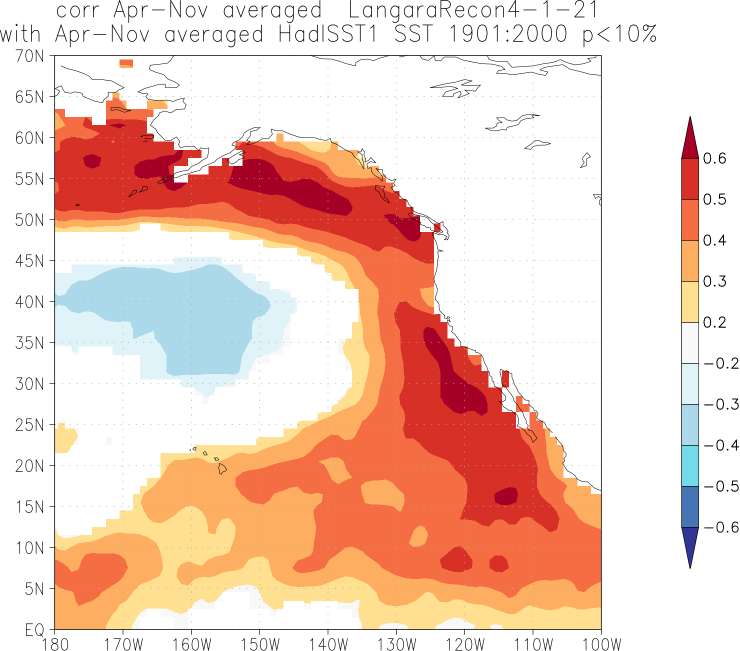
A

B

C

D

Figure A.S.7. Detrending of sample TN006a by standard RCS (top panels) and RCSapt (lower panels). A) Original ring-width series plotted over regional curve, B) ring width index derived from division of original series through the RC, C) power transformed series plotted over the corresponding RC, D) ring width index calculated by subtraction of power transformed series from corresponding RC.



Tree Nob

Langara

Figure A.S.8. Langara SST Reconstruction correlated with gridded, Northeast Pacific SST. The reconstruction is correlated with Hadley Center 1-degree gridded SST (Rayner et al., 2003) over the reconstruction target window (April-November) from 1900-2000.

# Appendix B

A MODERN MULTICENTENNIAL RECORD OF RADIOCARBON VARIABILITY FROM AN EXACTLY DATED BIVALVE CHRONOLOGY AT THE TREE NOB SITE (ALASKA COASTAL CURRENT)

Edge, D. C., Wanamaker, A. D., Staisch, L.M., Reynolds, D. J., Holmes, K.L. & Black, B. A. (in press). A modern multicentennial record of radiocarbon variability from an exactly dated bivalve chronology at the Tree Nob site (Alaska Coastal Current). Radiocarbon. DOI: 10.1017/RDC.2022.83

*The original article, “A modern multicentennial record of radiocarbon variability from an exactly dated bivalve chronology at the Tree Nob site (Alaska Coastal Current)”, is being published by The Licensed Content Publisher, Cambridge University Press, under the Creative Commons CC-BY 4.0 license. The manuscript is redistributed here under the terms of this license, including the attribution above, a link to the license (https://creativecommons.org/licenses/by/4.0/), and an indication of any changes made as follows: The content that follows is shared in the form accepted for publication, with edits made to conform to the format of this dissertation, prior to final processing by University of Cambridge Press.*

**A modern multicentennial record of radiocarbon variability from an exactly dated bivalve chronology at the Tree Nob site (Alaska Coastal Current)**

David C. Edge1, Alan D. Wanamaker2, Lydia M. Staisch3, David J. Reynolds4, Karine L. Holmes2, Bryan A. Black1

1Laboratory of Tree Ring Research, University of Arizona, Bryant Bannister Tree Ring Building, University of Arizona, 1215 E Lowell St, Tucson, AZ 85721, USA

2Department of Geological and Atmospheric Sciences, Iowa State University, 2237 Osborn Dr, Ames, IA 50011, USA

3Geology, Minerals, Energy, and Geophysics Science Center, United States Geological Survey, Moffett Field - Building 19, 345 Middlefield Road MS973, Menlo Park, CA 94025, USA

4Centre for Geography and Environmental Science, Department of Earth and Environmental Science, University of Exeter, Penryn Campus, Treliever Road, Penryn, Cornwall, TR10 9FE, UK

## B.1. Abstract

Quantifying the marine radiocarbon reservoir effect, offsets (ΔR), and ΔR variability over time is critical to improving dating estimates of marine samples while also providing a proxy of water mass dynamics. In the northeastern Pacific, where no high-resolution time series of ΔR has yet been established, we sampled radiocarbon (14C) from exactly dated growth increments in a multicentennial chronology of the long-lived bivalve, Pacific geoduck (*Paneopea generosa*) at the Tree Nob site, coastal British Columbia, Canada. Samples were taken at approximately decadal time intervals from 1725CE to 1920CE and indicate average ΔR values of 256±22 years (1σ) consistent with existing discrete estimates. Temporal variability in ΔR is small relative to analogous Atlantic records except for an unusually old-water event, 1802-1812. The correlation between ΔR and sea surface temperature (SST) reconstructed from geoduck increment width is weakly significant (r2 = .29, p = .03), indicating warm water is generally old, when the 1802-1812 interval is excluded. This interval contains the oldest (-2.1σ) anomaly, and that is coincident with the coldest (-2.7σ) anomalies of the temperature reconstruction. An additional 32 14C values spanning 1952-1980 were detrended using a northeastern Pacific bomb pulse curve. Significant positive correlations were identified between the detrended 14C data and annual El Niño Southern Oscillation (ENSO) and summer SST such that cooler conditions are associated with older water. Thus, 14C is generally relatively stable with weak, potentially inconsistent associations to climate variables, but capable of infrequent excursions as illustrated by the unusually cold, old-water 1802-1812 interval.

## B.2. Introduction

14C is produced in the upper atmosphere from the interaction between cosmic radiation and nitrogen atoms, and due to its predictable rate of decay, is widely used as a geochronometer for dating organic material (e.g., Schuur et al., 2016). The rate of 14C production, however, varies over time, as has been quantified by measuring levels in exactly dated tree rings over the past several millennia (Stuiver et al., 1986a). Information on this year-to-year variability in atmospheric 14C is now used to increase dating accuracy (Büntgen et al., 2018; Pearson et al., 2020; Reimer et al., 2020). In the marine system, 14C dating is complicated by the time necessary for atmospheric 14C to equilibrate across surface-ocean environments. Dating is further complicated due to the mixing of water masses, some of which may have been isolated from the surface and therefore relatively depleted in 14C (Stuiver et al., 1986b). This so-called marine radiocarbon reservoir effect (Stuiver et al., 1986b; Alves et al., 2018) can add 1000 or more years of uncertainty to dating estimates and varies considerably over a range of spatial scales. In the northeast (NE) Pacific north of 40°N, average radiocarbon ages are 600-1000 years older than contemporaneous terrestrial samples (McNeely et al., 1991). Significant spatial heterogeneity in 14Ccontent has been observed on the order of 200 radiocarbon years per 40 km, likely due to differences in upwelling strength (Hutchinson, 2020; McNeely et al., 1991; Robinson and Thompson, 1981; Jones and Jones, 1992; Panich et al., 2018). Thus, correction for this reservoir effect is critical for accurate radiometric dating of marine samples.

In addition to spatial variability, the marine radiocarbon reservoir effect at a given location also fluctuates over time with respect to currents, vertical mixing of deep, 14C -depleted water, and the volume and source of freshwater input, which in most cases mixes in 14C-enriched water. However, robust estimates of the temporal variability of the marine radiocarbon reservoir effect in many regions suffer from the pooling of samples across large geographic region with differing ocean dynamics, the difficulty of sampling consistently through time at a specific location (Ascough et al., 2005; Hutchinson, 2020), and the relative amount of time represented. Archaeological sites can provide some degree of repeated sampling if accurate dates can be established for terrestrial samples known to be contemporaneous with marine samples (Southon et el., 1992; Ascough et al., 2005). A relatively new approach to establishing 14C variability is to sample carbonate from the absolutely dated annual increments of long-lived marine bivalves (Butler et al., 2009; Scourse et al., 2012; Wanamaker et al., 2012; Lower-Spies et al., 2020). Indeed, annual increments in bivalves can be exactly placed in time via the dendrochronology technique of crossdating to generate continuous, annually resolved, multicentennial-length chronologies. From these measurements of 14C in bivalve increments, the marine radiocarbon reservoir effect can be quantified over time for the same location (Butler et al., 2009; Wanamaker et al., 2012; Lower-Spies et al., 2020). To date, this has been successfully applied in the North Atlantic to explore carbon cycling and ocean circulation by acting as a tracer of relatively “old” water depleted of 14C vs. relatively “young” water more recently mingled with the atmospheric 14C reservoir (Butler et al., 2009; Wanamaker et al., 2012; Lower-Spies et al., 2020).

In the northeast Pacific, temporal variability in the marine radiocarbon reservoir remains poorly quantified. To address this issue, we sample growth increments at approximately decadal intervals from a crossdated chronology of Pacific geoduck, a long-lived bivalve (*Paneopea generosa*) abundant from approximately Puget Sound, WA through Kodiak, AK, that occur from the intertidal zone to 60 meters depth. They are typically buried about 1 meter in mud-and-sand sediments and feed by extending a siphon into the water column (Goodwin, 1973). Geoduck shell growth, as measured in the hinge area, is rapid in the first 10-15 years of life, declining exponentially thereafter, while year-to-year growth responds to environmental conditions, primarily water temperature (Cerrato, 2000; Strom et al., 2004). The chronology was developed from samples collected near the Tree Nob Islands in northern British Columbia, Canada, and continuously spans 1725 to 2008 CE (Edge et al. 2021). Growth-increment widths from this chronology were used to develop a sea surface temperature reconstruction, which is closely tied to NE Pacific variability as reflected by a strongly positive correlation (r = 0.62, p < 1.0e-5) with the leading principal component of SST gridded data across the northeast Pacific (Edge et al. 2021). Indeed, the Tree Nob chronology has some of the strongest region-wide climate relationships of any of the network of eight geoduck chronologies developed to date (Strom et al., 2004, Black 2009, Black et al. 2010, Edge et al. 2021). Given this apparent sensitivity to regional climate variability as well as exceptional length, the Tree Nob chronology was chosen for assessing the relationships of local and basin-scale climate indicators with 14C reservoir variability. We also utilized a previously published series of 14C measurements (Kastelle et al., 2011) from these shells sampled through the “bomb pulse” interval (1950-1982) to provide a finer-scale assessment of the link between 14C variability and instrumental climate records. In total, the pre-bomb data augments the finer-scale modern data during the bomb-pulse to provide complementary and longer-term perspectives on marine radiocarbon reservoir variability and relationships to climate and ocean dynamics.

## B.3. Methods and Background

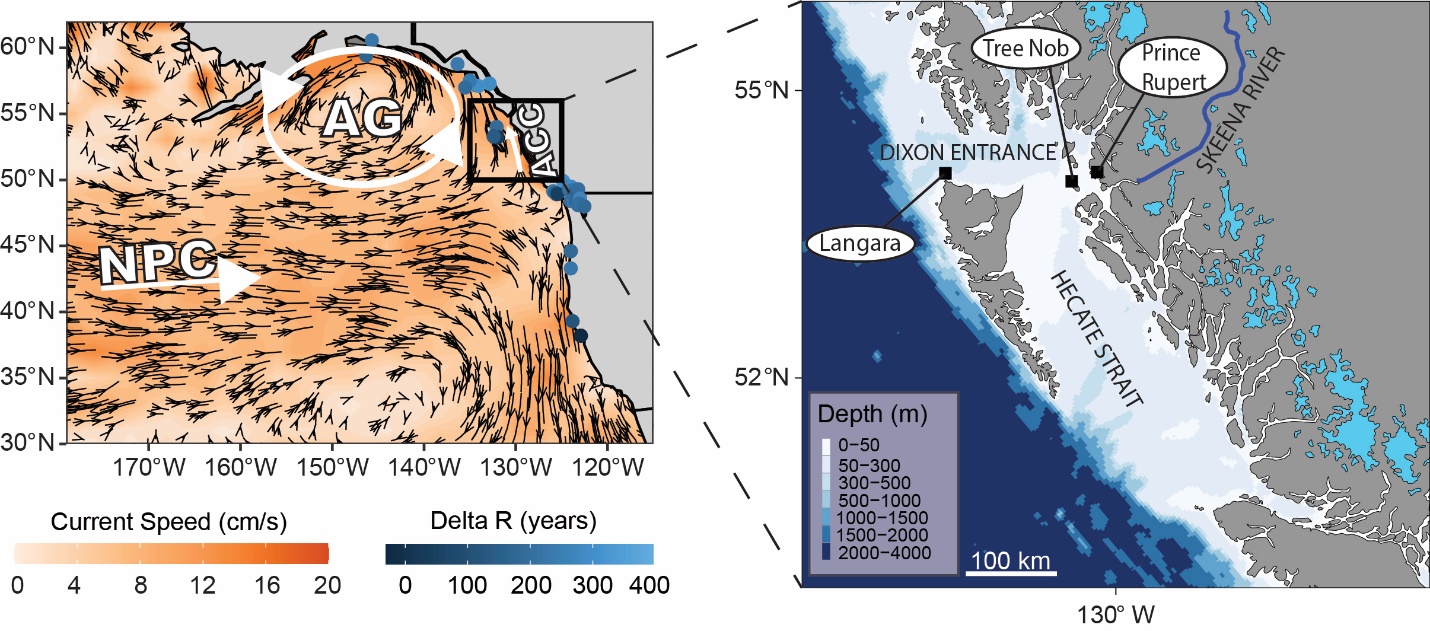
### B.3.1. Oceanographic Setting

The NE Pacific consists of a subpolar, cyclonic gyre in the Gulf of Alaska (GoA) and a subtropical, anti-cyclonic gyre (Fig B.1a). The subpolar Alaska Gyre (AG) consists of the North Pacific Current (NPC) in the south and the Alaska Coastal Current (ACC) along the North American coast, which quickens and narrows west of Kodiak Island to become the Alaska Stream (Dodimead and Hollister, 1958). Variability of transport within the AG is related to fluctuations in the Pacific Decadal Oscillation (PDO), the dominant mode of SST variability in the North Pacific (20-60N), and ultimately to the Aleutian Low (AL; Hristova et al., 2019; Newman et al., 2016). In addition to basin-scale phenomena, the GoA experiences local variability in the magnitude of spring runoff, up/down-welling, and mixing/stratification. Basin-scale patterns may contribute to fluctuations in the strength of the ACC, relative makeup of ACC source waters, up/down-welling in source-water regions, and vertical entrainment (Guilderson et al., 2006; Hristova et al., 2019; Hutchinson, 2020). Guilderson et al. (2006) have proposed a two-end-member mixing regime for the AG based on a linear relationship between 14C and potential density observed in samples collected during the summer of 2002, such that warm, 14C-enriched water enters the AG from the south, and the observed latitudinal gradient is due to vertical entrainment of 14C depleted water within the AG.

The primary, proximate source of water at Tree Nob is from the south via Hecate Strait (HS), a shallow strait that shoals from 200m in the south to just 50m at its northern extent (Fig B.1b). In 1983-84 several current meters were deployed across three transects of HS, which allowed for accurate measurement of flow and the development of surrogate measures of HS flow approximated by three sea level gauges, one to the west and, two to the east (r = 0.81; Crawford et al., 1988). When sea level is high in the east and low in the west, geostrophic flow induces a northward current through HS.

The Tree Nob Island Group lies at the far northeastern extent of Hecate Strait. The Islands are bounded to the north by Brown Passage and to the south by Bell Passage. These waterways connect Chatham Sound, to the east, with Hecate Strait, Dixon Entrance, and the open Northeast Pacific to the west. Due to high winds and strong tides, the Tree Nob site is well mixed with the open ocean (Trites, 1956; Lin and Fissel, 2018). And due to strong northerly flows, Tree Nob is not strongly impacted by freshwater inputs (Lin and Fissel, 2018). Although the study site lies in a quasi-estuarine environment, the strong relationship of geoduck growth increment width with regional- to basin-scale climatic indicators (Edge et al., 2021) suggests that the Tree Nob geoduck integrate environmental conditions across a broad region. Furthermore, the absolutely dated, annually resolved carbonate spans nearly three centuries to yield a marine archive with a uniquely long timespan, precision, and replication in the northeast Pacific.

Figure B.1. Study Site a. Mixed layer ocean currents in the NE Pacific Ocean. Black streamlines indicate general surface currents based on drifting buoy data from the Atlantic Oceanographic and Meteorological Library, National Oceanic and Atmospheric Administration. Individual marine radiocarbon reservoir offset (ΔR) measurements from previous studies shown as colored points (McNeely et al., 1991; Robinson and Thompson, 1981; Jones and Jones, 1992; Panich et al., 2018). General location and direction of the Alaska Gyre (AG), North Pacific Current (NPC), and Alaska Coastal Current (ACC) shown in white. b. Local bathymetry (Amante and Eakins, 2009). Tree Nob geoduck collection site and measurement locations for Langara sea surface temperature and salinity (SST and SSS) and Prince Rupert sea level. Glaciers shown in bright blue. Skeena River is in dark blue.



### B.3.2. Pre-bomb radiocarbon

Geoduck form annual increments (Shaul and Goodwin, 1982), with widths highly correlated to water temperature (Strom et al., 2004; Black et al., 2009, Edge et al., 2021) that can be assigned exact calendar years through crossdating (Black et al., 2008; Kastelle et al., 2011). A crossdated chronology developed from live-collected shells in the Tree Nob Islands (Fig B.1b), and later appended with dead-collected material, extends from CE 1725-2008 (Black et al., 2009; Edge et al., 2021). Live- and dead-collected shells were recovered in sand-and-mud substrate at approximately 10 m water depth. The carbonate of marine organisms is incorporated from the dissolved inorganic carbon (DIC) of ambient seawater and is thus expected to reflect local environmental conditions experienced during shell formation (Adkins et al., 2002; Beirne et al., 2012).

14C samples were obtained from the Tree Nob geoduck shells over the pre-bomb chronology interval of 1725-1920. Crossdated annual increments were sampled to a depth of ~600-800 μm from the shell hinge area using a Merchantek micromill and a Brasseler USAV scriber point (item #H1621.11.008). Samples were then pooled together to obtain ~ 10mg. The micromill was set to maximum drill speed, and several passes were performed ranging from 100–150 μm depth at 55 μm/s scan speed and 55 μm/s plunge speed. In total, 15 shell carbonate samples integrating ten to eleven annual increments were gathered. All samples were taken from the same cut plane used for increment-width measurement to ensure precise calendar-year dating of samples. Samples were sent to the National Ocean Sciences Accelerator Mass Spectrometry facility (NOSAMS Woods Hole, Massachusetts, USA) for 14C analysis.

Laboratory derived error was provided by NOSAMS based on 10 separate measurements of each sample. NOSAMS estimates an additional error of 2.6‰ for replicate samples due to variability in sample collection, processing, and homogeneity. The errors were combined to present a total measurement error of radiocarbon age (NOSAMS, 2020).

ΔR is given by the difference between measured and ‘expected’ radiocarbon age. Each milled geoduck shell sample spanned approximately 10 years, with an average date of formation which corresponds to a date on the Marine20 curve. The Marine 20 curve provides reservoir age estimates in ten-year intervals. Linear interpolation was used to better match these decadally reported radiocarbon age estimates to the average calendar year represented by each sample milled from the geoduck shells. The radiocarbon age of the sample as ‘expected’ by Marine20 was then subtracted from the value measured by NOSAMS to give the ΔR. (Stuiver, 1986; Stuiver and Braziunas, 1993; Heaton et al., 2020), where:

The Marine20 curve (Heaton et al., 2020) accounts for variability in atmospheric 14C production and climate as well as interactions among the ocean, atmosphere, and biosphere. Therefore, the ΔR time series may better represent changes in local 14C content than age-corrected Δ14C by removing as many other sources of variability as possible.

### B.3.3. Radiocarbon and climate covariability

For the pre-bomb dataset, 14Cvalues were compared by linear and polynomial regression to reconstructed, seasonal (mean Apr-Nov) sea surface temperature derived from geoduck growth-increment width as published in Edge et al. (2021). 14Cmeasurements were compared to the Northern Hemisphere (NH) volcanic explosivity index (VEI) by Pearson correlation given the likely influence of such events on ocean circulation and SST (Gao et al., 2008).

Unlike pre-bomb data, the bomb pulse 14Cdata could be compared directly to instrumental climate records. The first of the instrumental records is sea level, which serves as a proxy for HS flow as described by Crawford et al. (1988). Data were obtained from the Canadian Hydrographic Service. Only the Prince Rupert gauge covers the full period of the bomb-pulse data and is thus the only station used, though the pairwise correlations with the other two sites suggest this gauge is representative (r=0.949, 0.921, 0.892; p<0.00001). The upwelling index, as calculated by the National Oceanic and Atmospheric Administration (NOAA) Pacific Fisheries Environmental Laboratory (PFEL) was averaged across 51°N, 131°W and 54°N, 134°W, the two stations nearest to Tree Nob. Monthly freshwater discharge from the Skeena River, just 40 km distant and the second largest river draining British Columbia, was downloaded from the Department of Environment and Natural Resources, Canada (https://wateroffice.ec.gc.ca/report/data\_availability\_e.html?type=historical&station=08EF001&parameter\_type=Flow+and+Level). Monthly mean SST and sea surface salinity (SSS) data, collected at the Langara Lighthouse station, were obtained from Fisheries and Oceans Canada (https://open.canada.ca/data/en/dataset/719955f2-bf8e-44f7-bc26-6bd623e82884; Fig B.1b). Finally, bomb-pulse 14Cdata were compared to three basin-scale indices. Niño 3.4 is a measure of SST in the central, equatorial Pacific with strong connections to northeast Pacific SST and coastal sea surface heights (https://psl.noaa.gov/gcos\_wgsp/Timeseries/Data/Niño34.long.data). The North Pacific Index is a measure of sea surface pressure in the northeast Pacific which serves as a gauge of the strength of the AL (NPI; Trenberth and Hurrell, 1994; https://climatedataguide.ucar.edu/sites/default/files/npindex\_monthly.txt). Finally, the PDO index (https://www.ncdc.noaa.gov/teleconnections/pdo/) is the leading principal component of SSTs in the North Pacific and is closely linked to atmospheric pressure (Mantua et al., 1997; Newman et al., 2016).

Monthly climate data were averaged over 3-year intervals to match the temporal resolution of the 14C sampling (Kastelle et al., 2011) such that climate data for June of 1964 were represented by the average of June 1963, 1964, and 1965 and used for comparison with the 1964-centered 14Cvalue. Correlations to the bomb-pulse 14Cdata were performed in the R package TreeClim (Zang and Biondi, 2015). Significance (α = .01) was calculated by bootstrapping based on methods adapted from DENDROCLIM2002 (Biondi and Waikul, 2004).

### B.3.4. Bomb-pulse radiocarbon data

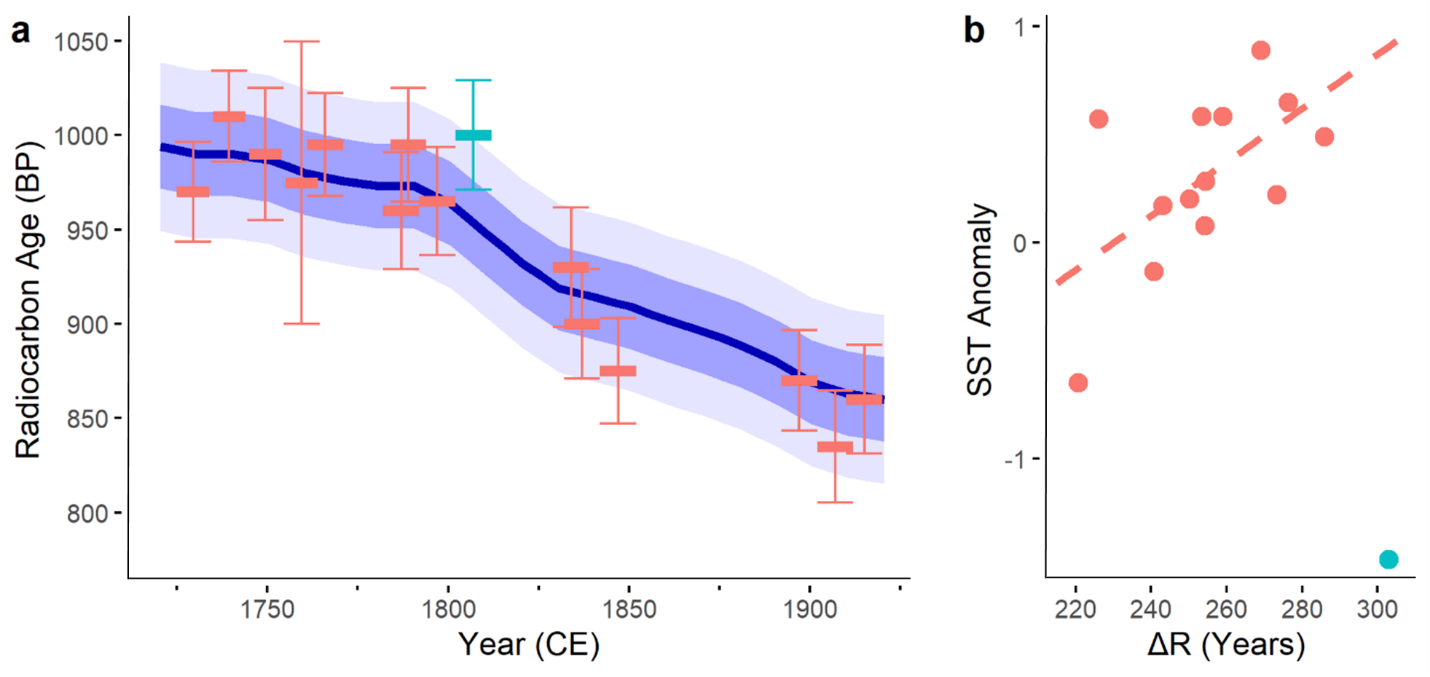
Samples spanning the bomb-pulse period were obtained from the Tree Nob geoduck shells in a previous study (Kastelle et al., 2011). The 32 bomb-pulse samples aggregate an average of three years of growth. The age-corrected Δ14C data were detrended by a latitude-specific, empirically derived marine radiocarbon curve (Helser et al, 2014) to remove the bomb signal. Given that individual, complete increments were not sampled, the mean calendar year represented by a given sample was often a decimal. To facilitate direct comparison to instrumental data, the detrended 14C data were combined in three-year bins by weighted mean over the interval from 1952-1972 (Fig B.S.1.). Weights were assigned proportional to the square of the temporal proximity such that a 14C datum centered at 1965.9, being 0.4 years from 1965.5, was assigned a weight of 2.

## B.4. Results

### B.4.1. Pre-bomb radiocarbon

The average ΔR for all pre-bomb (1725-1920) Tree Nob geoduck 14C samples is +256 years (σ = 22.3 yrs, n=15) and is relatively stable over time. Only one value differs from the mean by more than 2σ, corresponding to increments formed from 1802-1812 (Fig B.2a; Table B.1), though this sample is not an outlier with respect to the distribution of all samples based on Grubb’s test (critical value=2.55, N=15, α=.05). Three samples fall outside 1σ of the mean ΔR and correspond to the years 1784-1794, 1842-1852, and 1902-1912 while all other radiocarbon ages strongly agree with Marine 20 values after ΔR adjustment.

Figure B.2. Pre-bomb radiocarbon. a. Tree Nob radiocarbon ages relative to the Marine 20 curve; points above (below) line indicate older (younger) water. Red bars: radiocarbon age of Tree Nob samples that are not significantly different from Marine 20 (Heaton et al., 2020). Width of bar indicates the range of calendar years sampled. Green bar: 14C sample (middle year of 1807) that is significantly different from average ΔR (2σ). See Methods for error calculation. Blue line: Average radiocarbon age for the mixed layer given by Marine20, corrected by average ΔR of all samples (+256 years). Dark blue shading: 1σ sample error of ΔR; light blue shading is 2σ. b. ΔR vs SST. SST is the 11-year averaging of Langara SST reconstruction, z-scored (Edge et al., 2021). Coloring of points as in panel a. Red dashed line: least squares linear fit of red points.



### B.4.2. Radiocarbon and climate covariability

A linear regression of ΔR onto SST with the 1802-1812 sample removed is marginally significant and has a positive slope (r2=0.29, p = 0.03, Fig B.2b). The 1802-1812 interval contains the most extreme SST and 14C values in the record and suggests an opposing SST-radiocarbon relationship. The Tree Nob ΔR value for 1802-1812, which is greater than the mean ΔR by 2.1σ, coincides with the coldest period in the SST record reconstructed from geoduck (Figure B.3). Temperatures are below 2σ between the years 1808-1812 with the lowest value in 1810 of 3.4σ below the mean. This is coincident with a volcanic eruption in late 1808 with a VEI of 5.5, though other highly explosive eruptions appear to have no relationship to Tree Nob 14C (Gao et al., 2008; Guevara-Murua et al., 2014; Fig B.3).

### B.4.3. Bomb-pulse radiocarbon data

Among regional climate indicators, the relationship between geoduck 14Cin the bomb-pulse interval and SST is perhaps the strongest and most consistent with significant and positive correlations from July through September (Fig B.4). Sea level is positively correlated in February and March while Skeena River discharge negatively correlates in July and August (Fig B.4). The correlation with the NOAA upwelling index is inconsistent with a significant negative relationship for February (r=-0.77, p<0.01) but a positive relationship for June (r=0.65, p<0.01) (Fig B.4). The variable with strongest and most consistent relationships with geoduck 14Cis the Niño3.4 index, which positively correlates for almost every month from May through December and is the only variable with a significant annual relationship (r=0.62, p=0.0056; Fig B.4). The NPI negatively correlates in February and September while the PDO positively correlates in February and October, suggesting young water coincident with a deeper AL and positive PDO (Fig B.S.2.).

Figure B.3. Tree Nob and volcanic proxy records. (a) ΔR in Tree Nob shell material; horizontal uncertainty reflects the 10-11 years represented by each sample while vertical bars are laboratory error. Dark and light gray bands indicate 1- and 2-σ from the mean (white dotted line) of Marine 20. (b) Reconstructed Langara SST and 50% prediction interval from Edge et al. (2021). Dark and light gray bands indicate 1- and 2-σ from the mean (white dotted line) (c) NH VEI based on Gao et al. (2008). Yellow highlighting identifies an interval of synchronous anomalies across all three indicators and corresponds to the years 1802-1812.

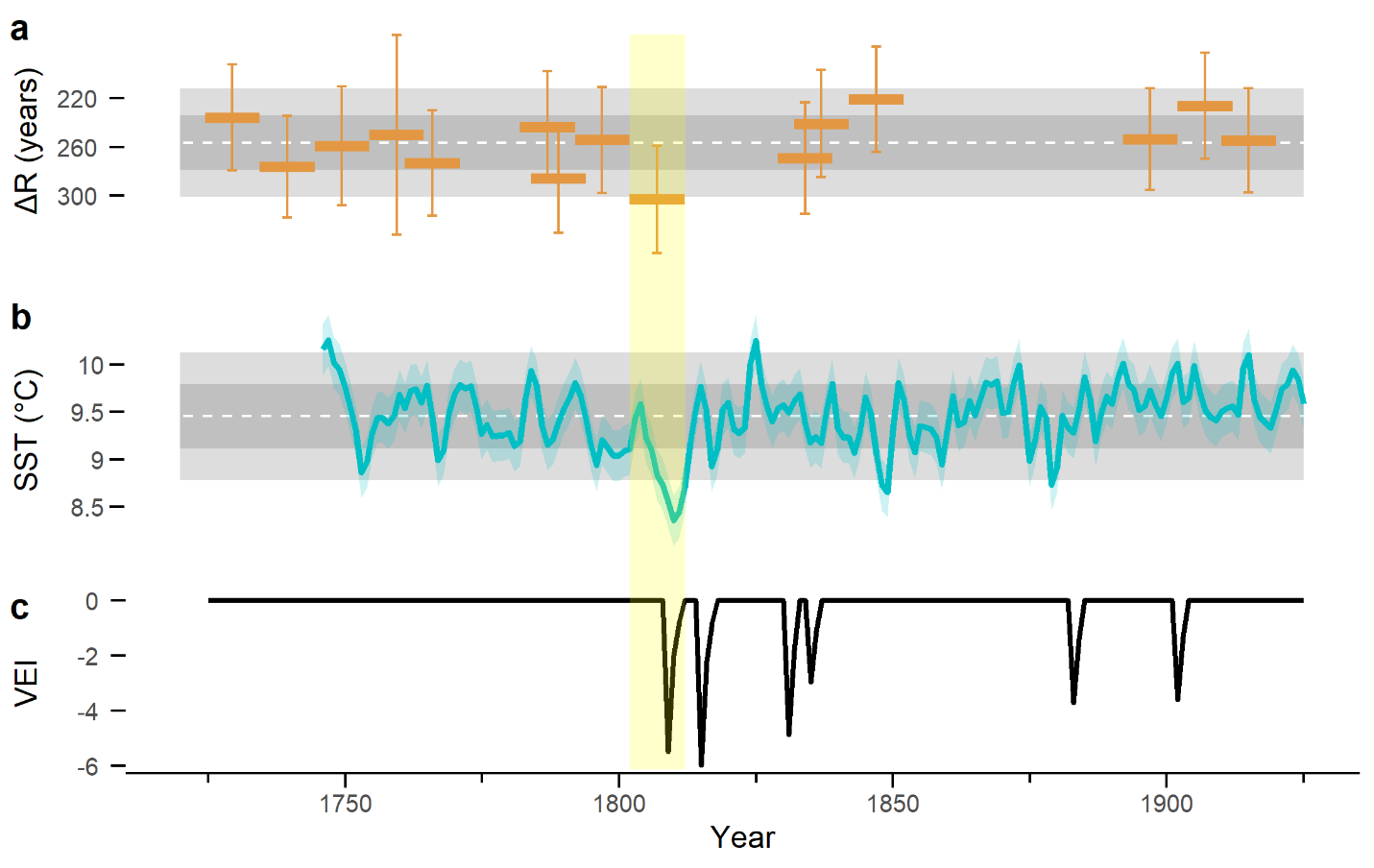
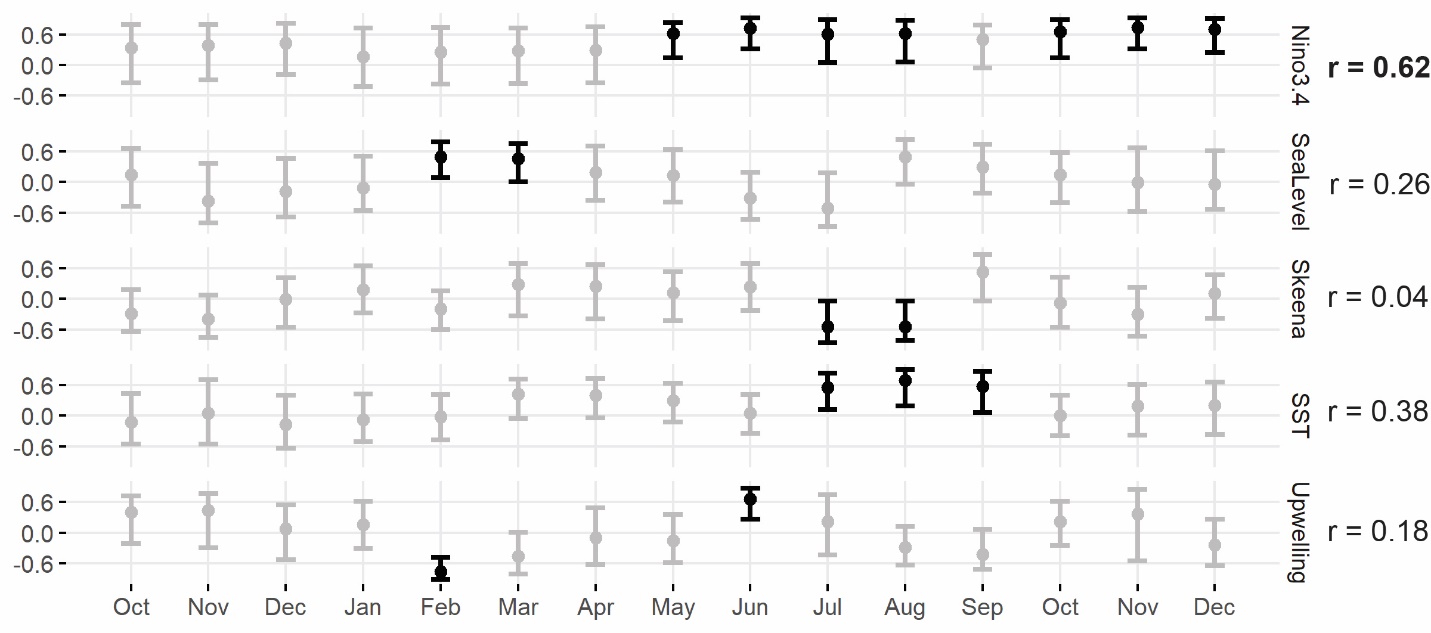


Figure B.4. Bomb 14C correlations with monthly-averaged climate indicators. The level of significance is p <.01 and is calculated by bootstrapping. Significant correlations are shown in black, non-significant in grey. Correlations with annually-averaged (Jan-Dec) climate values are shown to the right of plots, bold font indicates significance at p < .05. Climate data sources given in section 2.3



## B.5. Discussion

The ΔR value at Tree Nob of 256 ± 22 years is consistent with regional measurements in both the 20th century, 249±158 years (McNeely et al., 1991; http://calib.org/marine/, n=12, average distance of 280±138 km, recalculated with Marine20), and throughout the Holocene, 250±195 years (Schmuck et al., 2021). The Tree Nob average is also consistent with the geographically nearest 20th century measurement of 247 ± 50 (McNeely et al., 1991) and the geographically nearest archaeological site, a late-Holocene study only 30 km distant with an estimated 273±38-year reservoir (Edinborough et al., 2016). These Tree Nob values fit into a broader 14Cgradient along the NE Pacific Coast (Fig B.1a), with young water in the California Current (37-38°N, average ΔR = 26 years, σ = 103 years, n = 10; Robinson and Thompson, 1981; Ingram and Southon, 1996; Panich et al., 2018), older water along Vancouver Island to the north (48-49°N, average ΔR = 147 years, σ = 73 years, n = 10; McNeely et al., 1991; Robinson and Thompson, 1981), and even older water in the ACC (54-58°N, average ΔR = 346 years, σ = 83 years, n = 10; McNeely et al., 1991).

Although the Tree Nob sample site is outside the area recommended for use with the Marine20 curve given the potential impacts of sea ice (Heaton et al., 2020), the 14C values observed in the geoduck increments fit remarkably well with predicted values. Indeed, ΔR showed a notable lack of variability over time, straying relatively little outside sample error estimates. The only exception is the excursion during the early 1800s for which age calculations based only on 14C decay could lead to dating errors of 50 years or more (Table 1; Fig B.2a; Fig B.3a). Data from elsewhere in the GoA region also suggests little variability in ΔR over the Holocene, though accuracy in these studies may be impacted by relatively small sample sizes and collections over relatively large spatial domains (Southon et al., 1990; Hutchinson, 2020; Schmuck et al., 2021). Estimates of 14C variability at Tree Nob of 2.9‰ in Δ14C and 22 years ΔR (1σ) are also similar in magnitude to the 1σ decadal variability measured in tropical Pacific corals of 2-3‰ (eg. Druffel, 2001; Grottoli et al., 2003). This variability is much less than that of similar bivalve radiocarbon records from *Arctica islandica* shells in the North Atlantic where 1σ variability of ΔR is on the order of 40-60 years (eg. Butler et al., 2009; Wanamaker et al., 2012; Lower-Spies et al., 2021). Thus, ΔR at Tree Nob appears to be relatively stable, especially compared to published sclerochronological records from the North Atlantic, where water mass variability is likely much greater, but with the potential for significant, if infrequent, excursions as occurred in the early 1800s.

The coupled nature of the new R data and existing Tree Nob geoduck growth-increment width based seasonal (Apr-Nov) SST reconstruction (Edge et al., 2021) provides a unique opportunity to evaluate the connections between marine radiocarbon and climatic variability in this region. In coastal areas currents, vertical mixing, and the volume and source of freshwater runoff may cause temporal variability in 14C (Allen et al., 1995; Hickey and Banas, 2008; Schmuck et al., 2021). Although instrumental records of these more proximal drivers of 14C are not available prior to about 1950, SST is an environmental indicator likely influenced by some combination of these processes (Hristova et al., 2019; Lagerloef, 1995). The subtle variability in 14C over time, uncertainties in the 14C measurements and SST estimates, and decadal averaging of all data may, in part, mask the relationship between 14C and SST. Yet there was still a significant linear relationship between ΔR and SST when calculated without the highly anomalous 1802-1812 sample (Fig B.2b). This relationship suggests old water is relatively warm, counter to expectations, and may represent radiocarbon-old freshwater contributions from radiocarbon-depleted glacial melt or carbonate weathering, though no substantial connections were found to river flows in the higher resolution bomb-pulse data. In contrast, the 1802-1812 datum is consistent with the expectation that colder water, often of relatively deep or more northerly origin, is depleted in 14C. Indeed, anomalously old ΔR and slow geoduck growth in the early 1800s co-occurred with the most extreme cold climate event of the pre-bomb data.

The 1802-1812 14C anomaly is coincident with the largest Northern Hemisphere volcanic eruption of the Tree Nob 14C period of record, the “Unknown” eruption of 1808, which led to significant cooling of the Northern Hemisphere accompanied by other climatic anomalies (Moberg et al., 2005; Gao et al., 2008; Cole-Dai et al., 2009). Climate simulations of the North Pacific response to large tropical volcanic eruptions over the last 600 years show greatly enhanced upwelling-favorable winds at Tree Nob in the years following eruptions (Wang et al., 2012; Zanchettin et al., 2012). The Tree Nob region is strongly dominated by downwelling with an average annual upwelling index of -32 m3/s per 100m of coastline (max = -12 m3/s/100m, min = -69 m3/s/100m, 1947-2020; NOAA PFEL). Yet the 1802-1812 interval may have been associated with an anomalously strong period of upwelling, bringing cold, 14C-depleted water to the surface (Wang et al., 2012; Zanchettin et al., 2012). In addition, wind reversals in the Tree Nob region evident in model simulations of the Unknown eruption may have reduced the advection of warmer, 14C-enriched water into the region, further enhancing “old and cold” conditions. However, extreme 14C anomalies are not evident with other major volcanic events such as Tambora or Krakatoa, possibly due to the seasonality, location, or nature of the eruptions. Thus, despite the coincidence, the relationship between extreme 14C and the Unknown eruption of 1808 may be spurious. Ultimately, a mechanism for this “old and cold” event of the early 1800s cannot be identified from the radiocarbon history, though it does underscore that infrequent yet significant excursions in 14C can occur and appear to be coincident with climate extremes.

The bomb-pulse 14C data provides a complementary perspective on the relationship between 14C and climate at somewhat finer temporal scales within an era spanned by the instrumental record. Positive correlations between SST and 14C run counter to the relationship between SST and 14C over the pre-bomb record and are consistent with the “old and cold” hypothesis. Bomb-pulse 14C data are not significantly correlated with annual mean SST, but instead correlate during the summer months, which is when shells are most actively growing and incorporating carbonate (Shaul and Goodwin, 1982; Edge et al., 2021). Beyond temperature, positive ΔR correlation with sea level suggests that the advection of water masses into the region also influences 14C. High sea level anomalies indicate geostrophic flow from the south from where water is likely 14C-enriched (Fig B.1a). These significant correlations occur in the winter, which is when this transport is likely to be strongest (Crawford et al., 1988), moving water masses into the region that may persist into the growing season. The correlation between upwelling index and 14C is also significant in the winter and could reflect the importance of vertical water movements.This region is almost exclusively dominated by downwelling, which is at its most intense during the winter months with peak mean values from November through February. The negative correlation between February upwelling and geoduck ΔR is consistent with the tendency of warmer, shallower water to be 14C-enriched relative to upwelled water. The cause of the June positive correlation between upwelling and ΔR is less clear and may be spurious. One possibility is that upwelling may encourage stratification during the annual freshet, which typically peaks in June, while downwelling, especially during the annual freshet, forces lighter water under denser water to thereby enhance vertical mixing (Austin and Lentz, 2002). Finally, positive correlations with Skeen River discharge in July and August, the warmest and driest months of the year, may reflect inputs of 14C-depleted glacial melt. This relationship is not likely an artifact of freshwater stratification, as summer Skeena River flow is inversely correlated to local SST (r=-0.47, p=0.0078, 1940:2017 JJA Langara SST).

Correlations between basin-scale indicators and 14C are consistent with those between more local indicators and 14C. For example, positive Niño3.4 values indicate El Niño events, which are associated with warmer water in the region (r=0.62, p<0.00001, 1940-2017 Langara annual SST), and are thus consistent with positive correlations between 14C and SST (Fig B.4). Correlations with Niño3.4 persist through the growing season and beyond. However, lagged correlations into the fall are likely due to lags in climate signals from the tropical Pacific, where the index is calculated, from reaching the mid-latitudes of the NE Pacific. Furthermore, modelling work demonstrates a strengthening of the ACC during El Niño events (Melsom et al., 1999), which would increase the advection of more southerly, 14C-enriched water into the study region. A metanalysis of Holocene-timescale 14C variability in the northeast Pacific suggests ENSO may be the most predictive climate variable for coastal 14C, which is reflected by the strong correlations observed here (Hutchinson, 2020). In contrast, 14C correlations with PDO and NPI are considerably weaker and less consistent with ENSO. However, the nature of the correlations is consistent with overall patterns of temperature and transport in the NE Pacific. Positive values of the PDO are associated with lower atmospheric pressure over the NE Pacific and relatively strong advection of water from the south along the cost and thus through Hecate Strait. The NPI is closely related to the PDO, but more directly measures regional pressure, and is opposite in sign, explaining its negative correlation with 14C relative to a positive correlation with PDO. Thus, NPI and PDO likely reflect the influence of the Aleutian Low with its ties to both AG advection (Hristova et al., 2019) and SST (Newman et al., 2016). Indeed, the intensity of the Aleutian Low is greatest in the winter, which coincides with the seasonality of the relationship with 14C for both indicators.

Bomb-pulse 14C are important confirmatory data for the pre-bomb but must be interpreted with caution. There may be biases in the 14C bomb-pulse model used to detrend the data. Also, the bomb pulse data are from a very limited temporal window that spans a single cool regime in the North Pacific that began in 1946 and lasted through 1976 (Miller et al., 1994; Mantua et al., 1997). Thus, the bomb-pulse data lack the variability in environmental conditions covered by the pre-bomb data, absent the contrast of a warm ocean regime let alone climatic extremes such as the cold period of the early 1800s. This may help explain why the PDO, with energy in interdecadal timescales, did not correlate as strongly with the bomb-pulse data as ENSO, which has greater energy on interannual timescales. These differing temporal resolutions might also change the 14C-SST relationships. Finally, the bomb-pulse itself also changes the ocean-atmosphere exchange dynamics by enhancing Δ14C difference between the two reservoirs, which may affect relationships between 14C and climate. Because the bomb-pulse 14C time series is short (n=21) and a large number of correlation analyses were performed without a Bonferroni correction, the significance of these monthly correlations should be interpreted cautiously. Therefore, to reduce the number of spurious results, we implemented an α=.01 threshold for significance testing in TreeClim. Yet, despite these potential shortcomings, positive bomb-pulse 14C correlations with SST are consistent with the “old is cold” hypothesis in which colder water masses tend to be of deeper or more northerly origin and depleted of 14C. This relationship, as well as the correlations with sea level, NPI, PDO, and ENSO are also consistent with the two-end-member mixing regime proposed by Guilderson et al. (2006). Notably, this is opposite to the relationship in the pre-bomb data, wherein cold water is coincident with radiocarbon enrichment. Yet relationships between climate and ΔR are relatively weak, as is the variability in ΔR over time, suggesting that during most years radiocarbon is generally stable and minimally affected, if at all, by environmental variability at this site. The radiocarbon excursion in the early 1800s and co-occurrence with unusual cold does, however, indicate that the system is subject to anomalies consistent with the expectation of “old and cold” water masses.

Ultimately, the Tree Nob 14C time series suggests ΔR is relatively stable in the decadally averaged timescales sampled here, but with the potential for significant excursions under climatic extremes such as the cold period of the early 1800s. The 1802-1812 anomaly may be related to a brief, volcanic-induced climate excursion, an example of an event which may not be captured when sampling a lower temporal resolution. However, this pattern is limited to one location in the shallow nearshore environment and therefore may not well represent deeper or offshore locations. Other 14C archives may better address these locations to provide a contrast for the nearshore. For example, Pacific rockfish can live for a century or longer, form annual increments that can be crossdated, and thus could provide as source of absolutely dated, offshore carbonate that would pre-date the bomb pulse (Black et al., 2008; Sydeman et al., 2014, van der Sleen 2014 POP paper). Indeed, a network of rockfish and geoduck chronologies could be sampled for 14C along the NE Pacific to better quantify temporal and spatial patterns of 14C variability. The timescales involved in 14C analysis could also be refined if individual increments are sampled to reveal interannual variability rather than the decadal-scale resolution addressed here. Given that the Tree Nob chronology covers 58% of the past 1500 years, there is the possibility of greatly increasing the temporal depth of the 14C chronology as more subfossil shells are collected to fill gaps, which could provide further insight into 14C variability in the NE Pacific and potentially refine dating of other organic marine material of archaeological, geological, or climatic importance.

## B.6. Acknowledgements

This work is funded by the National Science Foundation (AGS Award Number: 1855628 to BAB; Award Number: 1602751 to ADW).

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Geological Survey or the United States Government. The publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

## B.7. References

Adkins, J. F., Griffin, S., Kashgarian, M., Cheng, H., Druffel, E. R. M., Boyle, E. A., et al. (2002). Radiocarbon dating of deep‐sea corals. Radiocarbon, 44(2), 567–580. https://doi.org/10.1017/S0033822200031921

Alves, E. Q., Macario, K., Ascough, P., & Bronk Ramsey, C. (2018). The worldwide marine radiocarbon reservoir effect: Definitions, mechanisms, and prospects. Reviews of Geophysics, 56, 278–305.

Amante, C., & Eakins, B. W. (2009). ETOPO1 arc-minute global relief model: procedures, data sources and analysis.

Ascough, P., Cook, G., & Dugmore, A. (2005). Methodological approaches to determining the marine radiocarbon reservoir effect. *Progress in Physical Geography*, *29*(4), 532-547.

Austin, J. A., & Lentz, S. J. (2002). The inner shelf response to wind-driven upwelling and downwelling. *Journal of Physical Oceanography*, *32*(7), 2171-2193.

Beirne, E. C., Wanamaker, A. D., & Feindel, S. C. (2012). Experimental validation of environmental controls on the δ13C of Arctica islandica (ocean quahog) shell carbonate. Geochimica et Cosmochimica Acta, 84, 395–409. https://doi.org/10.1016/j.gca.2012.01.021

Biondi, F., & Waikul, K. (2004). DENDROCLIM2002: A C++ program for statistical calibration of climate signals in tree-ring chronologies. *Computers & Geosciences*, *30*(3), 303-311.

Black, B. A., Copenheaver, C. A., Frank, D. C., Stuckey, M. J., & Kormanyos, R. E. (2009). Multi-proxy reconstructions of northeastern Pacific sea surface temperature data from trees and Pacific geoduck. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *278*(1-4), 40-47.

Black, B. A., Gillespie, D. C., MacLellan, S. E., & Hand, C. M. (2008). Establishing highly accurate production-age data using the tree-ring technique of crossdating: a case study for Pacific geoduck (Panopea abrupta). *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(12), 2572-2578.

Büntgen, U., Wacker, L., Galván, J. D., Arnold, S., Arseneault, D., Baillie, M., Beer, J., Bernabei, M., Bleicher, N., & Boswijk, G. (2018). Tree rings reveal globally coherent signature of cosmogenic radiocarbon events in 774 and 993 CE. *Nature communications*, *9*(1), 1-7.

Butler, P. G., Scourse, J. D., Richardson, C. A., Wanamaker Jr, A. D., Bryant, C. L., & Bennell, J. D. (2009). Continuous marine radiocarbon reservoir calibration and the 13C Suess effect in the Irish Sea: Results from the first multi-centennial shell-based marine master chronology. *Earth and Planetary Science Letters*, *279*(3-4), 230-241.

Cole‐Dai, J., Ferris, D., Lanciki, A., Savarino, J., Baroni, M., & Thiemens, M. H. (2009). Cold decade (AD 1810–1819) caused by Tambora (1815) and another (1809) stratospheric volcanic eruption. *Geophysical Research Letters*, *36*(22).

Crawford, W. R., Huggett, W. S., & Woodward, M. J. (1988). Water transport through Hecate Strait, British Columbia. *Atmosphere-ocean*, *26*(3), 301-320.

Dodimead, A., & Hollister, H. (1958). Progress report of drift bottle releases in the northeast Pacific Ocean. *Journal of the Fisheries Board of Canada*, *15*(5), 851-865.

Edge, D. C., Reynolds, D. J., Wanamaker, A. D., Griffin, D., Bureau, D., Outridge, C., Stevick, B. C., Weng, R., & Black, B. A. (2021). A Multicentennial Proxy Record of Northeast Pacific Sea Surface Temperatures From the Annual Growth Increments of Panopea generosa. *Paleoceanography and Paleoclimatology*, *36*(9), e2021PA004291.

Edinborough, K., Martindale, A., Cook, G. T., Supernant, K., & Ames, K. M. (2016). A marine reservoir effect∆ R value for kitandach, in Prince Rupert Harbour, British Columbia, Canada. *Radiocarbon*, *58*(4), 885-891.

Gao, C., Robock, A., & Ammann, C. (2008). Volcanic forcing of climate over the past 1500 years: An improved ice core‐based index for climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D23).

Guilderson, T. P., Roark, E. B., Quay, P. D., Page, S. R. F., & Moy, C. (2006). Seawater radiocarbon evolution in the Gulf of Alaska: 2002 observations. *Radiocarbon*, *48*(1), 1-15.

Heaton, T. J., Köhler, P., Butzin, M., Bard, E., Reimer, R. W., Austin, W. E., Ramsey, C. B., Grootes, P. M., Hughen, K. A., & Kromer, B. (2020). Marine20—the marine radiocarbon age calibration curve (0–55,000 cal BP). *Radiocarbon*, *62*(4), 779-820.

Helser, T. E., Kastelle, C. R., & Lai, H.-l. (2014). Modeling environmental factors affecting assimilation of bomb-produced Δ14C in the North Pacific Ocean: Implications for age validation studies. *Ecological modelling*, *277*, 108-118.

Hristova, H. G., Ladd, C., & Stabeno, P. J. (2019). Variability and trends of the Alaska Gyre from Argo and Satellite Altimetry. *Journal of Geophysical Research: Oceans*, *124*(8), 5870-5887.

Hutchinson, I. (2020). Spatiotemporal variation in ΔR on the West Coast of North America in the late Holocene: implications for dating the shells of marine mollusks. *American Antiquity*, *85*(4), 676-693.

Ingram, B. L., & Southon, J. R. (1996). Reservoir ages in eastern Pacific coastal and estuarine waters. *Radiocarbon*, *38*(3), 573-582.

Jones, T. L., & Jones, D. A. (1992). Elkhorn Slough revisited: Reassessing the chronology of CA-MNT-229. *Journal of California and Great Basin Anthropology*, *14*(2), 159-179.

Kastelle, C. R., Helser, T. E., Black, B. A., Stuckey, M. J., Gillespie, D. C., McArthur, J., Little, D., Charles, K. D., & Khan, R. S. (2011). Bomb-produced radiocarbon validation of growth-increment crossdating allows marine paleoclimate reconstruction. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *311*(1-2), 126-135.

Lower‐Spies, E. E., Whitney, N. M., Wanamaker, A. D., Griffin, S. M., Introne, D. S., & Kreutz, K. J. (2020). A 250‐year, decadally resolved, radiocarbon time history in the Gulf of Maine reveals a hydrographic regime shift at the end of the Little Ice Age. *Journal of Geophysical Research: Oceans*, *125*(9), e2020JC016579.

Lin, Y., & Fissel, D. B. (2018). The ocean circulation of Chatham Sound, British Columbia, Canada: Results from numerical modelling studies using historical datasets. Atmosphere-Ocean, 56(3), 129-151.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, *78*(6), 1069-1080.

McNeely, R., & McCuaig, S. (1991). Geological Survey of Canada radiocarbon dates XXIX.

Melsom, A., Meyers, S. D., O'Brien, J. J., Hurlburt, H. E., & Metzger, J. E. (1999). ENSO effects on Gulf of Alaska eddies. *Earth Interactions*, *3*(1), 1-30.

Miller, A. J., Cayan, D. R., Barnett, T. P., Graham, N. E., & Oberhuber, J. M. (1994). The 1976-77 climate shift of the Pacific Ocean. *Oceanography*, *7*(1), 21-26.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., & Karlén, W. (2005). Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data. *Nature*, *433*(7026), 613-617.

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., Mantua, N. J., Miller, A. J., Minobe, S., & Nakamura, H. (2016). The Pacific decadal oscillation, revisited. *Journal of Climate*, *29*(12), 4399-4427.

NOSAMS. (2020). *Radiocarbon Data and Calculations*. National Ocean Sciences Accelerator Mass Spectrometry. https://www2.whoi.edu/site/nosams/client-services/radiocarbon-data-calculations/

Panich, L. M., Schneider, T. D., & Engel, P. (2018). The marine radiocarbon reservoir effect in Tomales Bay, California. *Radiocarbon*, *60*(3), 963-974.

Pearson, C., Salzer, M., Wacker, L., Brewer, P., Sookdeo, A., & Kuniholm, P. (2020). Securing timelines in the ancient Mediterranean using multiproxy annual tree-ring data. *Proceedings of the National Academy of Sciences*, *117*(15), 8410-8415.

Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., Butzin, M., Cheng, H., Edwards, R. L., & Friedrich, M. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon*, *62*(4), 725-757.

Robinson, S. W., & Thompson, G. (1981). Radiocarbon corrections for marine shell dates with application to southern Pacific Northwest Coast prehistory. *Syesis*, *14*, 45-57.

Schmuck, N., Reuther, J., Baichtal, J. F., & Carlson, R. J. (2021). Quantifying marine reservoir effect variability along the Northwest Coast of North America. *Quaternary Research*, 1-22.

Schuur, E. A., Druffel, E. R., & Trumbore, S. E. (2016). Radiocarbon and climate change: Mechanisms, applications and laboratory techniques. Springer.

Scourse, J. D., Wanamaker, A. D., Weidman, C., Heinemeier, J., Reimer, P. J., Butler, P. G., et al. (2012). The marine radiocarbon bomb pulse across the temperate North Atlantic: A compilation of Δ14C time histories from Arctica islandica growth increments. Radiocarbon, 54(2), 165–186.

Shaul, W., & Goodwin, L. (1982). Geodock (Panope generosa: Bivalvia) age as determined by internal growth lines in the shell. *Canadian Journal of Fisheries and Aquatic Sciences*, *39*(4), 632-636.

Southon, J., Nelson, D., & Vogel, J. (1992). The determination of past ocean-atmosphere radiocarbon differences. NATO advanced research workshop on the last deglaciation: Absolute and radiocarbon chronologies.

Strom, A., Francis, R. C., Mantua, N. J., Miles, E. L., & Peterson, D. L. (2004). North Pacific climate recorded in growth rings of geoduck clams: a new tool for paleoenvironmental reconstruction. *Geophysical Research Letters*, *31*(6).

Stuiver, M., & Braziunas, T. F. (1993). Modeling atmospheric 14C influences and 14C ages of marine samples to 10,000 BC. *Radiocarbon*, *35*(1), 137-189.

Stuiver, M., Kromer, B., Becker, B., & Ferguson, C. W. (1986). Radiocarbon age calibration back to 13,300 years BP and the 14C age matching of the German oak and US bristlecone pine chronologies. *Radiocarbon*, *28*(2B), 969-979.

Stuiver, M., Pearson, G. W., & Braziunas, T. (1986). Radiocarbon age calibration of marine samples back to 9000 cal yr BP. *Radiocarbon*, *28*(2B), 980-1021.

Sydeman, W., García-Reyes, M., Schoeman, D. S., Rykaczewski, R., Thompson, S., Black, B., & Bograd, S. (2014). Climate change and wind intensification in coastal upwelling ecosystems. *Science*, *345*(6192), 77-80.

Trenberth, K. E., & Hurrell, J. W. (1994). Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics*, *9*(6), 303-319.

Trites, R. W. (1956). The oceanography of Chatham Sound, British Columbia. Journal of the Fisheries Board of Canada, 13(3), 385-434.

van der Sleen, P., Dzaugis, M. P., Gentry, C., Hall, W. P., Hamilton, V., Helser, T. E., Matta, M. E., Underwood, C. A., Zuercher, R., & Black, B. A. (2016). Long-term Bering Sea environmental variability revealed by a centennial-length biochronology of Pacific ocean perch Sebastes alutus. *Climate Research*, *71*(1), 33-45.

Wanamaker, A. D., Butler, P. G., Scourse, J. D., Heinemeier, J., Eiríksson, J., Knudsen, K. L., & Richardson, C. A. (2012). Surface changes in the North Atlantic meridional overturning circulation during the last millennium. *Nature communications*, *3*(1), 1-7.

Wang, T., Otterå, O. H., Gao, Y., & Wang, H. (2012). The response of the North Pacific Decadal Variability to strong tropical volcanic eruptions. *Climate Dynamics*, *39*(12), 2917-2936.

Zanchettin, D., Timmreck, C., Graf, H.-F., Rubino, A., Lorenz, S., Lohmann, K., Krüger, K., & Jungclaus, J. (2012). Bi-decadal variability excited in the coupled ocean–atmosphere system by strong tropical volcanic eruptions. *Climate Dynamics*, *39*(1), 419-444.

Zang, C., & Biondi, F. (2015). treeclim: an R package for the numerical calibration of proxy‐climate relationships. *Ecography*, *38*(4), 431-436.

## B.8. Supplemental Material

**Contents**

Figures B.S.1. to B.S.2.

**Introduction**

The Supplemental Material information consists of two figures: B.S.1) development of the bomb-pulse radiocarbon indices from raw measurements and detrending curve, B.S.2.) bomb-pulse radiocarbon relationships with basin-scale climate indices

Chart, scatter chart

Description automatically generated

**Figure B.S.1.** Bomb-pulse 14C standardization. a. black dots: individual 14C samples given in per mil 14C. red line: Northeast Pacific average. green line: latitude-corrected 14C average. blue line: upwelling-corrected 14C average. Curves based on Helser et al., 2014. b. black dots: residuals of measured 14C from latitude-corrected curve (green line in panel a). blue line: annual, weighted mean of residuals.

Chart, box and whisker chart

Description automatically generated

**Figure B.S.2.** Correlations between 3-year averaged basin-scale indices and Tree Nob 14C. Monthly indices are averaged over three years to meet the temporal resolution of 14C. Significance and confidence intervals bootstrapped in TreeClim at p < 0.01. Significant correlations are shown in black, non-significant in grey. Annual correlations given to right of plots and considered significant at p < .05.

# Appendix c

ESTIMATING UNCERTAINTIES IS CLIMATE RECONSTRUCTIONS FROM CROSSDATED PROXIES

Edge, D.C., Meko, D.M., Thompson, D.M., Trouet V.M., Black, B.A.

*The following article has been prepared for submission to the journal Dendrochronologia. The work has not yet been submitted and is therefore under no prior copyright agreement.*

**Estimating Uncertainties in Climate Reconstructions from Crossdated Proxies**

David C. Edge1,2, David M. Meko1, Diane M. Thompson1, Valerie M. Trouet1, Bryan A. Black1

1Laboratory of Tree Ring Research, University of Arizona, 1215 E Lowell St, Tucson, AZ 85721, USA

2School of Earth and Sustainability, Northern Arizona University, 624 Knoles Dr, Flagstaff, AZ 86011, USA

## C.1. Abstract

Accurate estimates of uncertainty provide a gauge of signal-to-noise in paleoclimate reconstructions, which is critical for the detection and attribution environmental change. The reconstruction uncertainties derived from crossdated archives, including trees and bivalves, represent the envelope within which some percentage of the true climate values will be located in the pre-instrumental interval. The representativeness of a chronology value of the population in a given year is variable and this uncertainty must be included in the reconstruction uncertainty. To address this issue, we calculated pre-instrumental confidence intervals (uncertainties) based on chronologies from marine bivalve (*Panopea generosa*) ring widths, mountain hemlock (*Tsuga mertensiana)* maximum latewood density, and blue oak (*Quercus douglasii*) ring widths and their respective climate targets. For each of the three original chronologies, 100 synthetic chronologies were developed in which chronology sample depth, average correlation between series in a chronology (rbar), first-order autocorrelation (AR1), and correlation to climate target were varied. For all 303 chronologies (300 synthetic, 3 original), 90%- and 50% confidence intervals were calculated using all combinations of three different chronology bootstrapping techniques and two different reconstruction error measures. For each reconstruction, the percentage of target values captured by the confidence intervals was measured in a period independent of the reconstruction calibration and confidence interval calculations. The percentage of climate target values captured by these confidence intervals was also found to covary with the rbar, such that chronologies with lower rbar produced excessively wide confidence intervals. Chronology properties had no impact on the reliability of confidence intervals by maximum entropy bootstrapping (MEboot). The robustness of MEboot confidence intervals across a range of chronology properties and chronology-target relationships in these trials lends a reasonable expectation of accuracy for uncertainties calculated in this way.

## C.2. Introduction

In dendroclimatology, annual reconstructions of past climate typically consist of “estimates” of a particular climate variable at a particular location, for example, mean July-September air temperature. These estimates are based on a calibration relationship between the tree-ring record and the measured July-September temperature over an interval when both records are available, which we call the instrumental overlap interval (IOI). Using this relationship, the values of the tree-ring record are transformed into equivalent temperature units, called the reconstruction, providing estimates of temperature over the full extent of the tree-ring chronology (Fritts, 1976).

A reconstruction that fits a target climate variable closely in the IOI may be a very good proxy for that target. However, it may demonstrate similar character by chance, or the reconstruction model may be overfitted. Several methods have been developed to test the “skill” of a reconstruction, or how the reconstruction compares with a chance relationship. Many reconstructions are evaluated by splitting the IOI into independent calibration and verification intervals. Researchers utilize these intervals to measure the coefficient of determination (R2), Reduction of Error (RE), and the Coefficient of Efficiency (CE) to assess reconstruction skill (see Fritts, 1976; Briffa et al., 1988; Cook and Kairiukstis, 1990 and references therein). The skill/uncertainty can be ascribed to the domains of 1) the biological response to climate as captured by the proxy, 2) the measurement and manipulation of the proxy data, and 3) the transformation of the proxy data into units of climate (Fritts, 1976). In compliment to measures of reconstruction skill, reconstruction confidence intervals (hereafter, confidence intervals) often accompany dendroclimatic reconstructions, for example a band around the reconstruction estimated to envelope 90% of the true climate values (Cook and Kairiukstis, 1990; National Research Council, 2006).

Confidence intervals purport to show the quality of fit of a reconstruction. A recent reconstruction of drought severity translates the uncertainty domains described above into practical distinctions among error sources from detrending, chronology, and calibration (Esper et al., 2007). Detrending error may be the primary source of uncertainty in the low-frequency domain when the climate variable of interest contains sufficient low-frequency variability as extracting the climate signal from biological systems can be quite complex (Cook et al., 1995; Melvin & Briffa, 2008). The percent common signal (Cropper, 1982), Expressed Population Signal (EPS) and Subsample Signal Strength (SSS; Wigley et al., 1984; Buras, 2017) are all used to gauge chronology error. In the drought reconstruction mentioned above, a bootstrapping technique incorporates chronology error while the standard error of the estimate captures calibration error (Esper et al., 2007). There are several bootstrapping and regression error methods used in the dendroclimatic community, but we were motivated to investigate two particular bootstrapping methods and two calibration error methods based on their description in a recent streamflow reconstruction (Cook et al., 2013). Here we test several combinations of these methods using real and synthetic data to measure performance.

## C.3. Methods

We calculated confidence intervals and tested their performance using 303 chronologies (3 original, 300 synthetic) and their corresponding target climate variables (hereafter “targets”). We use the term chronology to refer to the assemblage of crossdated indices, which is the primary form in which we utilize these data. We use the term mean-value (MV) chronology to discuss the time series of the annual chronology average, calculated by robust biweight mean (Mosteller and Tukey, 1977).

### C.3.1. Chronologies

We calculated and tested a set of annually resolved chronologies, including two tree-ring chronologies, one bivalve sclerochronology, and 300 synthetic chronologies. All real chronologies were developed from crossdated, replicated, annually resolved datasets. For simplicity, we selected chronologies that produced reconstructions of a climate variable by simple linear regression and contained at least 60 years of IOI to ensure sufficient data for independent intervals for calibration, confidence interval calculation, and confidence interval testing. We selected original, and constructed synthetic, chronologies to represent a range of values for sample depth (i.e., the number of time series representing a single year), rbar, first-order autocorrelation (AR1), and correlation to target.

The Tree Nob chronology consists of crossdated, detrended growth increments of Pacific geoduck from coastal British Columbia, Canada, and extends continuously from 1725 to 2008 (Edge et al., 2021a). We used the chronology as produced for the original study, which were detrended by regional-curve standardization and then log transformed. The Tree Nob sea surface temperature (SST) reconstruction targets April-November SST at Langara Island. The length of the IOI is 62 years, from 1940-2001, with a minimum sample depth of 11 measurement series in this interval.

The Rock Springs Ranch chronology is based on the tree-ring widths of Blue Oak in San Benito County, California, USA (Stahle & Griffin, 2012). The chronology spans 1379 to 2003 and is highly sensitive to the local Jan-Feb rainfall (Stahle et al., 2013; Griffin and Anchukaitis, 2014). We downloaded the raw ring widths and detrended individual ring-width series with 2/3-length, 50% frequency cutoff cubic splines. We then compared the MV chronology to monthly precipitation data from the PRISM analysis 4k (Di Luzio et al., 2008). Based on significant correlation values, we selected a target of total (sum) Jan-Feb precipitation for reconstruction. The IOI extends from 1895-2003, 109 years, with an minimum sample depth of 30 measurement series in this interval.

The Arrowsmith Mountain maximum latewood density (MXD) chronology is based on mountain hemlock on Vancouver Island, British Columbia, Canada and extends from 1629 to 1983 (Schweingruber 1988, Schweingruber et al., 1991; Briffa et al., 1992; Wiles et al., 1996; Briffa & Schweingruber, 2002). We combined the raw MXD time series assemblage by robust biweight mean to produce the MV chronology. In order to define a target season for reconstruction, we compared the MXD chronology to monthly air temperature values from the nearest grid box of the HADCRUT5 air temperature dataset (Morice et al., 2021) . Based on significant correlation values, we selected a target of average April-October temperature. Our target is similar to the Apr-Sep target for a reconstruction which utilized this and two additional MXD chronologies in the region (Wiles et al., 1996). The chronology-target overlap extends from 1857-1983, 127 years, with a minimum sample depth of 28 measurement time series in this interval.

### C.3.2 Regression Assumptions

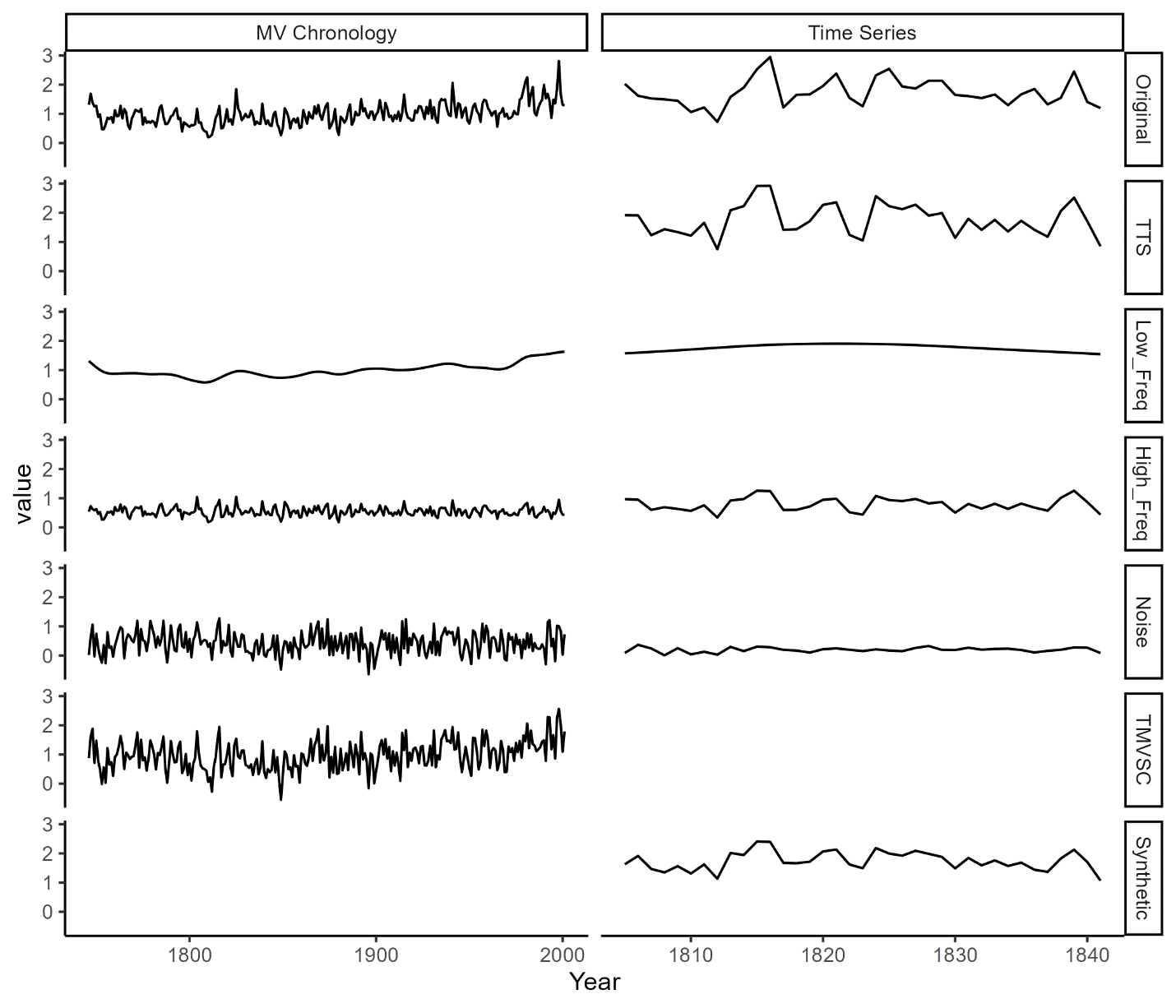
We regressed each chronology onto its target over the full IOI and tested the assumptions of regression (autocorrelation, normality, and homoscedasticity) for the residuals. We utilized the Durbin Watson test of autocorrelation in regression residuals using the lmtest package in r with a threshold of 0.05 based on Pan’s algorithm (Durbin and Watson, 1950; Farebrother, 1980; Zeileis and Hothorn 2002). We tested normality of the residuals with the Shapiro-Wilk test, using the stats package in the R programming language, with p-values less than 0.05 considered to reject the null of normality (R Core Team; 2022). We tested homoscedasticity with the Goldfeld-Quandt test provided by the lmtest package in R, with p-values less than 0.05 considered to reject the null of homoscedasticity (Goldfield and Quandt; 1965; Zeileis and Hothorn 2002).

### C.3.3 Synthetic Chronologies

In addition to the three real chronologies, we developed 100 synthetic chronologies for each chronology-target pair using the R programming language (Fig 1). Synthetic chronologies provided additional opportunities to test the confidence interval methods for each climate target and helped determine which chronology properties had the largest influence on the fidelity of the confidence intervals. We varied key properties of the synthetic chronologies in each iteration, including the correlation to target, first-order autocorrelation, sample depth, and rbar, defined as the mean value of all pairwise correlations among samples. We constrained the sample depths of the three sets of synthetic chronologies to either half, double, or identical to that of the original chronology. We constrained the AR1, rbar, and correlation to the climate target by the character of the original chronology but also added a random component.

We designed an algorithm to build synthetic chronologies based on an original chronology, its climate target, and four chronology properties (Fig C.1). The synthetic chronology algorithm (hereafter, algorithm) built chronologies in two steps, first building a temporary mean value synthetic chronology (TMVSC), then building component time series based on the TMVSC. The component time series of the new synthetic chronology were built based on the TMVSC, but each of the final synthetic MV chronologies varied slightly from the TMVSC due to noise imparted in the construction of the individual time series. Four perturbation parameters, selected randomly when the synthetic chronology algorithm was initiated, adjusted the sample depth, rbar, AR1, and correlation to target. The algorithm randomly selected an rbar adjustment factor, which adjusted the random noise imparted to individual time series. The sample depth was set to change for 20% of synthetic chronologies, causing those chronologies to halve or double in sample depth with equal probability. An AR1 adjustment factor adjusted the variance of the white noise imparted to the TMVSC, changing the AR1 of the resulting chronology. The algorithm adjusted the correlation to the target, a random value ranging from 50% to 130% of the original chronology-target correlation.

Figure C.1 Synthetic chronology construction. The MV Chronology panel shows the process of building a TMVSC from the original MV chronology from the low/high frequency and noise components. One TMVSC is produced for each synthetic chronology. The Time Series panels shows the development of one synthetic chronology time series from the TMVSC segment, where the Original Time Series is the TMCSV in the interval covered by a particular time series. The TTS is broken into the low/high frequency components. Noise is added to the high frequency before the two components are recombined to form the synthetic time series.



The algorithm constructed the TMVSC from three components that included the low- and high-frequency components of the original MV chronology and white noise. The algorithm first deconstructed the original MV chronology into high- and low-frequency components. The MV chronology was fit with a 50% frequency cubic smoothing spline of a length randomly selected from a uniform distribution 1/10-length to the full length of the MV chronology. The residuals, obtained by dividing the original MV chronology by the fitted-spline values served as the high frequency series. Next, the algorithm generated a white noise series with a standard deviation identical to the high frequency series. The algorithm then combined the high frequency and noise series in a proportion randomly assigned, with each component representing 30-70% of the total variance. The AR1 adjustment factor then determined the scaling of this new series relative to the standard deviation of the original high frequency series, ranging from 10% to 300%. The algorithm then added this scaled series to the low frequency series to create the TMVSC. The algorithm repeated the process of building a TMVSC until the desired correlation to the target was reached.

The time series construction component of the algorithm proceeded similarly to the TMVSC algorithm with high/low frequency and noise components. First, a temporary time series (TTS) was defined as the sum of the TMVSC and white noise, with a variance set by the random rbar adjustment factor. The white noise imparted to each TTS was different, but its variance was set by the rbar adjustment factor, thus imparting a similar amount of noise to each. This TTS was deconstructed into high- and low-frequency components with a cubic smoothing spline as described above, with a spline of the same length as the TTS. The algorithm constructed the final time series from the sum of the spline, the residuals (high frequency series), and a new white noise series. The ratio of original high frequency to white noise was calculated as the square root of the rbar of the original indices.

### C.3.4. Bootstrapping

We bootstrapped all chronologies, real and synthetic, using both traditional (Efron, 1979)) and maximum entropy bootstrapping methods (Vinod, 2006; Cook et al., 2013). We performed traditional bootstrapping by resampling one year at a time, from all possible values for that year of the chronology to produce a bootstrapped chronology with sample depths identical to the original chronology.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Original Series** | **Sorting Order** | **Consecutive Distances** | **Sorted Series**  V**alues** | **Intermediate Values** | **Ensemble Distributions** | **Example Bootstrapped Series** |
|  |  |  |  | (0.9-**0.45**)=0.45 |  |  |
| 1.1 | 2 |  | 0.9 |  | (1.0, 1.15) | 1.1137492 |
|  |  | abs(1.1-1.3)=0.2 |  | ((0.9+1.1)/2)=1 |  |  |
| 1.3 | 4 |  | 1.1 |  | (1.25, 1.5) | 1.4529844 |
|  |  | abs(1.3-0.9)=0.4 |  | ((1.1+1.2)/2)=1.15 |  |  |
| 0.9 | 1 |  | 1.2 |  | (0.45, 1) | 0.4946209 |
|  |  | abs(0.9-1.7)=0.8 |  | ((1.2+1.3)/2)=1.25 |  |  |
| 1.7 | 5 |  | 1.3 |  | (1.5, 2.15) | 2.0572953 |
|  |  | abs(1.7-1.2)=0.5 |  | ((1.3+1.7)/2)=1.5 |  |  |
| 1.2 | 3 |  | 1.7 |  | (1.15, 1.25) | 1.2029572 |
|  |  |  |  | (1.7+**0.45**)=2.15 |  |  |
| \*Consecutive Distances: (0.2, 0.4, 0.8, 0.5), sorted: (0.2, 0.4, 0.5, 0.8), trimmed: (0.4, 0.5), trimmed mean: (**0.45**) | | | | | | |

We performed MEboot on each individual timeseries of a chronology using the MEboot package in R (Table C.1.; Vinod and López-de-Lacalle, 2009). The MEboot algorithm creates surrogates based on the distances between ordered time series values such that all MEboot surrogate time series, when ordered by value, will have an identical sorting order to that of the original time series (Vinod, 2006). For instance, a time series 1.1, 1.3, 0.9, 1.7, 1.2 has consecutive distances of 0.2 (1.3-1.1), 0.4 (1.3-0.9), 0.8 (1.7-0.9), 0.5 (1.7-1.2). The 25% trimmed mean, the mean after removing 25% of values at each extreme, of these distances is 0.45. The trimmed mean of the consecutive distances is used to extrapolate the bounds of the ensemble range for the minimum and maximum values of the series, such that the minimum ensemble range for the lowest value in the time series is 0.9-0.45 and the maximum value at the maximum value in the time series is 1.7 + 0.45. All other ranges are given by the averages of the ordered values, so after sorting the time series by value to 0.9, 1.1, 1.2, 1.3, 1.7, we find midpoint values of 1.0, 1.15, 1.25, 1.5. Therefore, all MEboot ensemble members would be drawn from the uniform distributions: (1.0,1.15), (1,25, 1.5), (0.45, 1), (1.5, 2.15), (1.15, 1.25). The table shows that any surrogate created from these distributions with have the same sorting order as the original time series.

Table C.1. MEboot (adapted from Vinod, 2006 Table 1). Maximum entropy bootstrapping development from original time series to uniform distributions. A random draw from the distributions produces an example of a bootstrapped series. “abs” indicates absolute difference.

For each bootstrapping method, we produced 1000 sets of bootstrapped indices from each of the 303 chronologies. We developed a bootstrapped MV chronology for each set of indices based on the annual robust biweight mean. We then retained 5th and 95th, as well as 25th and 75th, percentile values at each year for 90th and 50th percentile chronology confidence intervals from the pool of 1000 MV bootstrapped chronologies for each bootstrapping method.

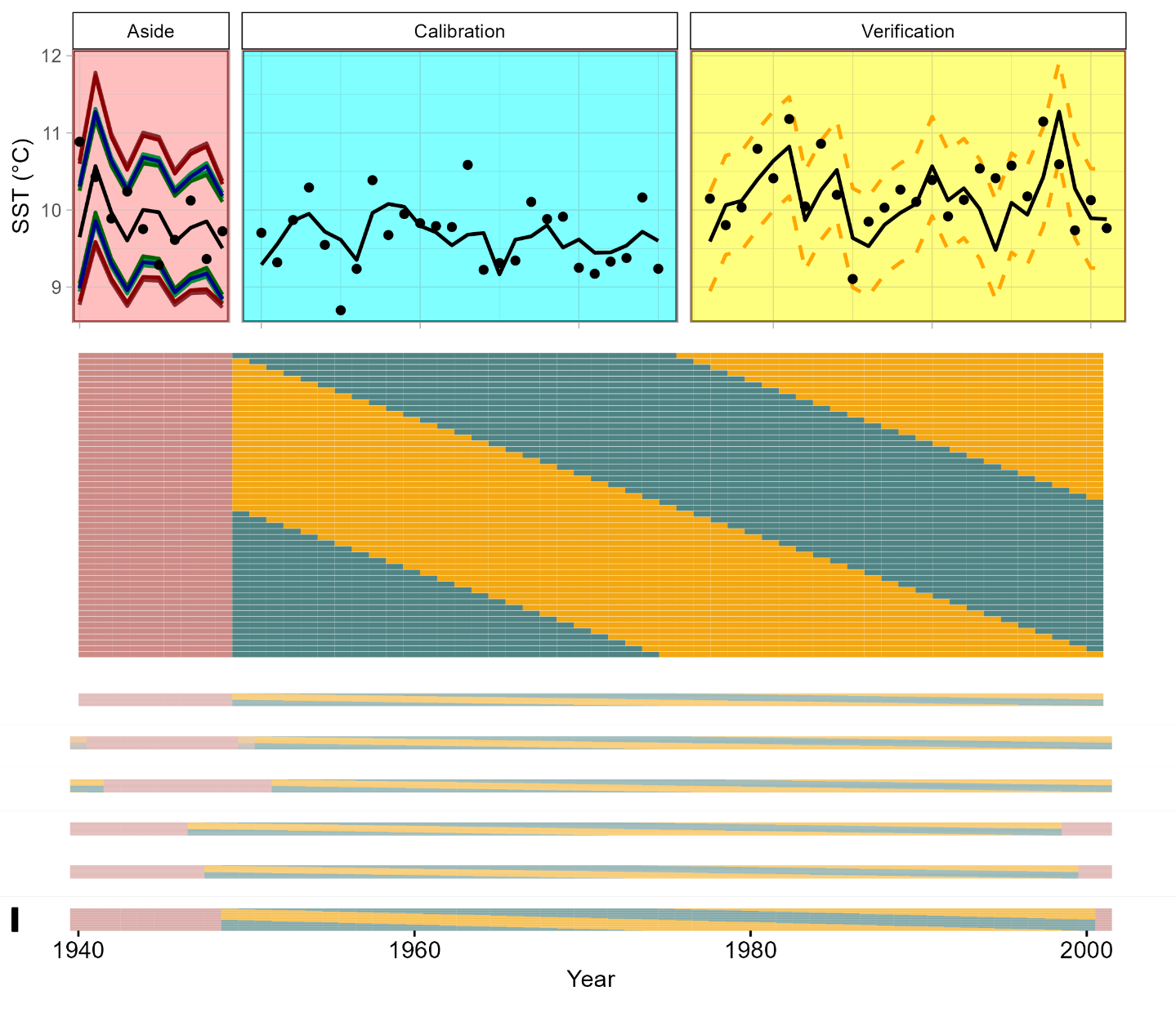
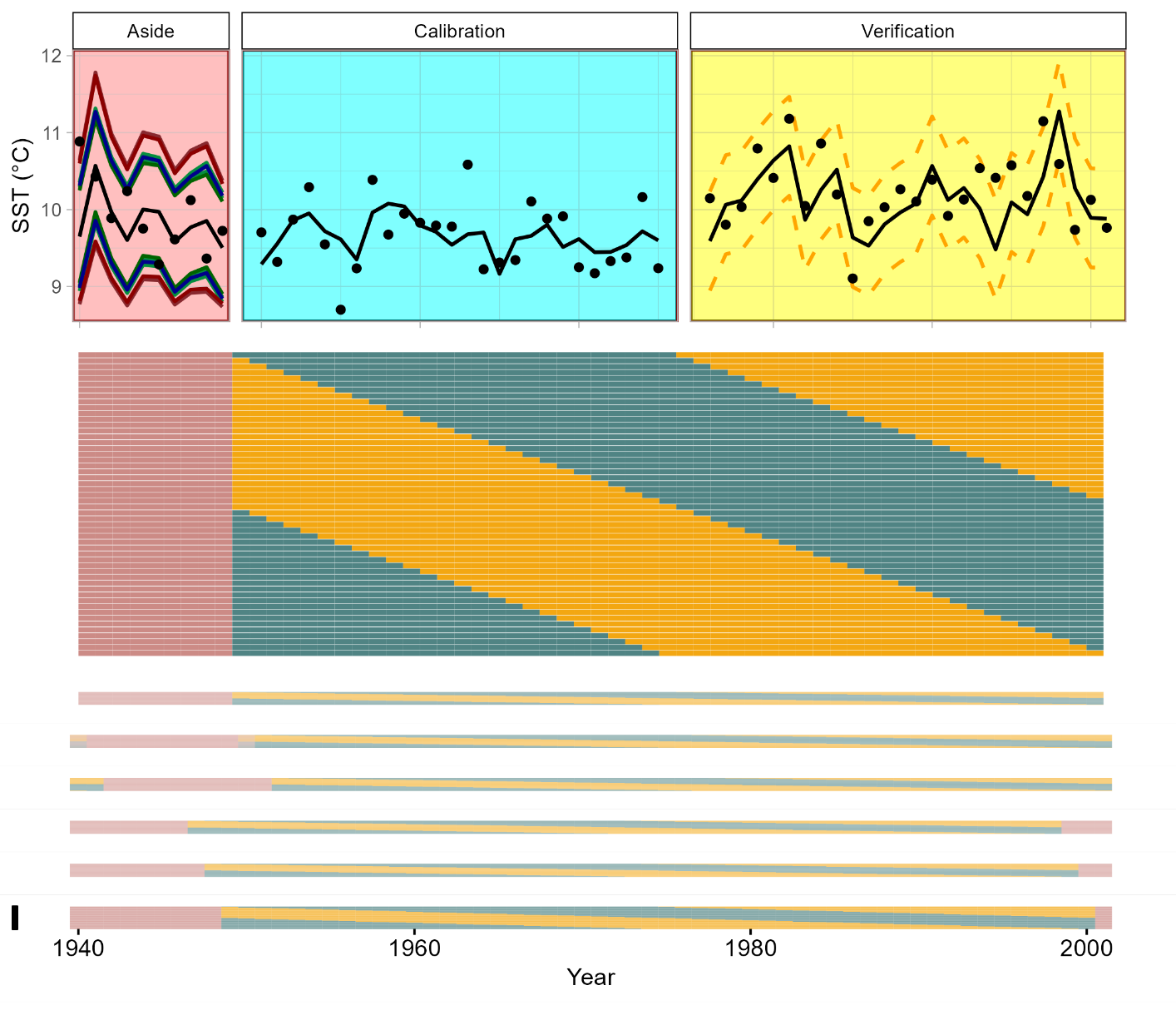
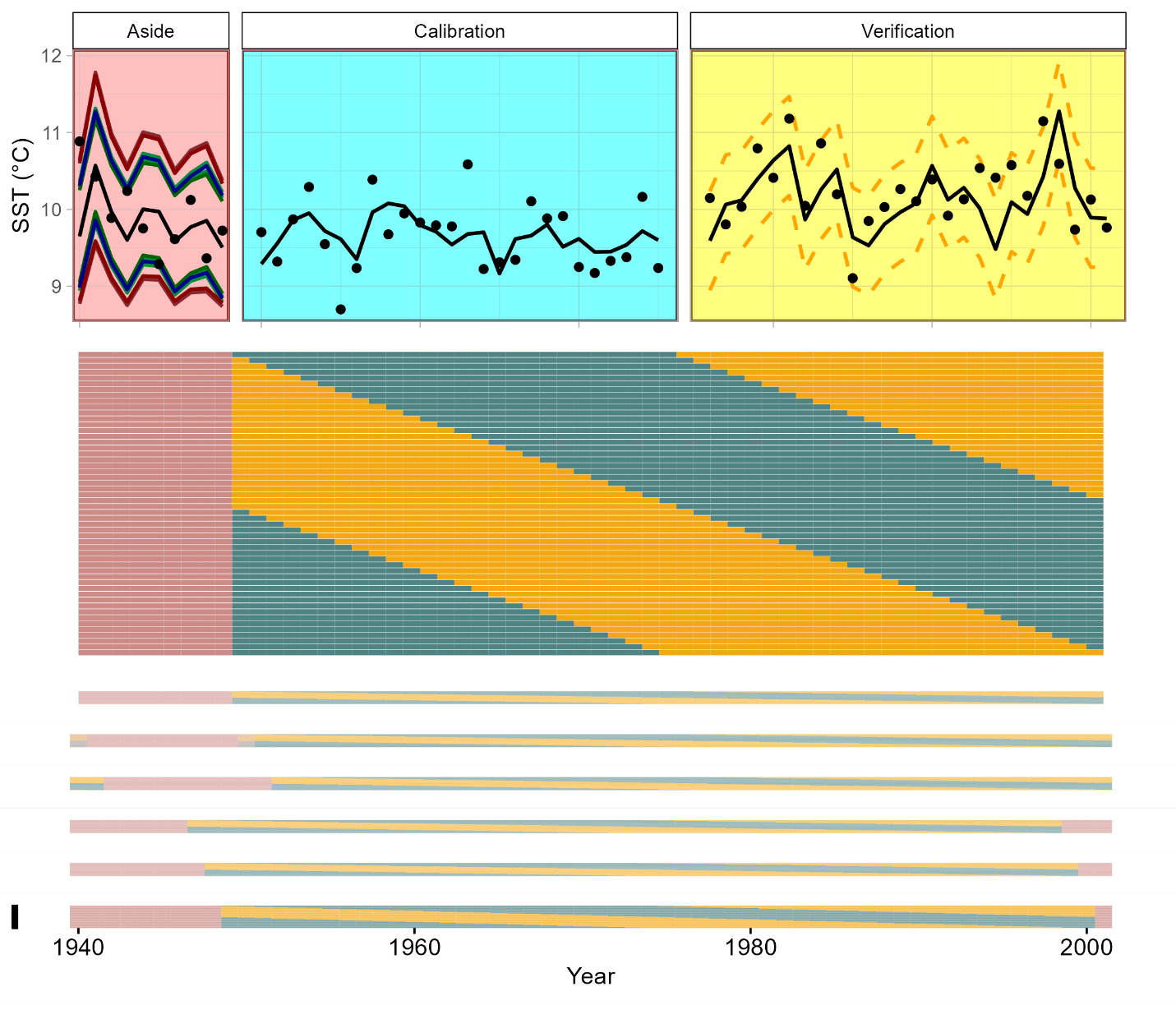
### C.3.5. Confidence Interval Calculation

Prior to calibration of the reconstruction, we first set aside a ten-year interval (named the set-aside interval (SAI)) of the for independent testing of the confidence intervals (Fig C.2.). We repeated the following calculations for each SAI in all possible continuous intervals of the IOI (Fig C.2.c). We used all possible continuous intervals of the remaining period as calibration and verification intervals (CVI), allowing for the calibration and verification intervals to wrap from the last year of the period back to the first year (Fig C.2.b). We then split the CVI in half and used the early portion, named the calibration interval, for calibrating the regression (Fig C.2.a). We also regressed the bootstrapped chronology confidence intervals onto the target in the calibration interval. We used the latter portion, named the verification interval, to calculate the verification error (VE), which is the absolute value of the difference between the reconstructed and target value over the verification interval for each year.

For each CVI, we used the set of VE values to calculate the empirical (non-parametric) 50th and 90th percentile verification intervals (VEe50, VEe90) as the ascending order 50th and 90th percentile values from the VE set. To calculate the theoretical (parametric) VE values, we assumed that the differences are normally distributed, and thus that the VE approximate a half-normal distribution with a lower bound of 0. We calculated theoretical 50th and 90th percentile errors (VEt50, VEt90) after adjusting these intervals for θ of the sample, analogous to the sample standard deviation, using the qhalfnorm function in the R package fdrtool (Strimmer, 2008). The VEt50 is typically indistinguishable from the VEe50, so we do not make a distinction in our tests.

We combined the VEe and VEt values with the regression coefficients from both the MV chronology and bootstrapped chronology confidence intervals to calculate the reconstruction confidence intervals in the SAI. We first built a new reconstruction for the SAI by regressing the MV chronology onto the target in the full CVI. We calculated non-bootstrapped confidence intervals by adding the VEe and VEt values to this new reconstruction. We calculated the bootstrapped reconstruction confidence intervals by the same process, substituting the regression coefficients from the bootstrapped chronology confidence intervals. The possible permutations of these bootstrapping (None, MEboot, Traditional) and VE (VEe50, VEt50, VEt90, VEe90) options produce 12 sets of confidence intervals; but with theoretical and empirical VE identical for the 50% intended capture, we produced only 9 unique confidence intervals per SAI.

Figure C.2. Confidence Interval Testing: Intervals and Methods. Panel a shows the calculations for one CVI and the testing of one SAI. Panel b shows the relationship between CVI calculations and a single SAI. Panel c shows six of the 62 SAIs for this example. a) **Calibration**: black line: target values regressed onto chronology, black points: target values. This regression defines the reconstruction for calibration and verification. **Verification**: orange dashed line: 90-percentile empirical error, black points: target values. VE measured as absolute difference between black line and black points. **Aside**: black line: reconstruction based on regression in full CVI. Colored lines: 90th-percentile empirical and theoretical confidence intervals, green: no bootstrapping. blue: MEboot, red: traditional bootstrapping. CIC is measured for each method as the percentage of target points falling within the confidence intervals. b) Pink box: Aside interval as in panel a. Blue/yellow lines: calibration/verification intervals, 52 of each in this example. Average VEE and VET from the 52 possible verification intervals used to build confidence intervals in a single SAI. c) SAI and CVI, the first three and last three of 62 (for this example). The mean CIC (CICm) across all SAI for each confidence interval method was used as the CICm value for the chronology.



**a**

**b**

**c**

### C.3.6. Confidence Interval Testing

For each of the 9 confidence intervals per SAI, we calculated the confidence interval capture (CIC), the percentage of target values that fall within a set of confidence intervals in a given period. We calculated CIC in each SAI by dividing the total number of target values captured by the confidence intervals by 10 (the SAI length). For example, if 9 of the 10 SAI climate target values fall within the confidence intervals, 90% of the target values were “captured”, in line with the 90% intended capture (IC) rate. This test was repeated for all possible continuous 10-year intervals, such that a chronology-target pair with a 62-year IOI contains 62 possible, overlapping testing intervals (Fig 2c). We then calculated the mean CIC from all trials (CICm) for a given chronology and confidence-interval method.

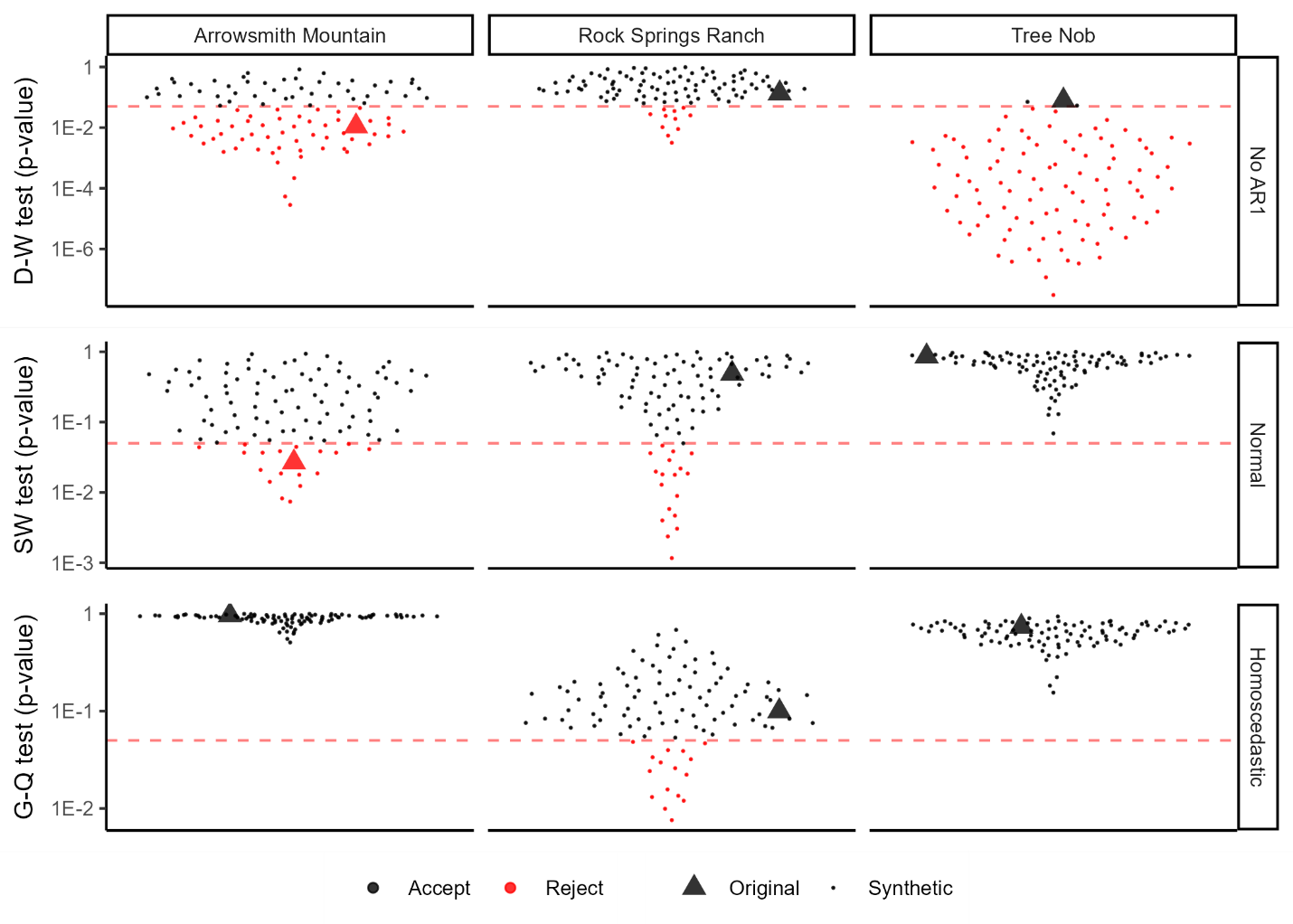
We then compared CICm for the four chronology properties (correlation to climate target, first-order autocorrelation, sample depth, and rbar) for all 303 chronologies. Strong correlations between chronology properties and CIC for a given confidence interval method suggest a bias of the method that will produce confidence intervals too wide or narrow based on idiosyncrasies of individual chronologies.

## C.4. Results

### C.4.1. Regression Assumptions

Of the 300 synthetic and three original chronologies, 200 synthetic and one original rejected the null hypothesis of one or more of the regression assumptions tested (Fig C.3., red points). The regression residuals from the Arrowsmith Mountain chronology and 54 of 100 of the associated synthetic chronologies rejected the null of no autocorrelation. 19 of the Arrowsmith Mountain chronologies had regression residuals that were not normally distributed. Among the Rock Springs Ranch synthetic chronologies, 11, 16, and 18 rejected the null hypotheses of No AR1, homoscedasticity, and normality respectively. At Tree Nob, only the original and two synthetic chronologies failed to reject the null hypothesis of no AR1.

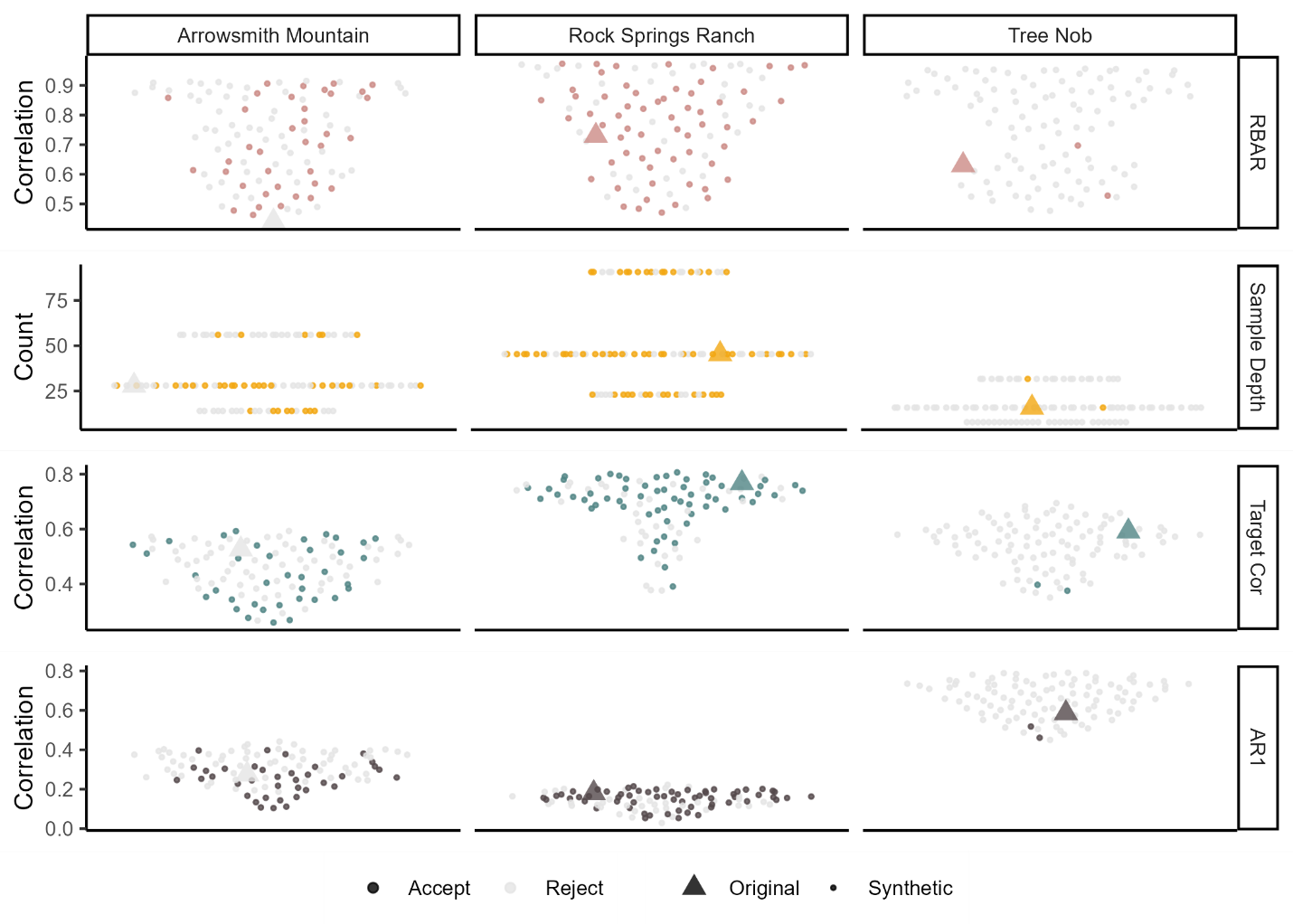
Figure C.3. Regression Assumptions Testing. **No AR1:** results of the Durban Watson test; null hypothesis of no AR1 in residuals rejected at p<0.05. **Normality:** normal distribution of residuals tested by the Shapiro-Wilk test; null hypothesis of normal residuals rejected at p<0.05. **Homoscedasticity**: trend in variance of residuals as tested by the Goldfeld-Quandt test; null hypothesis of homoscedastic residuals rejected at p<0.05. Red points indicate chronologies rejecting a given null hypothesis by the standards listed above. Testing thresholds are demarked by red, dashed lines.



### C.4.2. Synthetic Chronologies

The three sets of synthetic chronologies varied considerably in the range and distribution of the four characteristics parameterized (Fig C.4.), but the rbar, sample depth, AR1, and correlation to target of each of the three original chronologies fell within the spread of the associated synthetic chronologies. The rbar distributions ranged from less that 0.5 to greater than 0.9 for all three synthetic chronology sets. The AR1 distributions were more constricted by the AR1 of the original chronologies with some Tree Nob synthetic chronologies nearing an AR1 of 0.8, while no Rock Springs Ranch synthetic chronologies rose above 0.3. The upper bounds of the target correlation distributions were constrained by that of the original chronology, such that the original chronologies sat near the tops of the three distributions. The chronologies which rejected the null hypothesis of one of the three regression assumptions were not generally of distinct distributions for any of the chronology properties. However, the three chronologies remaining at Tree Nob sat near the lower portion of the AR1 and rbar distributions.

Figure C.4. Chronology Properties. 101 points represent the 100 synthetic and one original chronologies for each cluster of points which are jittered about the x-axis for visibility. Points shown in grey represent chronologies which rejected the null hypothesis of one of the three regression assumptions, see Fig C.3. **RBAR**: average pairwise correlation between individual chronology series. **AR1**: first-order autocorrelation of the MV chronology. **Target Cor**: correlation between chronology and climate target over the full interval of overlap (IOI). **Sample Depth**: Average number of series in the chronology during the interval of overlap with the target. Note that all sample depth values for a single chronology site fall on one of three values, the original chronology sample depth, or half or double that value. Small circles represent values for synthetic chronologies, large triangles represent values of the original chronologies.



### C.4.3 Bootstrapping

Figure C.5. Confidence Interval Performance. Each point represents the CICm of a chronology. Each color represents a chronology site with associated original and synthetic chronologies, with 12 violin plots corresponding to the 12 confidence interval methods. Each grouping of three plots represents the CICm results of one confidence interval method. Columns distinguish bootstrapping methods: None, MEboot, or Traditional. Rows distinguish error measurements: Ninety, Empirical (VEE90), Ninety, Theoretical (VET90), Fifty, Empirical (VEE50), Fifty, Theoretical (VET50).

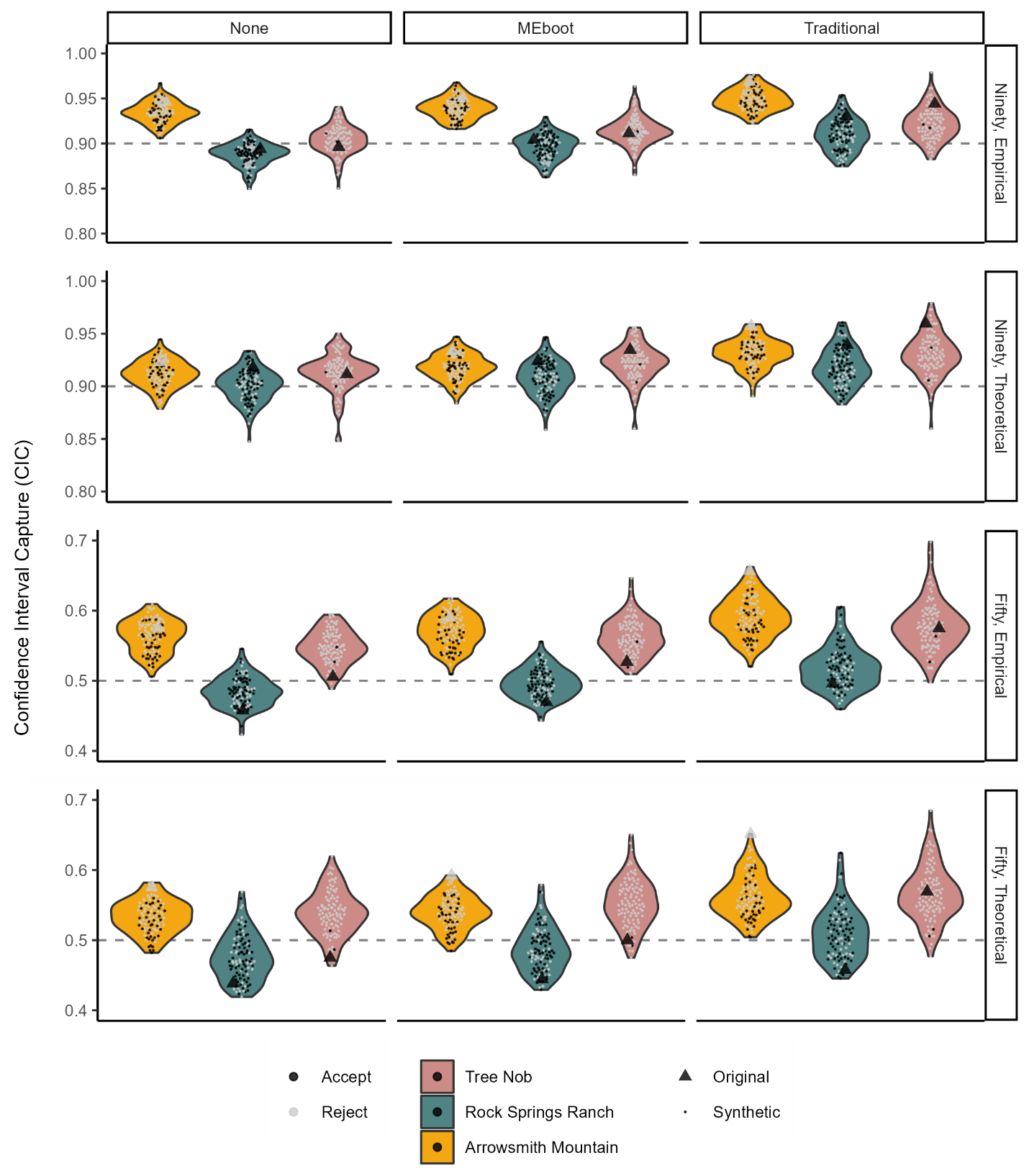
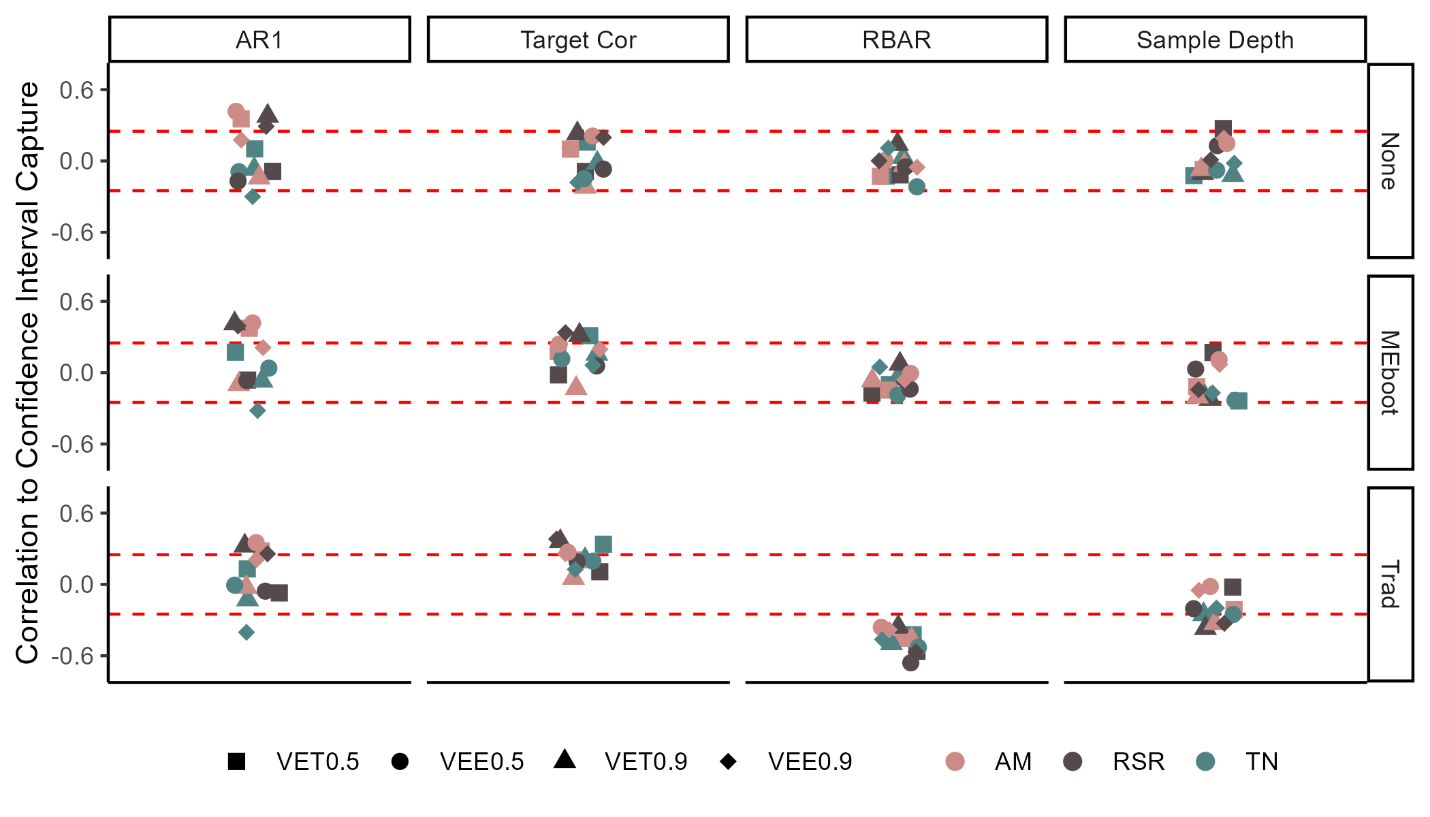


Figure C.6. Correlations between Chronology Properties and Confidence Interval Capture. Individual points represent a correlation between the rate of capture for a confidence interval method at a particular site (CICm) and the chronology properties of the chronologies utilized (n=101 each). The red dashed lines at r=±0.25 represent a threshold for relationships of interest.



The width of traditionally bootstrapped intervals was consistently greater than the MEboot intervals at both 50- and 90-percetile (Fig C.5.). Both techniques produced variable chronology error through time, unequal positive and negative errors at any given year, and generally larger error ranges at more extreme chronology values.

### C.4.4 Confidence Interval Testing

The various confidence interval methods showed variable success in capturing the IC (e.g., 0.9 for 90th percentile confidence intervals; see dashed lines, Fig C.5.). Of the 12 methods implemented, the Non-bootstrapped VEE 90th percentile confidence intervals produced an average CICm closest to the IC at 90.6%, with and interquartile range of 3.7%. If only the chronologies that failed to reject the null hypotheses of all regression assumptions are considered (ignoring the grey points in Fig C.5.), the Meboot VET 50th percentile confidence interval performed best, with an average CICm of 49.9%, IQR = 4.6%. The CICm values for Arrowsmith Mountain chronologies were generally higher than the IC and higher than the other two sites across all methods. The difference between the average CICm values across sites is reduced with VET as opposed to VEE, while the range of CICm values for a given site is reduced with VEE relative to VET. The 90th percentile theoretical errors (VEt90) produced confidence intervals with CICm slightly above the IC, but the CICm distributions were consistent across the three sites. The combination of MEboot, empirical reconstruction errors, and 90-percentile confidence intervals was most effective at minimizing CICm - IC for Tree Nob and Rock Springs Ranch chronologies and produced the tightest grouping of CICm (Fig C.5.). CICm was successively wider from no bootstrapping, to MEboot, to traditional bootstrapping, with traditional bootstrapping CICm generally greater than IC. Individual CICm distributions were generally wider for traditional bootstrapping compared with no bootstrapping and MEboot.

The CICm for traditionally bootstrapped chronologies was negatively correlated to the rbar of all chronologies for all methods (Fig C.6.). the correlation to target and sample depth values were positively and negatively correlated to CIC, respectively, for some methods and chronology properties. The CICm, based on methods without bootstrapping or by MEboot, show little or no consistent relationship to any chronology property.

## C.5. Discussion

Although the chronology simulation algorithm produced chronologies with a wide range of properties and affected confidence interval width, the methods of bootstrapping and error measurement had the most significant impact on the reliability of the confidence intervals. Traditional bootstrapping produces excessively wide confidence intervals, and this is due to the over-sensitivity of this technique to outlier values with biases that are especially pronounced in datasets with low rbar values (Fig C.6.). The MEboot method assures greater fidelity to the properties of the original series (Cook et al. 2013). MEboot preserves autocorrelation structure while reliably producing median, as well as minimum and maximum, values of the dataset. This is distinct from traditional bootstrapping, which shows a bias toward extreme values when selecting the 5- and 95-percentile representatives (Cook et al. 2013).

We designed the chronology simulation algorithm to vary important chronology properties in order to measure their impacts on the reliability of the confidence intervals. The resulting synthetic chronologies are based on the original chronologies and are not independent. The CIC distributions of each site, for a given method, are distinctive. However, the trends across methods are also noteworthy (Fig C.5.). The intended function of the simulations is demonstrated in Figure 6 wherein the variability of an important chronology property, rbar, produces biased confidence intervals when traditional bootstrapping is used. The consistency of this result across the three sites and the strength of the correlation suggest the bias is inherent to the method. The direction of the bias is also consistent with a conceptual understanding of the method wherein the 90th-percentile values in the bootstrapped ensemble necessarily overrepresent chronology indices with more extreme values. In a chronology with low rbar, the distribution of values in a typical year is, by definition, wider, exacerbating the problem inherent to the method. Conversely, a chronology with a high rbar, in which index values of a typical year are tightly grouped, traditional bootstrapping produces 90th-percentile intervals much closer to the central tendency of all values.

Some of the CICm distributions which best relate to their respective IC are not bootstrapped. When considering sources of reconstruction uncertainty, we made a distinction between chronology error and calibration error. We intended to capture chronology error with bootstrapping and calibration error with either VEE or VET. The methods for calculating these errors were chosen based on the principle that calibration ought to be independent from the measurement of skill, such as in the calculation of the CE metric. Our testing method suggests that the VEE and VET alone produce sufficiently wide confidence intervals without the benefit of bootstrapping. However, our testing method does not account for substantial changes in sample depth and makeup within a chronology. It is notable, therefore, that when the Arrowsmith Mountain chronology, which contains no change in sample makeup for the entire IOI, is removed from the analysis, MEboot with VEE produces the best outcomes (average CICm = 89.9%, 49.7%; IQR = 1.8%, 2.7%).

Constructing confidence intervals requires independent time periods in which to calibrate the reconstruction and measure the reconstruction errors. Because many climate proxies derived from crossdated archives contain first-order autocorrelation, consecutive values are not independent. For this purpose, the SAIs could either comprise years chosen at random to incorporate the greatest diversity of possible values or could only be continuous in order to maintain the persistence structure. The random selection option has the advantage of nearly infinite possible calibration intervals, particularly if resampling is permitted, while a continuous selection limits the number of possibilities. We chose an intermediate solution, wherein all calibration and verification intervals are continuous with the exception of up to two breaks, and SAIs are continuous with the exception of up to one break (Fig C.2.b,c). This allows for a large number of intervals, such that outliers, which are uncommon but can have considerable leverage, are unlikely to produce spurious results (Rousseeuw & Leroy, 2005). This also allows for testing a calibration interval in a manner analogous to the final reconstruction, in which the full IOI is utilized, by maintaining independence of calibration and verification intervals while also accounting for persistence.

The autocorrelation of many of the chronologies also reduced the independence of the reconstruction and confidence intervals in the SAI, which reduces the validity as a test of confidence interval performance when the associated targets also have high AR1. We measured the importance of this phenomenon by comparing confidence interval performance to MV chronology AR1 (Figure C.6.) and found no consistent relationship, suggesting that the independence of the SAI is not significantly impacted by autocorrelation. At Tree Nob, the site with the highest AR1 values, CIC and AR1 are negatively correlated, but this relationship is not consistent across confidence interval methods.

We have provided evidence for the unreliability of confidence intervals constructed from traditional bootstrapping and some evidence to support MEboot. However, the simple linear regression method we used to develop the 303 reconstructions utilized is an uncommon approach in the development of climate reconstructions. Multiple predictors and distributed lag regression models as well as principal components are commonly employed (National Research Council, 2006). Nevertheless, we have utilized a diversity of real chronology-target pairs with a range of properties alongside synthetic chronologies with properties of still greater diversity. The approach we have employed can easily be extended to more complex reconstruction methods. Although the robustness of the methods that performed well in these simulations may not hold for all reconstruction methods, the simulations suggest that some methods, including traditional bootstrapping, should not be used. Traditional bootstrapping contains assumptions inappropriate for pairing with crossdated archives. This method consistently produced confidence intervals wider than intended in our trials (Fig C.5.). Furthermore, the consistent negative correlation between CIC and rbar for traditionally bootstrapped chronologies (Fig C.6.) suggests consistent, predictable bias inherent to this method.

We did not exhaustively test all possible bootstrapping methods. Resampling complete individual time series would also maintain the autocorrelation structures of the series while varying the subsets of the sampled population. We did not test this technique here because measurement time series may not overlap to a high degree in tree-ring datasets. Moreover, MEboot provides the benefit of faithfully reproducing data structure while maintaining sample depth characteristics and avoiding issues related to resampling a low number of measurement time series in poorly replicated portions of chronologies.

## C.6. Conclusion

A reliable estimated error range inherent to any paleoclimate reconstruction is essential to the end-user. Confidence intervals provide at-a-glance information of the relative signal-to-noise ratio of the reconstruction, conveying both what is known and unknown. Confidence intervals are best calculated by measuring errors outside the period of calibration, as the errors of interest also lie outside the calibration period. The exact methodology chosen for producing confidence intervals will impact both the precision and accuracy of the results, and the best methods will likely conform to the method’s appropriateness in relation to the dataset. We find that for many proxies derived from crossdated archives, MEboot is likely to be the most appropriate bootstrapping method. Additionally, utilization of empirical errors rather than assuming a normal distribution of errors may be prudent in many circumstances. While future work is needed to extend this approach to reconstructions with multiple predictors and other complex reconstruction methods, our results suggest that MEboot paired with VEE or VET provide the most robust confidence intervals for crossdated paleoclimate proxies. Furthermore, we have introduced a new approach to synthetic chronology construction will allow for the testing of many methods with similar simulation experiments.

## C.7. References

Ahad, N. A., Yin, T. S., Othman, A. R., & Yaacob, C. R. (2011). Sensitivity of normality tests to non-normal data. *Sains Malaysiana*, *40*(6), 637-641.

Briffa, K., Jones, P., Pilcher, J., & Hughes, M. (1988). Reconstructing summer temperatures in northern Fennoscandinavia back to AD 1700 using tree-ring data from Scots pine. Arctic and Alpine Research, 20(4), 385-394.

Briffa, K. R., Jones, P. D., & Schweingruber, F. H. (1992). Tree-ring density reconstructions of summer temperature patterns across western North America since 1600. *Journal of Climate*, *5*(7), 735-754.

Briffa, K.R.; Schweingruber, F.H. (2002). NOAA/WDS Paleoclimatology - Briffa - Arrowsmith Mountain - TSME - ITRDB CANA113. NOAA National Centers for Environmental Information. https://doi.org/10.25921/r7fr-m938. Accessed 21 October 2022.

Buras, A. (2017). A comment on the expressed population signal. Dendrochronologia, 44, 130-132.

Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A., & Funkhouser, G. (1995). The'segment length curse'in long tree-ring chronology development for palaeoclimatic studies. The Holocene, 5(2), 229-237.

Cook, E. R., & Kairiukstis, L. A. (1990). *Methods of dendrochronology: applications in the environmental sciences*. Springer Science & Business Media.

Cook, E. R., Palmer, J. G., Ahmed, M., Woodhouse, C. A., Fenwick, P., Zafar, M. U., Wahab, M., & Khan, N. (2013). Five centuries of Upper Indus River flow from tree rings. *Journal of hydrology*, *486*, 365-375.

Di Luzio, M., Johnson, G. L., Daly, C., Eischeid, J. K., & Arnold, J. G. (2008). Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. Journal of Applied Meteorology and Climatology, 47(2), 475-497.

Draper, N. R., & Smith, H. (1998). Applied regression analysis (Vol. 326). John Wiley & Sons.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. Biometrika, 37(3/4), 409-428.

Edge, D. C., Reynolds, D. J., Wanamaker, A. D., Griffin, D., Bureau, D., Outridge, C., Stevick, B. C., Weng, R., & Black, B. A. (2021). A Multicentennial Proxy Record of Northeast Pacific Sea Surface Temperatures From the Annual Growth Increments of Panopea generosa. *Paleoceanography and Paleoclimatology*, *36*(9), e2021PA004291.

Edge, D.; Reynolds, D.J.; Wanamaker, A.D.; Griffin, D.; Bureau, D.; Outridge, C.; Stevick, B.C.; Weng, R.; Black, B.A. (2021). NOAA/WDS Paleoclimatology - British Columbia 2,900 Year Bivalve Sclerochronology and Sea Surface Temperature Reconstruction. NOAA National Centers for Environmental Information. https://doi.org/10.25921/ms94-wy29. Accessed 21 October 2022.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569-593). Springer.

Farebrother, R. W. (1980). Algorithm AS 153: Pan's procedure for the tail probabilities of the Durbin-Watson statistic. Journal of the Royal Statistical Society. Series C (Applied Statistics), 29(2), 224-227.

Fritts, H. (1976). Tree rings and climate. Academic Press. New York.

Goldfeld, S. M., & Quandt, R. E. (1965). Some tests for homoscedasticity. Journal of the American statistical Association, 60(310), 539-547.

Griffin, D., & Anchukaitis, K. J. (2014). How unusual is the 2012–2014 California drought? *Geophysical Research Letters*, *41*(24), 9017-9023.

Melvin, T. M., & Briffa, K. R. (2008). A “signal-free” approach to dendroclimatic standardisation. Dendrochronologia, 26(2), 71-86.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., ... & Simpson, I. R. (2021). An updated assessment of near‐surface temperature change from 1850: the HadCRUT5 data set. Journal of Geophysical Research: Atmospheres, 126(3), e2019JD032361.

Mosteller, F., & Tukey, J. W. (1977). Data analysis and regression. A second course in statistics. Addison-Wesley series in behavioral science: quantitative methods.

National Research Council. (2006). Surface temperature reconstructions for the last 2,000 years. National Academies Press.

Rousseeuw, P. J., & Leroy, A. M. (2005). Robust regression and outlier detection. John wiley & sons.

Schweingruber, F. (1988). A new dendroclimatic network for western North America. *Dendrochronologia*, *6*, 171-180.

Schweingruber, F. H., Briffa, K. R., & Jones, P. (1991). Yearly maps of summer temperatures in Western Europe from AD 1750 to 1975 and Western North America from 1600 to 1982. *Vegetatio*, *92*(1), 5-71.

Stahle, D.W.; Griffin, R.D. (2012). NOAA/WDS Paleoclimatology - Stahle - Rock Springs Ranch - QUDG - ITRDB CA646. NOAA National Centers for Environmental Information. https://doi.org/10.25921/mz3p-1721. Accessed 21 October 2022.

Stahle, D., Griffin, R., Meko, D., Therrell, M., Edmondson, J., Cleaveland, M., Stahle, L., Burnette, D., Abatzoglou, J., & Redmond, K. (2013). The ancient blue oak woodlands of California: Longevity and hydroclimatic history. *Earth Interactions*, *17*(12), 1-23.

Strimmer, K. (2008). fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics, 24(12), 1461-1462.

Team, R. C. (2013). R: A language and environment for statistical computing.

Trouet, V., & Van Oldenborgh, G. J. (2013). KNMI Climate Explorer: a web-based research tool for high-resolution paleoclimatology. *Tree-Ring Research*, *69*(1), 3-13.

Vinod, H. D. (2006). Maximum entropy ensembles for time series inference in economics. *Journal of Asian Economics*, *17*(6), 955-978.

Vinod, H. D., & López-de-Lacalle, J. (2009). Maximum entropy bootstrap for time series: the meboot R package. *Journal of Statistical Software*, *29*, 1-19.

Wigley, T. M., Briffa, K. R., & Jones, P. D. (1984). On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. Journal of Applied Meteorology and Climatology, 23(2), 201-213.

Wiles, G. C., D'Arrigo, R. D., & Jacoby, G. C. (1996). Temperature changes along the Gulf of Alaska and the Pacific Northwest coast modeled from coastal tree rings. *Canadian Journal of Forest Research*, *26*(3), 474-481.

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships.