Homogenization of Daily Temperature Data

Anuradha P. Hewaarachchi

Department of Statistics and Computer Science, University of Kelaniya, Kelaniya, Sri Lanka

Yingbo Li

Department of Mathematical Sciences, Clemson University, Clemson, South Carolina, and
Department of Statistical Science, Southern Methodist University, Dallas, Texas

Robert Lund

Department of Mathematical Sciences, Clemson University, Clemson, South Carolina

Jared Rennie

Cooperative Institute for Climate and Satellites—North Carolina, North Carolina State University, Raleigh, and
NOAA/National Center for Environmental Information, Asheville, North Carolina

ABSTRACT

This paper develops a method for homogenizing daily temperature series. While daily temperatures are statistically more complex than annual or monthly temperatures, techniques and computational methods have been accumulating that can now model and analyze all salient statistical characteristics of daily temperature series. The goal here is to combine these techniques in an efficient manner for multiple changepoint identification in daily series; computational speed is critical as a century of daily data has over 36 500 data points. The method developed here takes into account 1) metadata, 2) reference series, 3) seasonal cycles, and 4) autocorrelation. Autocorrelation is especially important: ignoring it can degrade changepoint techniques, and sample autocorrelations of day-to-day temperature anomalies are often as large as 0.7. While daily homogenization is not conducted as commonly as monthly or annual homogenization, daily analyses provide greater detection precision as they are roughly 30 times as long as monthly records. For example, it is relatively easy to detect two changepoints less than two years apart with daily data, but virtually impossible to flag these in corresponding annually averaged data. The developed methods are shown to work in simulation studies and applied in the analysis of 46 years of daily temperatures from South Haven, Michigan.

1. Introduction

Climate time series often exhibit artificial discontinuities induced by station relocations, gauge changes, observer changes, and so on. Such changes may impart statistical discontinuities in associated data and are called changepoints (or breakpoints, or mean shifts). Mitchell (1953) estimates that U.S. temperature series experience about six breakpoints per century on average. Some, but not necessarily all, of these times induce mean shifts in the series. While the times of some gauge changes, station relocations, and other events are documented in station history logs (called metadata), these records are notoriously incomplete, and many breakpoint times are undocumented.

This paper seeks to identify all changepoint times in a daily temperature record while accounting for four critical aspects: metadata, a reference series, a seasonal cycle, and autocorrelation. While Li and Lund (2015) and Li et al. (2016) consider these features in annual and monthly series, this paper modifies the methods to accommodate the more complex features seen in daily data. Analyses of a single daily series by some existing methods may take days of computation time as a century of daily data has over 36 500 entries. Our methods are illustrated on single series only; homogenization of a temperature series network or comparison to other homogenization

Corresponding author e-mail: Robert Lund, lund@clemson.edu