TemperatureImputations

Maxime Rischard

December 7, 2016

## Contents

1 Introduction

- explain the problem we're trying to solve

1.1 bias in recorded daily minima and maxima induced by time of measurement

- explain
- demonstrate using hourly temperatures from one station: reduce to daily min and max and show difference as a function of measurement hour
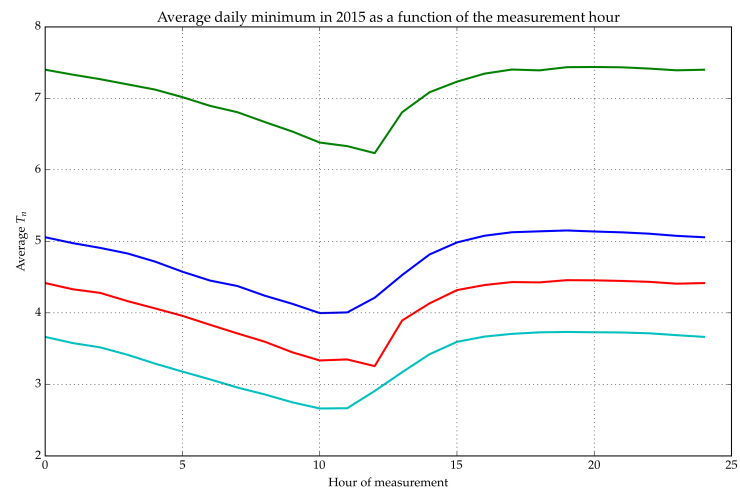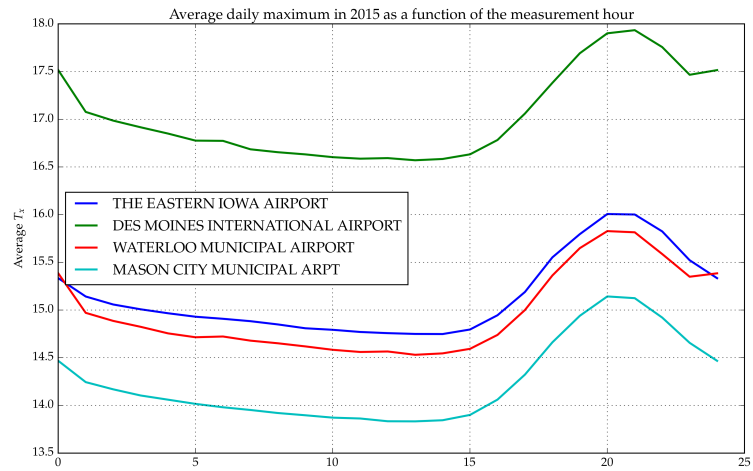
We illustrate the measurement bias in daily maxima and minima with ten days of hourly temperature measurements from the Waterloo Municipal Airport station in Iowa. Ideally, T_x measurements should capture the peak of each diurnal cycle, and T_n its trough. In Figure X, those ideal measurements are indicated by the red and blue triangles respectively. The actual measurements are obtained by dividing the data into 24 hour measurement windows, and extracting the minimum and maximum. For each window, we plot these extrema with a red and blue horizontal line.
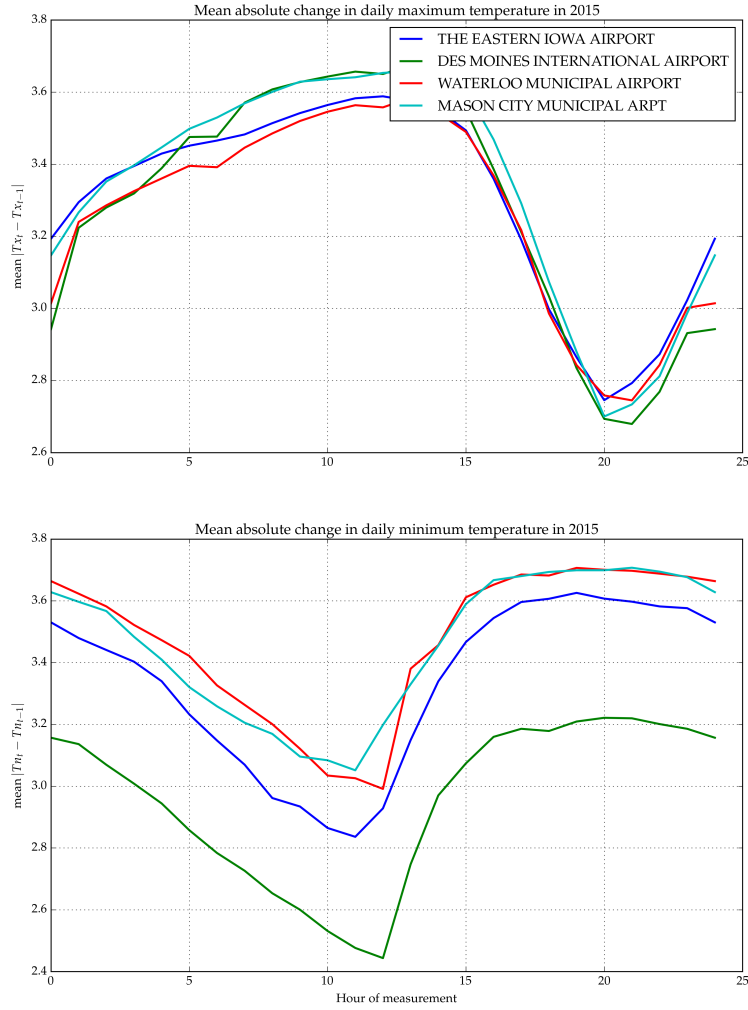
Most of the time, the ideal measurement and the actual measurement coincide: the triangle is on that day's line. But there are also several misses. The most blatant example occurs on April 3rd, where the peak of the diurnal cycle is 7.2°C and occurs at 21:00 UTC. However, because the previous day was much warmer, the day's T_x record of 15.6°C is reached immediately after the previous day's measurement. The measured T_x therefore overestimates the diurnal cycle's peak by 8.4°C.
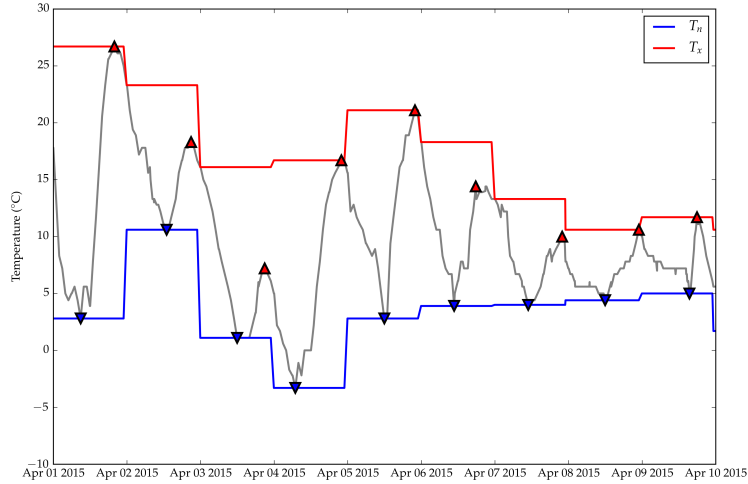
This subtle bias in the daily records can in turn bias long-term summary statistics that are of climatological interest. A measure as simple as the average daily maximum temperature for an entire year (2015) increases by over 1°C if the measurements are made at the warmest time of day 21:00 UTC rather than 14:00 UTC (see Figure X). Conversely, the average T_n is colder by over 1°C if T_n is measured at 10:00 UTC (the coldest time of day on average) rather than 17:00 UTC.

A climatologist studying weather variability might be interested in summary statistics such as the average absolute change in the daily temperature maxima and minima from one day to the next. The answer to that question too depends on the time of day at which the temperatures are recorded. Collecting the measurements at the hottest time of day means that the peaks on a warm day gets recorded twice, erasing the diurnal peaks of the following colder day, and hence the variability gets underestimated. We can see this in Figure X, where the respective variability estimates drop if the maxima get measured at the warmest time, or if the minima get measured at the coldest time.

Average daily maximum in 2015 as a function of the measurement hour

Average daily minimum in 2015 as a function of the measurement hour

Mean absolute change in daily maximum temperature in 2015

Mean absolute change in daily minimum temperature in 2015

2   Model Specification

1. start with a simple SE kernel in both space and time

$$k_{st}(s, s', t, t') = k_{time}(t, t') \times k_{space}(s, s') + k_{mean}(s, s')$$

2. observation model

   - not much to say… just clipping

3   Implementation

1. timeseries model

   - fitting hyperparameters
   - chunks
   - show variogram

2. spatiotemporal model

4

- fitting hyperparameters
- chunks

3. imputations

- Stan
- softmin and softmax
- observation noise
- reparametrization

4 Improving model

1. focused on timeseries model

- kernel components
- diurnal cycle
- show improved variograms

2. spatiotemporal model

- variograms and cross-variograms
- trace evolution
  - product kernel
  - sum of products with variance 1
  - sum of products with free variance
- for each model, report marginal likelihood, and predictive diagnostic in a table
- discuss importance of getting uncertainty right

5 Analysis

- show imputations on interesting days
- show imputations can capture two possible explanations for a measurement
- discuss possibility of inferring measurement time