Change point detection in monthly mean air temperature observations in Latvia

Svetlana Aniskevich[1], Viesturs Zandersons[2]
[1]svetlana.aniskevica@lvgmc.lv, [2]viesturs.zandersons@lvgmc.lv
Latvian Environment, Geology and Meteorology Centre, Riga, Latvia

Introduction

Nowadays with increasing length of meteorological observation time series it is possible to comprehensively monitor climate change, therefore it is crucial to provide qualitative and homogeneous data sets. With swiftly developing methods, sensors and surrounding conditions it is more challenging to maintain comparable observations without significant shifts due to nonnatural causes. Data homogenization procedure is meant to overcome such problems by using various statistical tools.

Homogenization process consists of two main stages - identification of change points and observation data correction. In order to perform appropriate data modification it is vital to estimate exact time when any significant shift happened, and later calculate the magnitude of shift. There is a broad list of change point detection methods including various statistical tests and models,

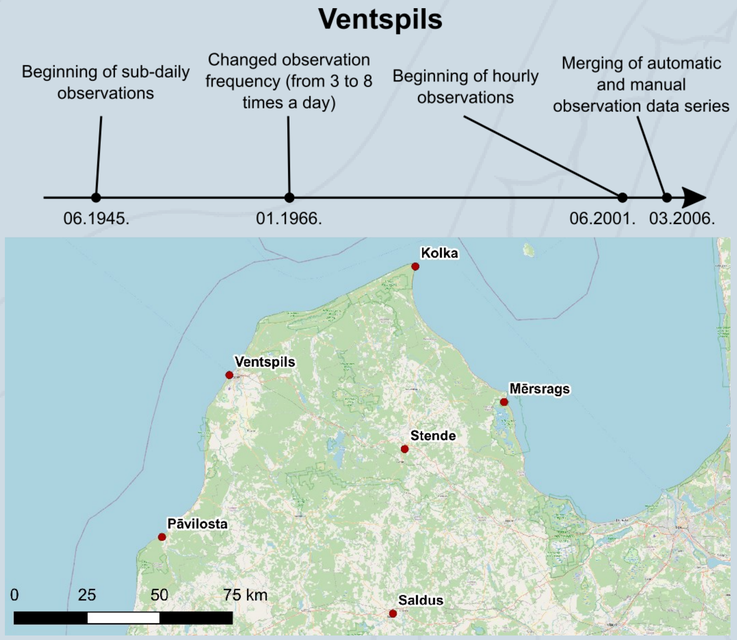

Fig 1. Station location and instrumentation changes in Ventspils observation station

Materials and methods

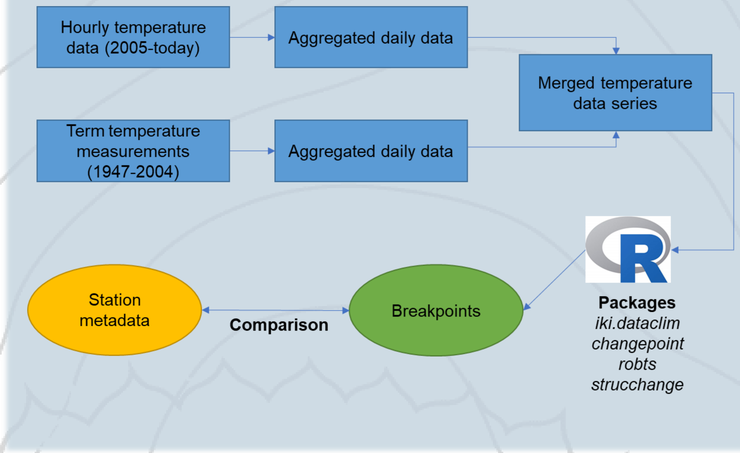

Fig 2. Datasets and methods used

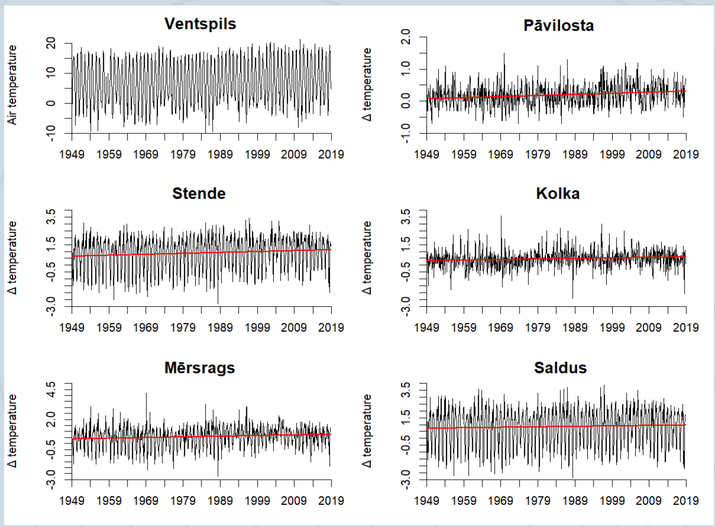

Fig 3. Monthly air temperature and difference series of Ventspils and neighboring meteorological stations. Red line indicates statistically significant trend in difference series

Results

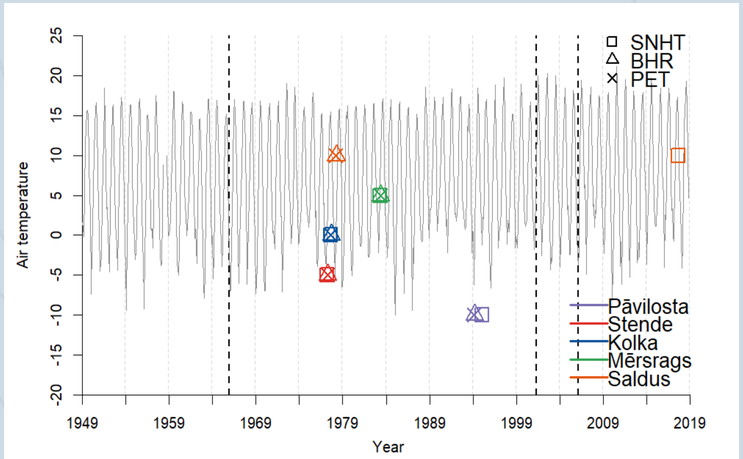

Fig 4. Observed breakpoints in Ventspils meteorological station monthly air temperature series data. Colors indicate difference series between various neighboring stations, dashed lines indicate breakpoints observed from metadata, shapes - different methods used: SNHT - Standard normal Homogeneity test, BHR - Buishand range test, PET - Pettitt test.

For the most significant changes (i.e. changes in station location) that happened not at the very end of the observation period (e.g. Riga station) almost all considered methods worked well. However, changes in station location that happened in 2016 (e.g. Daugavpils and Jelgava stations) were only identified by decision trees. Yet despite successfully identifying changes at the end of the period, machine learning algorithms identified more suspicious break points compared to metadata. All methods (except Pruned Exact Linear Time Method) identified changes in the very middle of the time period at some stations (e.g. Daugavpils, Liepāja, Rēzekne, Ventspils). While metadata recorded no changes in these stations, Mann-Kendall test showed that difference datasets had a significant trend. Therefore these change points could be falsely identified due to non-eliminated tendencies in the observations or could characterize nonrecognized slowly developing differences in the stations (e.g. urbanization).

We plan to look into more detail at the observation data properties and its influences on the results, perform similar analysis for simulated homogeneous and non-homogeneous observations with analogous properties and consider iterative procedure for single change point detection methods.