Mann-Kendall test under the long-term persistence hypothesis

Hristos Tyralis, Aris Tegos

Wednesday, February 08, 2017

Table of Contents

# Introduction

In this report we apply the Mann-Kendall test under the long-term persistence (LTP MKt) hypothesis (Hamed 2008). We show how we can apply the function MannKendallLTP of the "HKprocess" package, to infer about the trend observed in a time series. Furthermore we apply the test to 10 time series of potential evapotranspiration (PET) at various locations in Greece.

We will use the following packages for the analysis.

library("devtools")  
library("FGN")  
library("HKprocess")  
library("knitr")  
library("zoo")

Some information about the current session follows.

session\_info()

## Session info --------------------------------------------------------------

## setting value   
## version R version 3.3.2 (2016-10-31)  
## system x86\_64, mingw32   
## ui RTerm   
## language (EN)   
## collate English\_United States.1252   
## tz Europe/Istanbul   
## date 2017-02-08

## Packages ------------------------------------------------------------------

## package \* version date source   
## akima 0.6-2 2016-12-20 CRAN (R 3.3.2)  
## backports 1.0.5 2017-01-18 CRAN (R 3.3.2)  
## coda 0.19-1 2016-12-08 CRAN (R 3.3.2)  
## devtools \* 1.12.0 2016-06-24 CRAN (R 3.3.1)  
## digest 0.6.12 2017-01-27 CRAN (R 3.3.2)  
## evaluate 0.10 2016-10-11 CRAN (R 3.3.1)  
## FGN \* 2.0-12 2014-05-16 CRAN (R 3.3.1)  
## gtools 3.5.0 2015-05-29 CRAN (R 3.3.0)  
## HKprocess \* 0.0-2 2016-01-15 CRAN (R 3.3.1)  
## htmltools 0.3.5 2016-03-21 CRAN (R 3.3.1)  
## knitr \* 1.15.1 2016-11-22 CRAN (R 3.3.2)  
## lattice 0.20-34 2016-09-06 CRAN (R 3.3.1)  
## ltsa 1.4.6 2015-12-21 CRAN (R 3.3.0)  
## magrittr 1.5 2014-11-22 CRAN (R 3.3.1)  
## MASS 7.3-45 2016-04-21 CRAN (R 3.3.1)  
## Matrix 1.2-8 2017-01-20 CRAN (R 3.3.2)  
## MatrixModels 0.4-1 2015-08-22 CRAN (R 3.3.1)  
## mcmc 0.9-4 2015-07-17 CRAN (R 3.3.1)  
## MCMCpack 1.3-9 2017-01-26 CRAN (R 3.3.2)  
## memoise 1.0.0 2016-01-29 CRAN (R 3.3.1)  
## quantreg 5.29 2016-09-04 CRAN (R 3.3.1)  
## Rcpp 0.12.9 2017-01-14 CRAN (R 3.3.2)  
## rmarkdown 1.3 2016-12-21 CRAN (R 3.3.2)  
## rprojroot 1.2 2017-01-16 CRAN (R 3.3.2)  
## sp 1.2-4 2016-12-22 CRAN (R 3.3.2)  
## SparseM 1.74 2016-11-10 CRAN (R 3.3.2)  
## stringi 1.1.2 2016-10-01 CRAN (R 3.3.1)  
## stringr 1.1.0 2016-08-19 CRAN (R 3.3.1)  
## withr 1.0.2 2016-06-20 CRAN (R 3.3.1)  
## yaml 2.1.14 2016-11-12 CRAN (R 3.3.2)  
## zoo \* 1.7-14 2016-12-16 CRAN (R 3.3.2)

To reproduce the computations, change the in\_dir variable after unzipping the "supplementary information.7z" file.

in\_dir <- "T:"

The following directories define the locations of the data and the results of the code.

work\_dir <- paste(in\_dir, "/supplementary information/code and data/code",  
 sep = "")  
  
data\_dir <- paste(in\_dir, "/supplementary information/code and data/data",  
 sep = "")  
setwd(work\_dir)

work\_dir

## [1] "T:/supplementary information/code and data/code"

data\_dir

## [1] "T:/supplementary information/code and data/data"

We load the PET data which will be used in the case study.

setwd(data\_dir)  
df\_PET\_Greece <- read.csv("PET\_Greece.csv")

We set a seed, to reproduce the analysis.

set.seed(12345)

# Simulation study

## Simulation

To apply the test we simulate n\_test = 5 time series exhibiting Hurst-Kolmogorov (HK) behaviour using the "FGN" package (McLeod et al. 2008). In some of them we add a trend component. The simulated time series are generated from a HK process with = 0 and = 1. The Hurst parameter are set with the H\_values column of the sim\_values data frame. The added trend component is set with the trend\_values column of the sim\_values data frame. Finally we produce the sim\_series list, with 5 items, each one being a simulated time series.

sim\_values <- data.frame(matrix(nrow = n\_test, ncol = 3))  
names(sim\_values) <- c("H\_values", "n\_values", "trend\_values")  
  
sim\_values$H\_values <- c(0.5, 0.8, 0.9, 0.95, 0.5)  
# Values of H parameters used in the simulation  
  
sim\_values$n\_values <- c(100, 100, 100, 100, 100)  
# Length of the simulated time series  
  
sim\_values$trend\_values <- c(0, 0, 0, 0, 0.01)  
# Values of added trend component used in the simulation  
  
sim\_series <- list()  
  
for (i in 1:n\_test)  
{  
 sim\_series[[i]] <- SimulateFGN(n = sim\_values$n\_values[i], H =  
 sim\_values$H\_values[i]) + sim\_values$trend\_values[i] \* ((1 - 50):  
 (sim\_values$n\_values[i] - 50))  
}

Now we construct the df\_results data frame. Its first three columns contain the parameters used for the generation of the simulated time series.

df\_results <- data.frame(matrix(nrow = n\_test, ncol = 8))  
  
names(df\_results) <- c("H\_values", "n\_values", "trend\_values", "H\_estimate",  
 "trend\_estimate", "Mann\_Kendall\_2\_sided\_pvalue",  
 "Significance\_of\_H\_2\_sided\_pvalue", "Mann\_Kendall\_LTP\_2\_sided\_pvalue")

df\_results$H\_values <- sim\_values$H\_values  
df\_results$n\_values <- sim\_values$n\_values  
df\_results$trend\_values <- sim\_values$trend\_values

We present the df\_results data frame, with its first three columns filled.

df\_results

## H\_values n\_values trend\_values H\_estimate trend\_estimate  
## 1 0.50 100 0.00 NA NA  
## 2 0.80 100 0.00 NA NA  
## 3 0.90 100 0.00 NA NA  
## 4 0.95 100 0.00 NA NA  
## 5 0.50 100 0.01 NA NA  
## Mann\_Kendall\_2\_sided\_pvalue Significance\_of\_H\_2\_sided\_pvalue  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## Mann\_Kendall\_LTP\_2\_sided\_pvalue  
## 1 NA  
## 2 NA  
## 3 NA  
## 4 NA  
## 5 NA

## Application of the test

Now we

* Model the time series with an HK process and we estimate the parameters. We use the Maximum Likelihood Estimator as presented in Tyralis and Koutsoyiannis (2011) and we use the "HKprocess" package. We do this even for the simulated time series with the added trend component.
* Apply the LTP MKt in all simulated time series using the "HKprocess" package.
* Fit a linear model to all simulated time series.

mleHK\_values <- list()  
MannKendallLTP\_values <- list()  
linear\_values <- list()  
linear\_x <- list()  
  
for (i in 1:n\_test)  
{  
 mleHK\_values[[i]] <- mleHK(sim\_series[[i]])  
 # Estimation of the parameters of the HK process  
  
 MannKendallLTP\_values[[i]] <- MannKendallLTP(sim\_series[[i]])  
 # Application of the LTP MKt  
  
 linear\_x[[i]] <- 1:sim\_values$n\_values[i]  
  
 linear\_values[[i]] <- lm(sim\_series[[i]] ~ linear\_x[[i]])  
 # Linear regression on the time series  
}

We fill the df\_results data frame with the values of interest.

for (i in 1:n\_test)  
{  
 df\_results$H\_estimate[i] <- mleHK\_values[[i]][3]  
 df\_results$trend\_estimate[i] <- linear\_values[[i]]$coefficients[2]  
  
 df\_results$Mann\_Kendall\_2\_sided\_pvalue[i] <-  
 MannKendallLTP\_values[[i]]$Mann\_Kendall[6]  
  
 df\_results$Significance\_of\_H\_2\_sided\_pvalue[i] <-  
 MannKendallLTP\_values[[i]]$Significance\_of\_H[2]  
  
 df\_results$Mann\_Kendall\_LTP\_2\_sided\_pvalue[i] <-  
 MannKendallLTP\_values[[i]]$Mann\_Kendall\_LTP[2]  
}

The df\_results data frame now contains all information of interest for the current report. The values of interest for the LTP MKt (i.e. values in the last three columns) will be explained in the next Section.

df\_results

## H\_values n\_values trend\_values H\_estimate trend\_estimate  
## 1 0.50 100 0.00 0.5718440 -0.008047717  
## 2 0.80 100 0.00 0.7887196 0.008649039  
## 3 0.90 100 0.00 0.9001521 -0.004291134  
## 4 0.95 100 0.00 0.7603781 0.001514865  
## 5 0.50 100 0.01 0.6158808 0.009397300  
## Mann\_Kendall\_2\_sided\_pvalue Significance\_of\_H\_2\_sided\_pvalue  
## 1 0.037917319 3.951581e-01  
## 2 0.004453685 1.132089e-05  
## 3 0.107144596 6.306999e-09  
## 4 0.304208481 3.449777e-05  
## 5 0.009993012 3.568642e-01  
## Mann\_Kendall\_LTP\_2\_sided\_pvalue  
## 1 0.10568776  
## 2 0.34427404  
## 3 0.71070572  
## 4 0.71620495  
## 5 0.04842272

## The three steps of the Mann-Kendall test under the long-term persistence hypothesis

The LTP MKt consists of three steps, namely **O**, **H** and **M**. These three steps are presented in Hamed (2008). Each step is a hypothesis test. Let *H*0*i* denote the null hypothesis of each test and let *H*1*i* denote the alternative hypothesis, where *i* = **O**, **H**, **M** denotes the step of the LTP MKt.

Then:

* Step **O**

*H*0**O**: No significant trend under the independence assumption.

*H*1**O**: Significant trend under the independence assumption.Proceed to step **H**.

* Step **H**

*H*0**H**: No LTP. It can be concluded that there exists a trend.

*H*1**H**: LTP. Proceed to step **M**.

* Step **M**

*H*0**M**: No significant trend.

*H*1**M**: Significant trend exists.

The three steps are sumarrized by the following sequences

1. {*H*0**O**}, no significant trend.
2. {*H*1**O** , *H*0**H**}, significant trend exists.
3. {*H*1**O** , *H*1**H** , *H*0**M**}, no significant trend.
4. {*H*1**O** , *H*1**H** , *H*1**M**}, significant trend exists.

In LTP MKt we compute the *p*-value for each step. The *p*-value for each step is presented in the last three columns of the df\_results data frame. A *p*-value lower than a predefined level (e.g. = 0.05) gives evidence that *H*1 is true. If the *p*-value is higher than , then we accept that *H*0 is true.

## Application example

The part of interest of the df\_results data frame follows. This part contains the *p*-values.

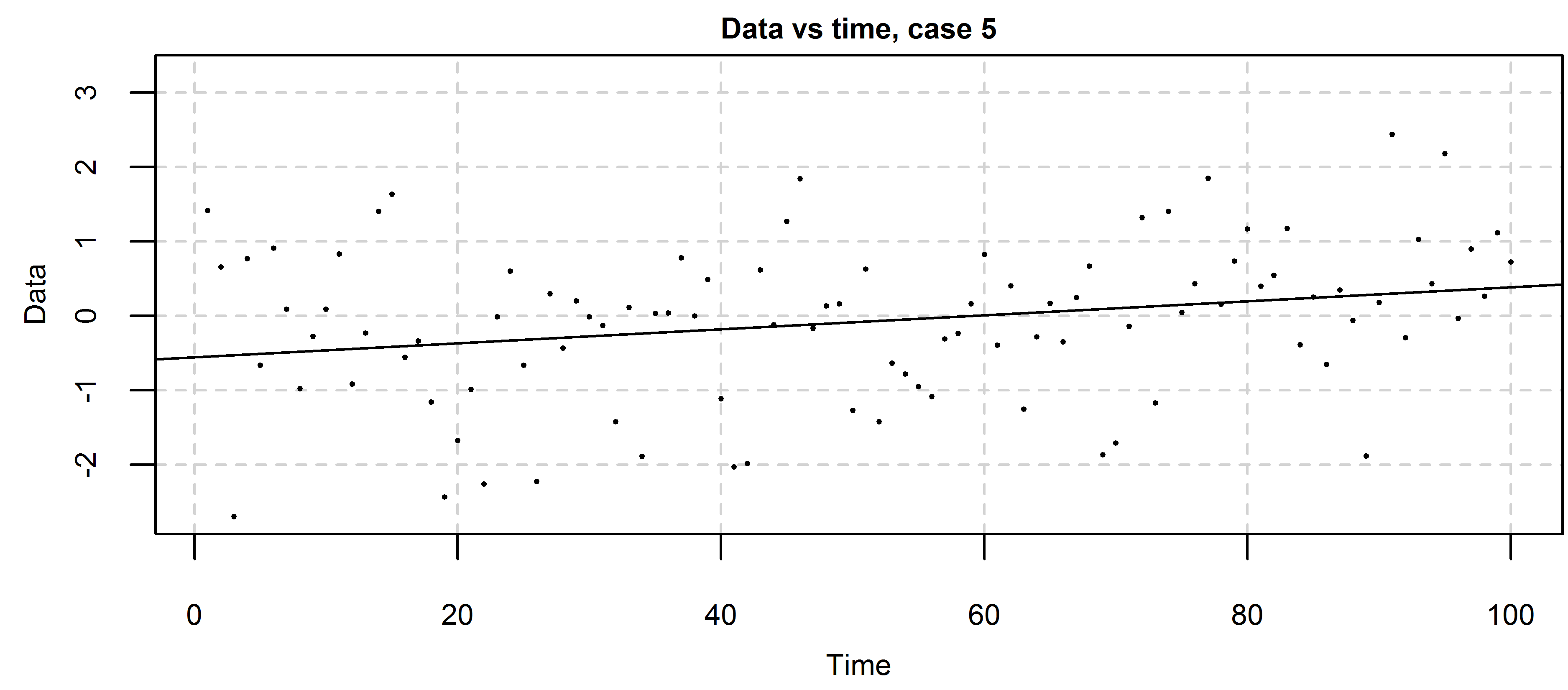
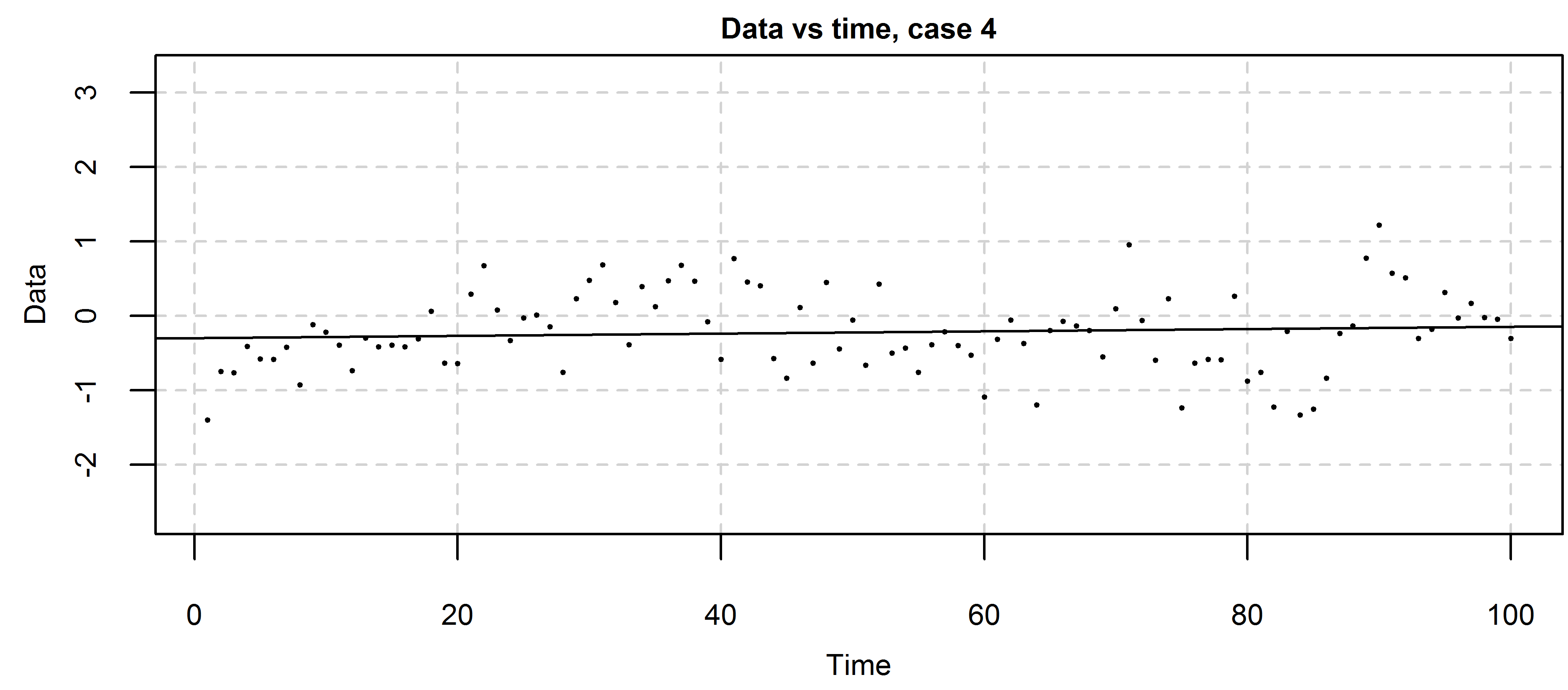
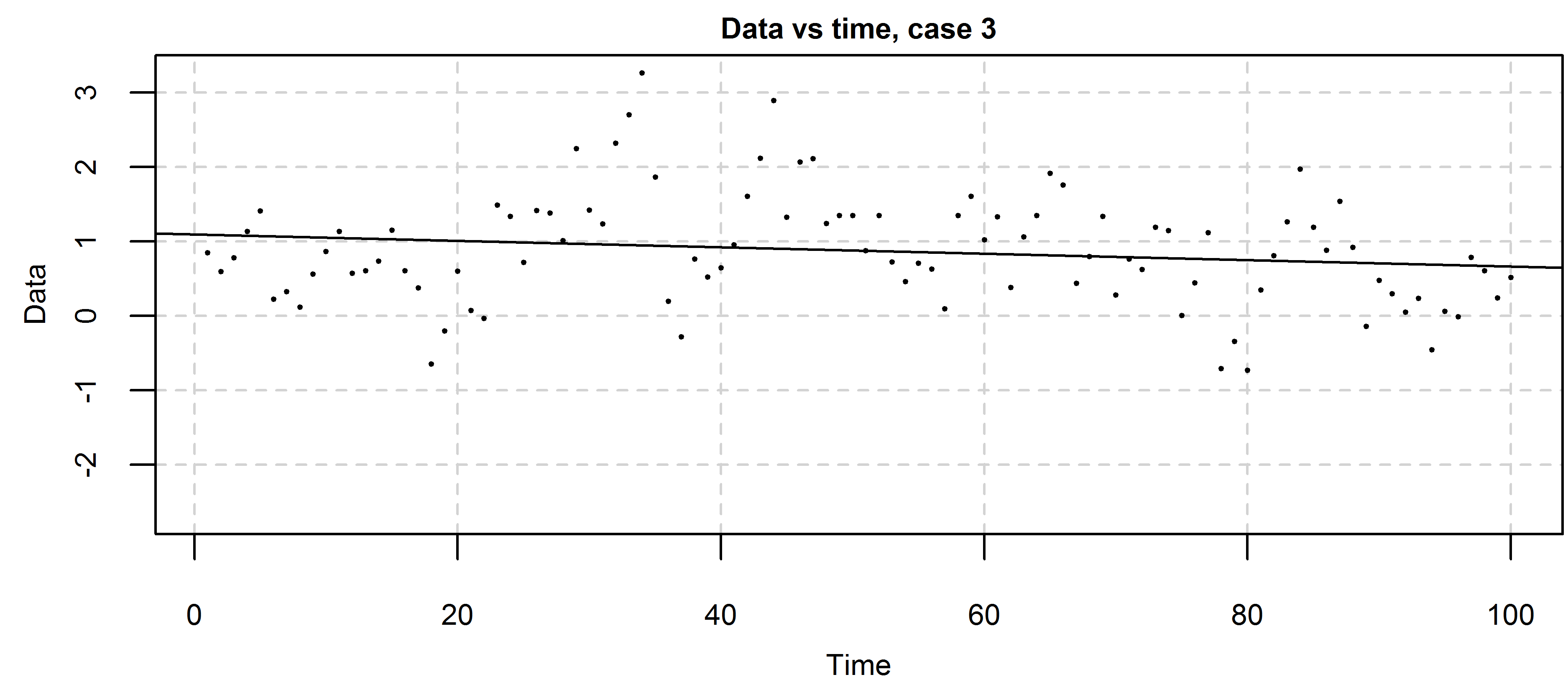
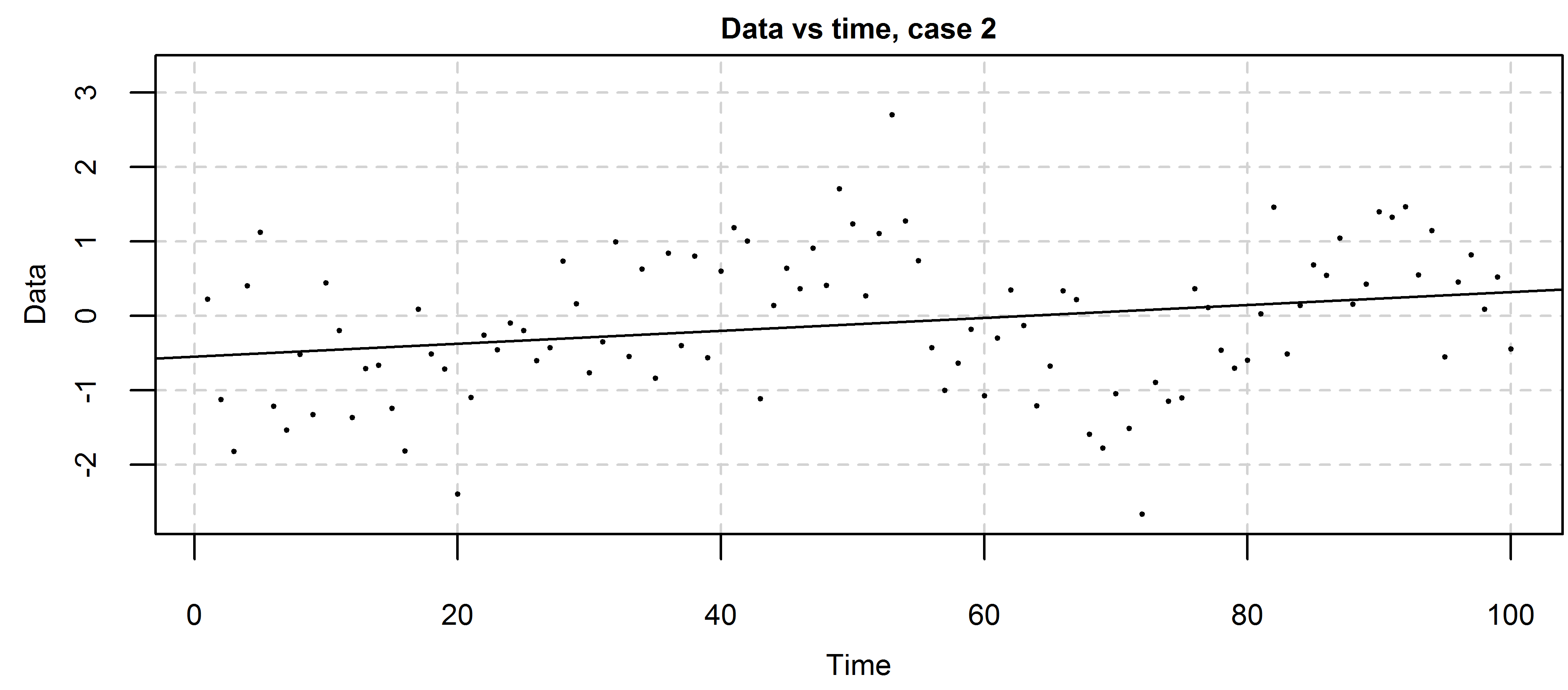
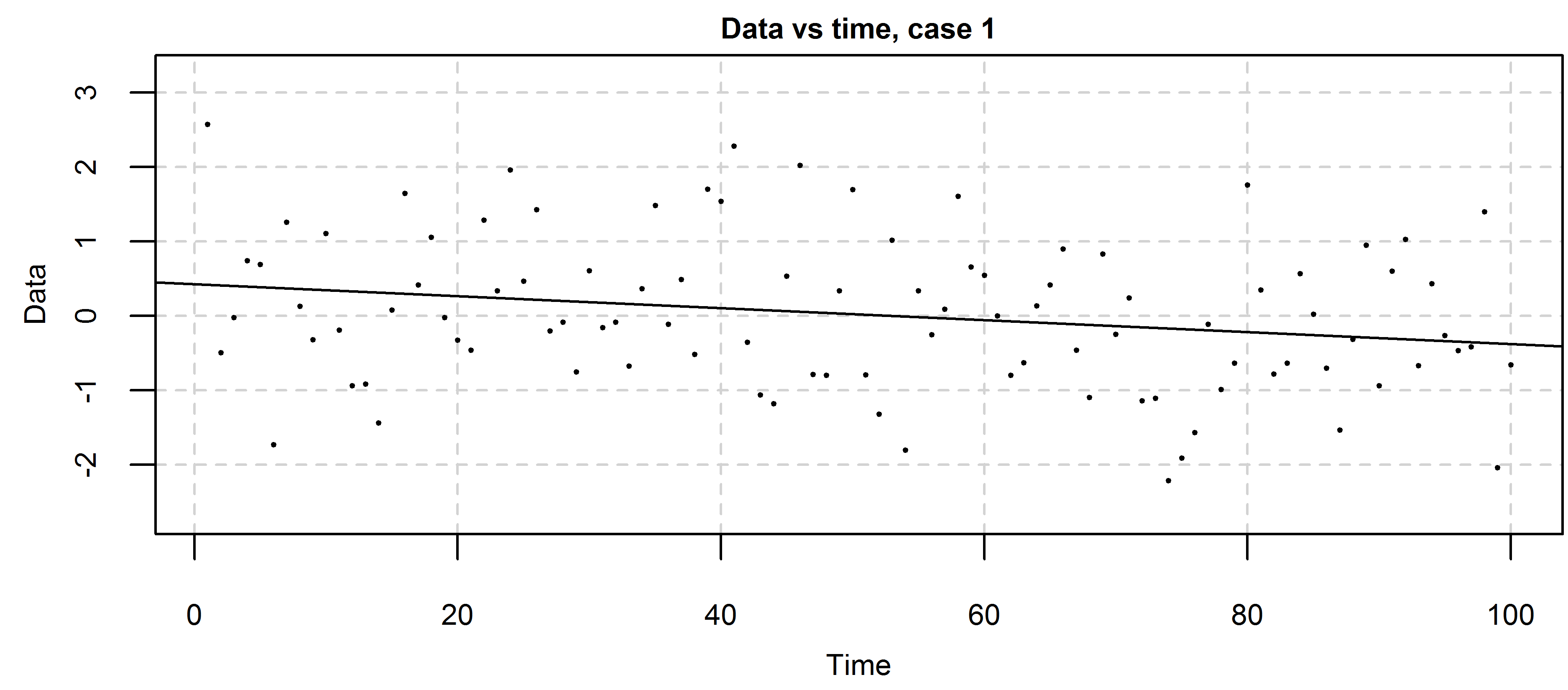
|  |  |  |
| --- | --- | --- |
| Mann\_Kendall\_2\_sided\_pvalue | Significance\_of\_H\_2\_sided\_pvalue | Mann\_Kendall\_LTP\_2\_sided\_pvalue |
| 0.0379173 | 0.3951581 | 0.1056878 |
| 0.0044537 | 0.0000113 | 0.3442740 |
| 0.1071446 | 0.0000000 | 0.7107057 |
| 0.3042085 | 0.0000345 | 0.7162050 |
| 0.0099930 | 0.3568642 | 0.0484227 |

Let = 0.05. The application of the LTP MKt results in the following.

1. {*H*1**O** , *H*0**H**}, significant trend exists.
2. {*H*1**O** , *H*1**H** , *H*0**M**}, no significant trend.
3. {*H*0**O**}, no significant trend.
4. {*H*0**O**}, no significant trend.
5. {*H*1**O** , *H*0**H**}, significant trend exists.

## Figures

We visualize each case to better understand the findings.



# Case study

## Application of the Mann-Kendall test under the long-term persistence hypothesis

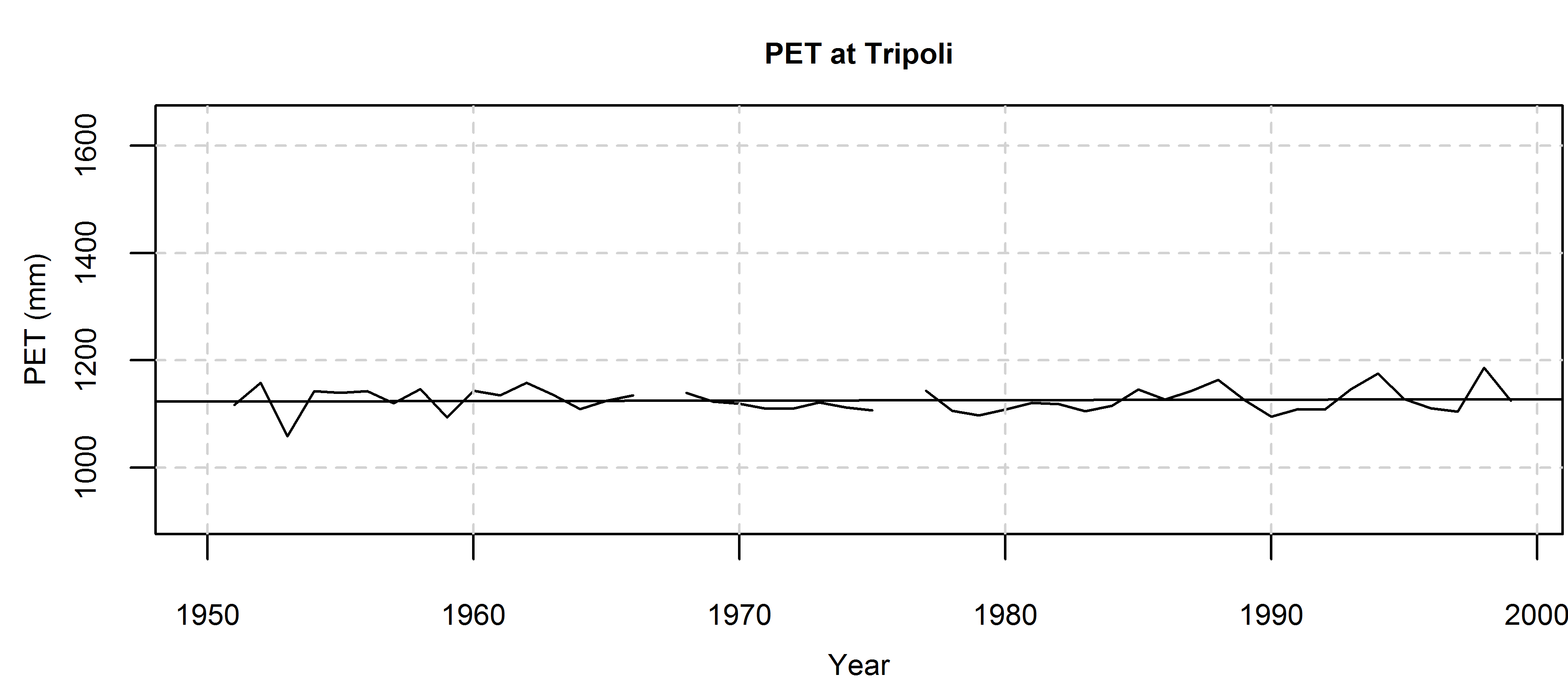
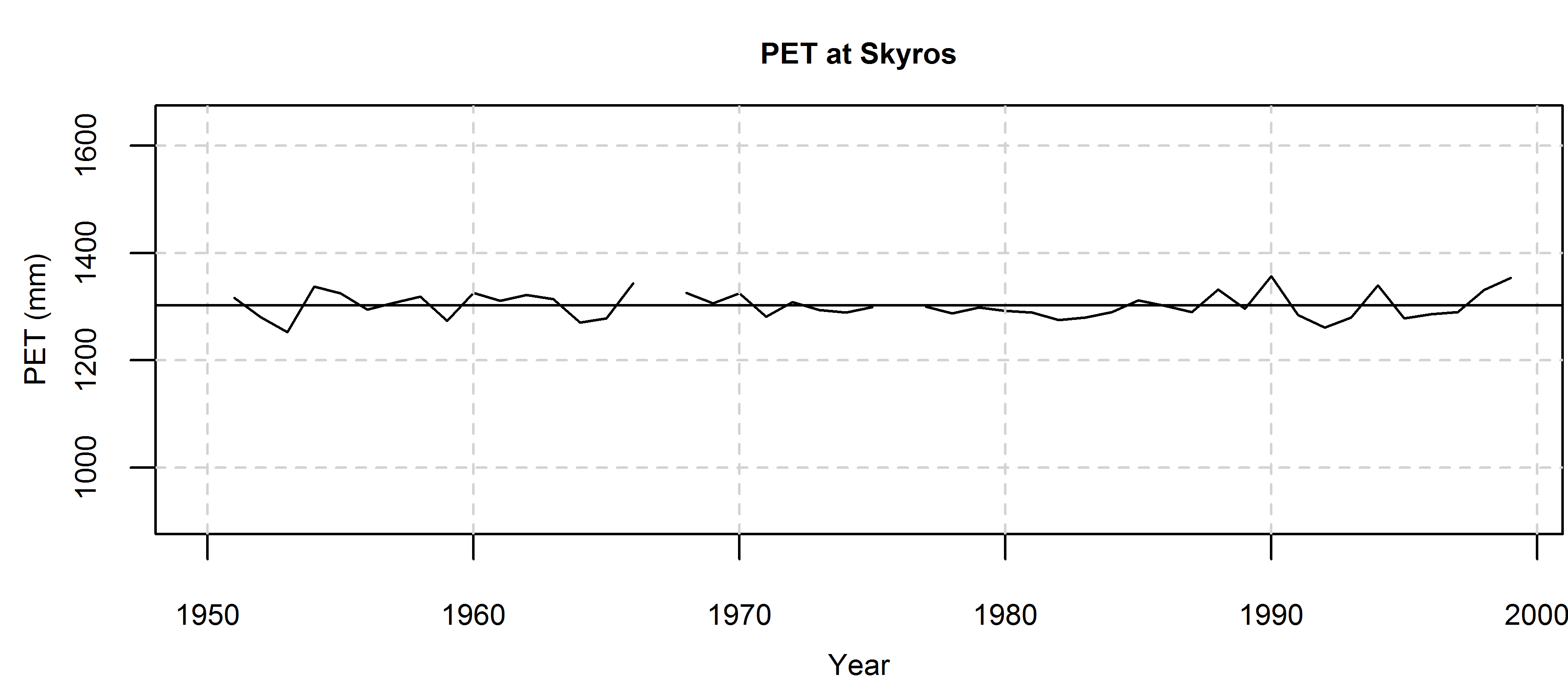
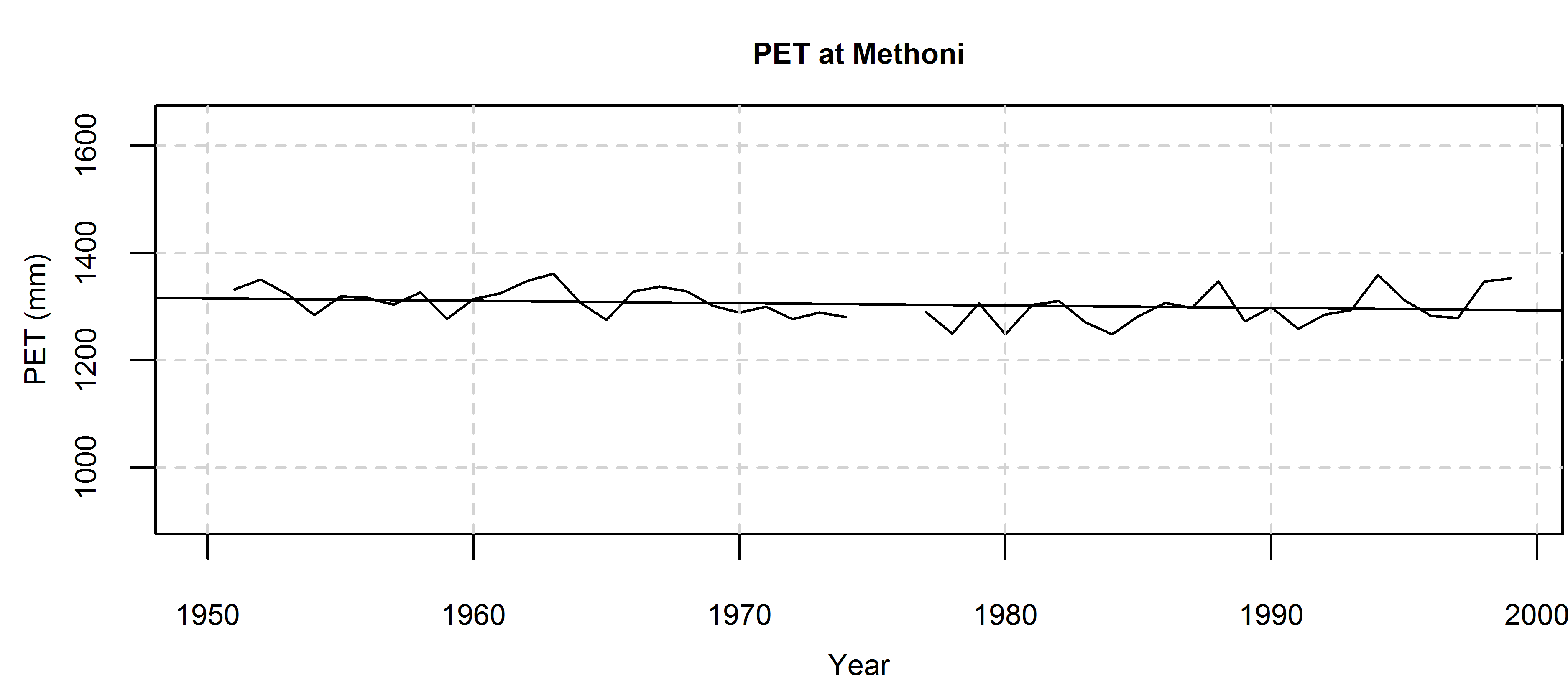
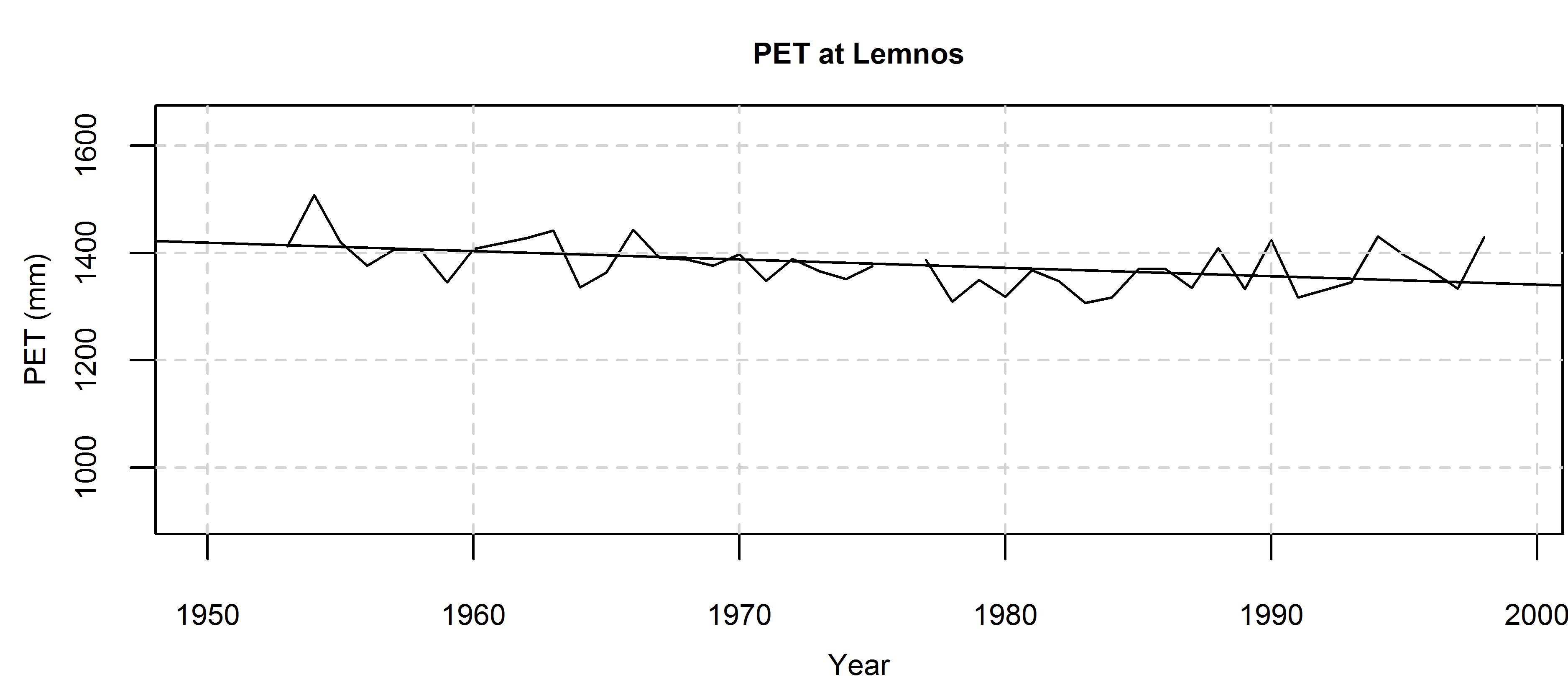
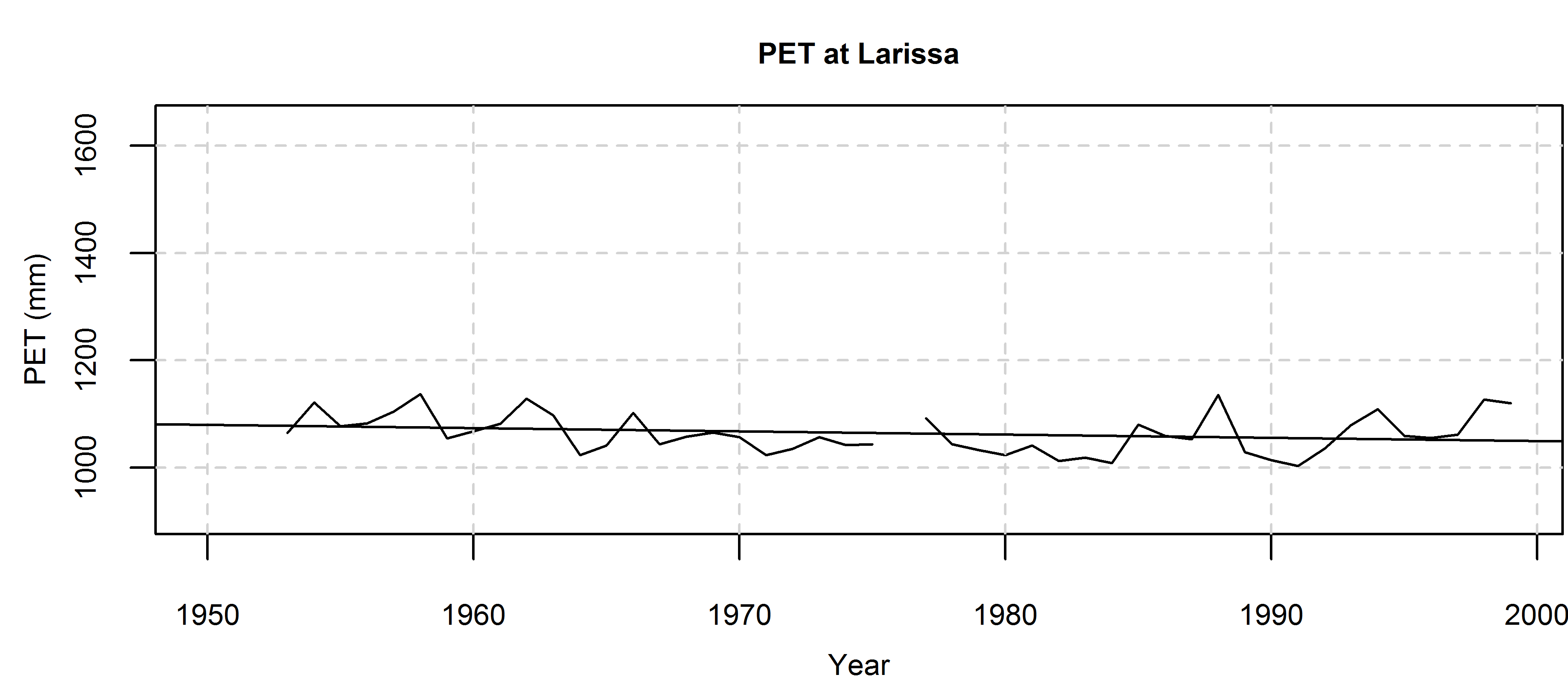
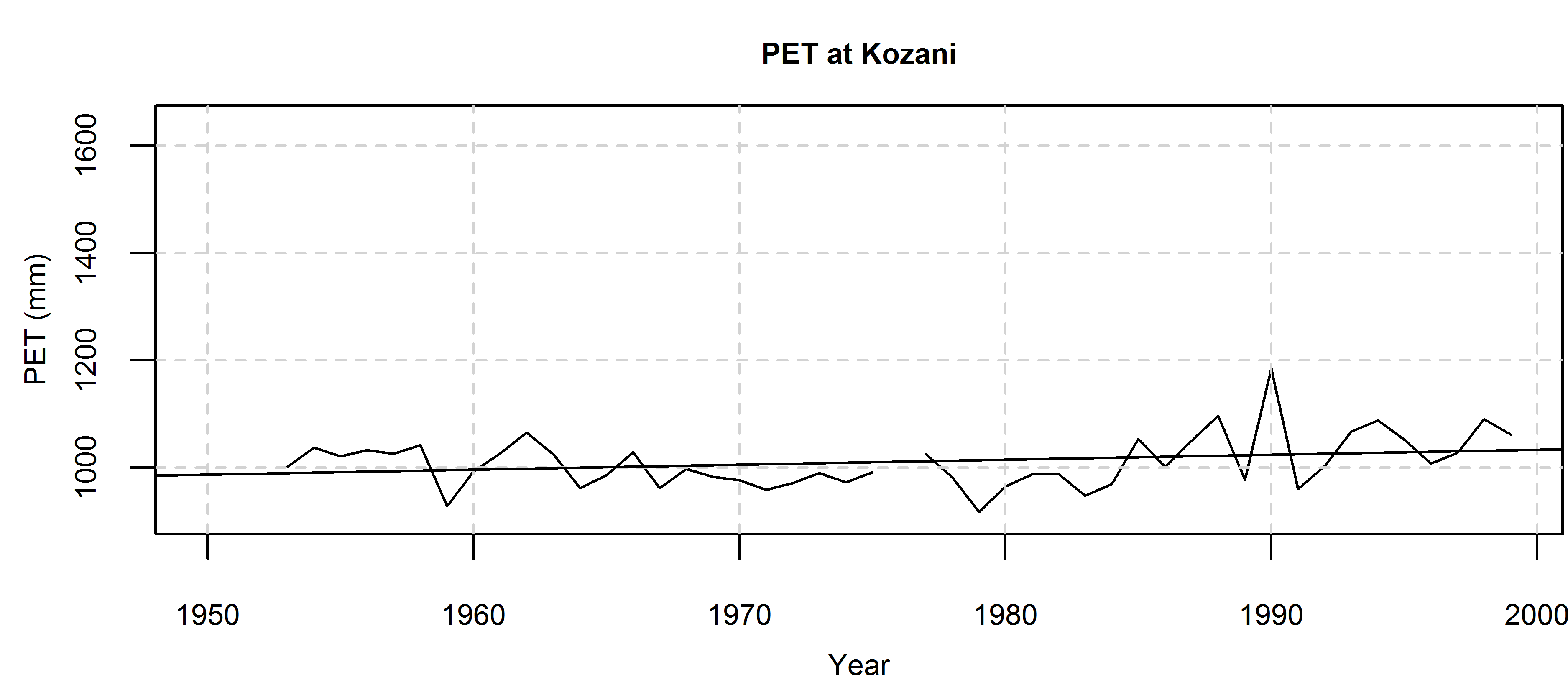
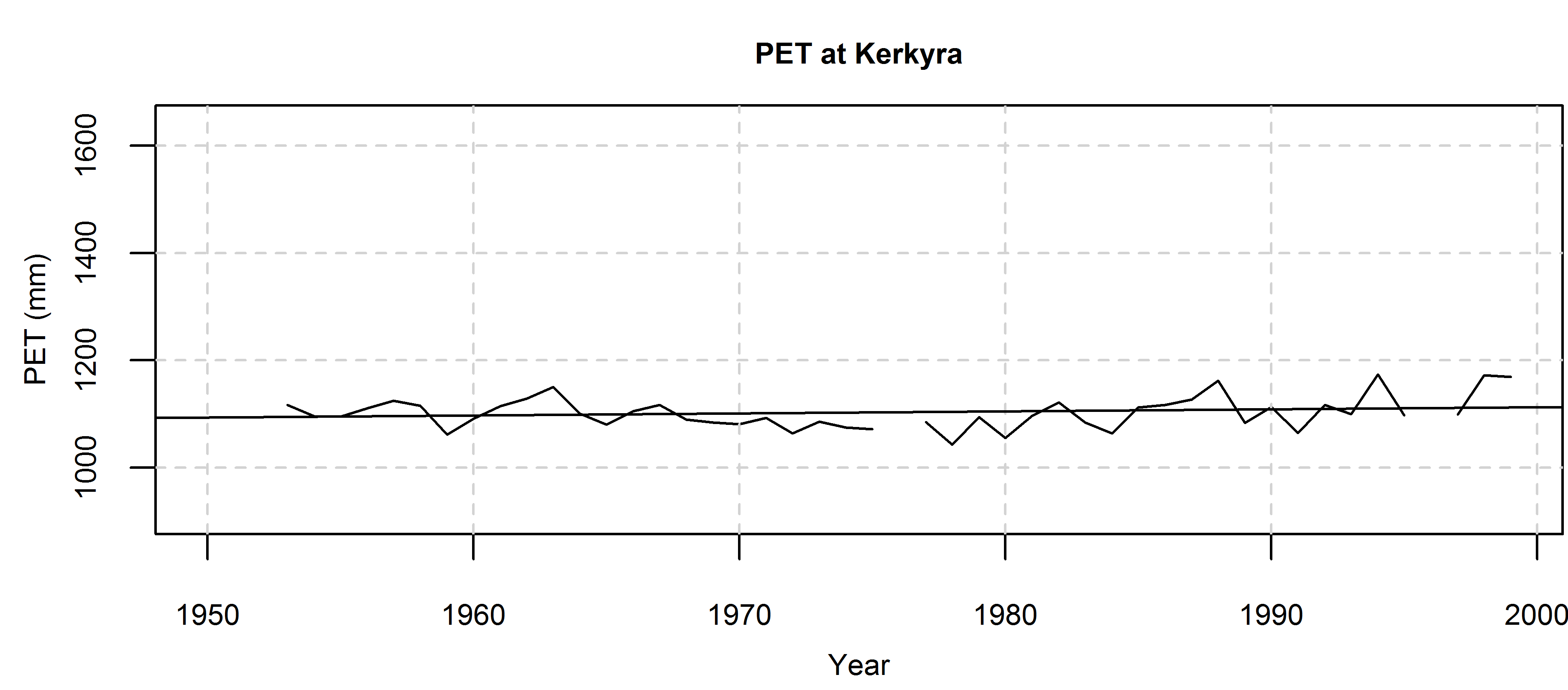
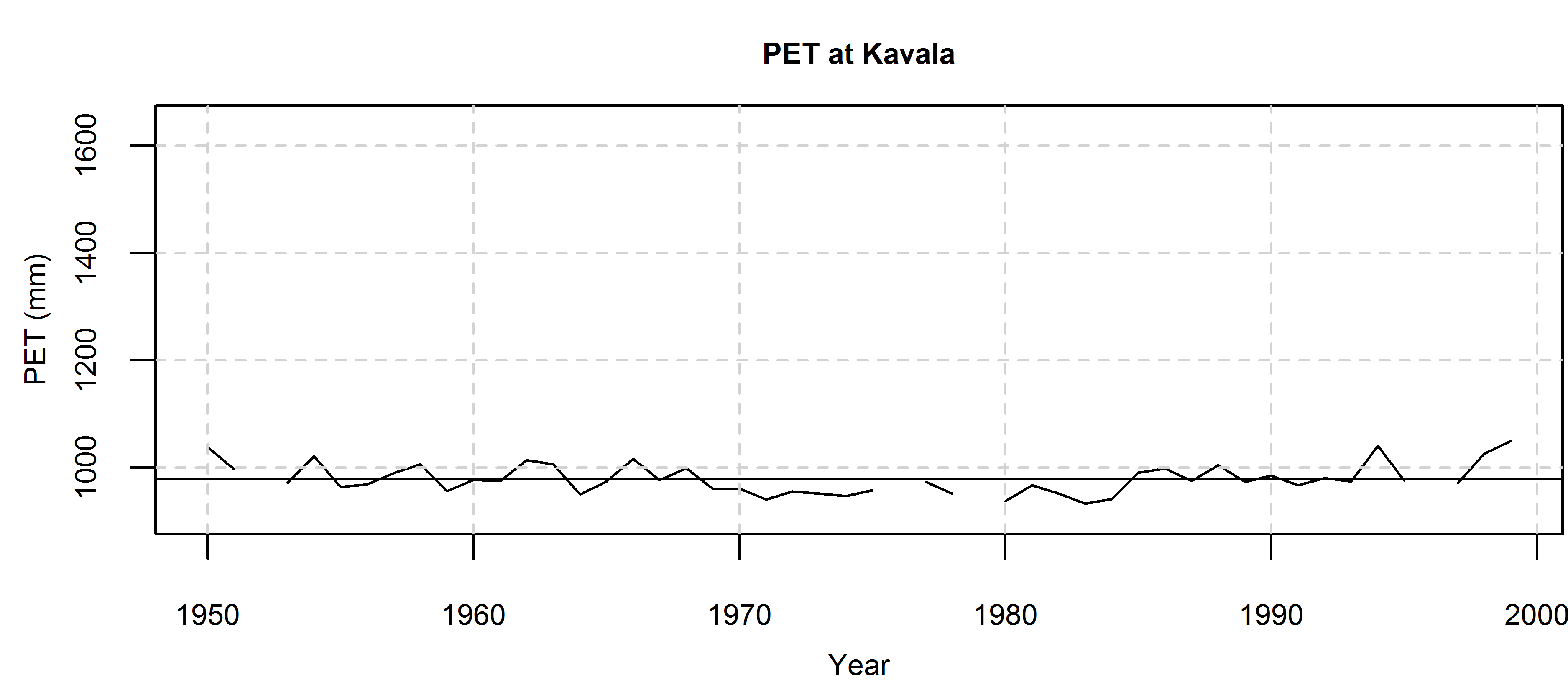
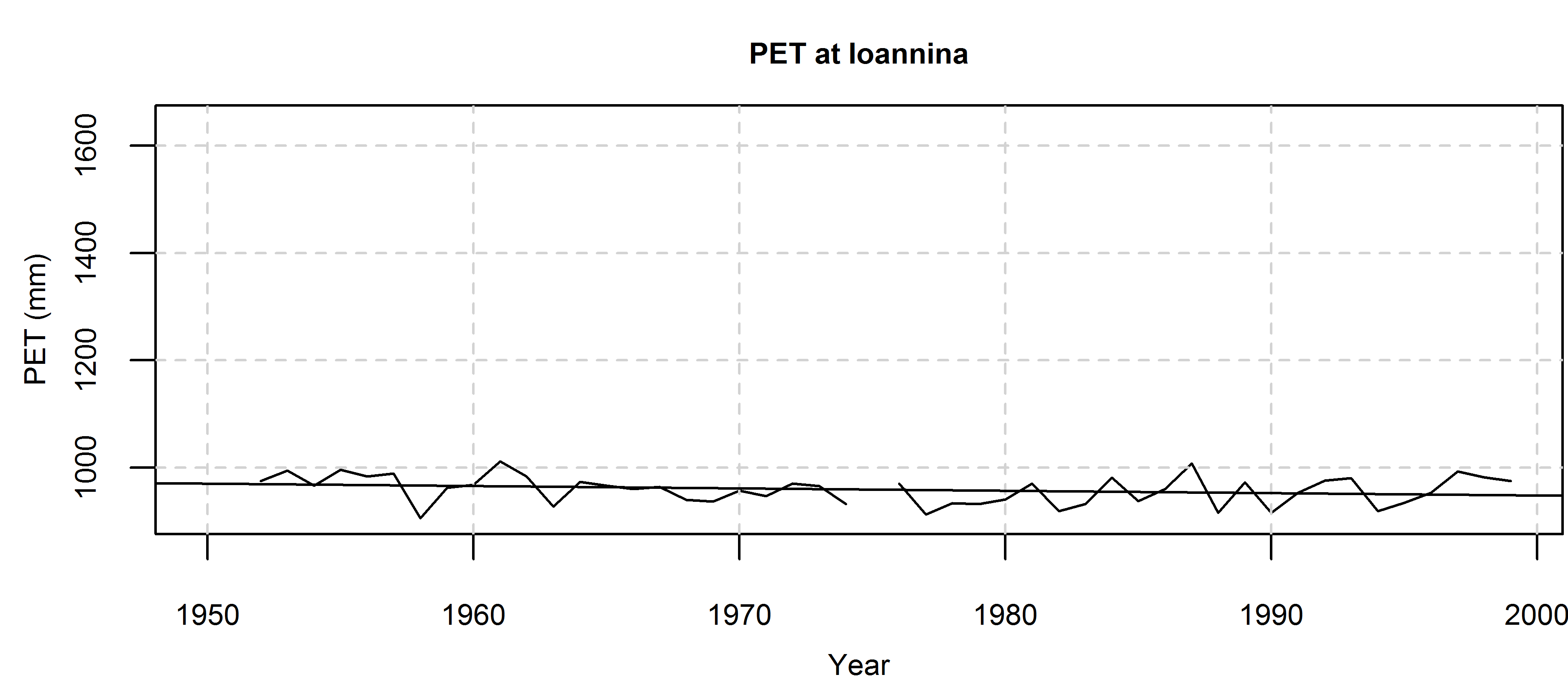
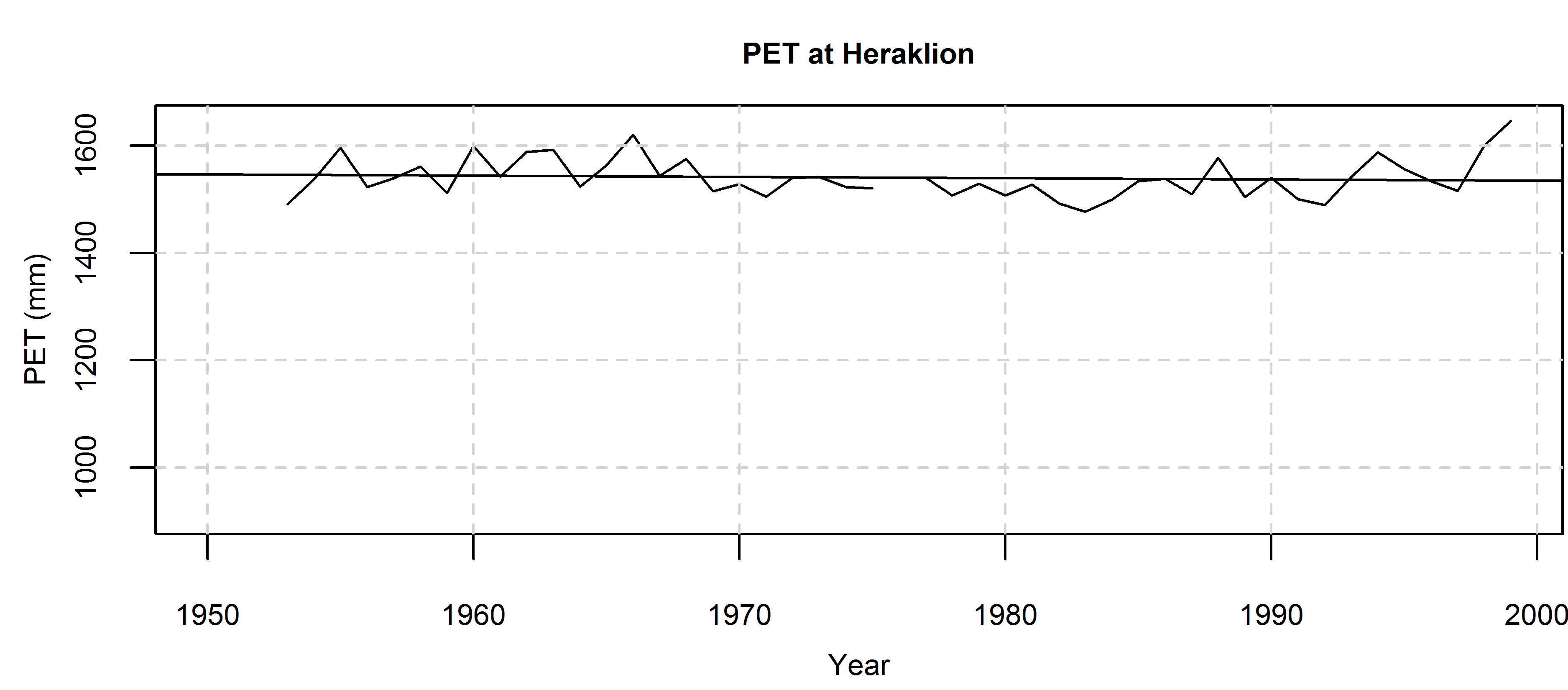
In this Section we apply the LTP MKt to 10 time series of PET at various locations in Greece.

The table with the results, when setting = 0.05 for all steps of the LTP MKt are presented in the following table.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Station | appr\_time\_series\_length | H\_estimate | trend\_estimate | Mann\_Kendall\_2\_sided\_pvalue | Significance\_of\_H\_2\_sided\_pvalue | Mann\_Kendall\_LTP\_2\_sided\_pvalue | Trend\_existence |
| Heraklion | 49 | 0.672397 | -0.2377255 | 0.3090865 | 0.02283812 | 0.6298758 | {H0O}, no trend |
| Ioannina | 50 | 0.5801406 | -0.4381775 | 0.04558661 | 0.2720165 | 0.179922 | {H1O,H0H}, trend exists |
| Kavala | 50 | 0.7614821 | 0.01110683 | 0.6275616 | 0.0002330466 | 0.8765082 | {H0O}, no trend |
| Kerkyra | 49 | 0.708458 | 0.3846636 | 0.8971222 | 0.004979742 | 0.9579901 | {H0O}, no trend |
| Kozani | 49 | 0.6303466 | 0.9253746 | 0.3132045 | 0.0368701 | 0.6140011 | {H0O}, no trend |
| Larissa | 49 | 0.7599349 | -0.5992912 | 0.03938509 | 0.003407958 | 0.4171651 | {H1O,H1H,HOM} no trend |
| Lemnos | 48 | 0.7359009 | -1.565468 | 0.0005824204 | 0.2593468 | 0.0230132 | {H1O,H0H}, trend exists |
| Methoni | 49 | 0.6934991 | -0.4428614 | 0.05678219 | 0.03758929 | 0.3396473 | {H0O}, no trend |
| Skyros | 49 | 0.4615806 | -0.005355828 | 0.4030851 | 0.8524974 | 0.4575056 | {H0O}, no trend |
| Tripoli | 49 | 0.4278933 | 0.08245386 | 0.46375 | 0.6086242 | 0.5571946 | {H0O}, no trend |

## Figures

We visualize each case to better understand the findings.



# References

Hamed KH (2008) Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis. Journal of Hydrology 349(3-4):350-363. doi: [10.1016/j.jhydrol.2007.11.009](http://dx.doi.org/10.1016/j.jhydrol.2007.11.009)

McLeod AI, Yu H, Krougly ZL (2008) Algorithms for Linear Time Series Analysis: With R Package. Journal of Statistical Software 23(5)

Tyralis H, Koutsoyiannis D (2011) Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process. Stochastic Environmental Research and Risk Assessment 25(1):21-33. doi: [10.1007/s00477-010-0408-x](http://dx.doi.org/10.1007/s00477-010-0408-x)