CLIMATE
RESEARCH
FOUNDATION

9th EUMETNET Data Management Workshop

*6th-8th November 2013, El Escorial, Madrid*

# Detection of inhomogeneities on daily data: a test based on the Kolmogorov-Smirnov test

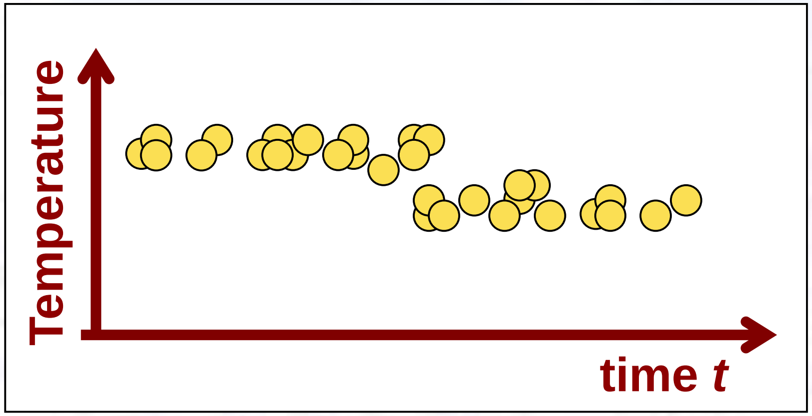

**Robert Monjo, Javier Pórtoles and Jaime Ribalaygua**

**Climate Research Foundation**
**(FIC, Fundación para la Investigación del Clima), Madrid**
**(fic@ficlima.org)**

# Contents

# Introduction





# Introduction



## Problem: Abrupt change in observations

- Changes in the weather shelter (best / worst ventilation, white painting, ...)
- Changes in meteorological sensors
- Changes in the location of weather stations
- Changes in the environment (vegetation, buildings, ...)

## Current solutions

- SNHT: Standard Normal Homogeneity Test (by Alexandersson, 1986)
- Others methods at **monthly scale**: Buishand range test, Pettitt test, von Neumann ratio tests,…
- Some methods at **daily scale** (mean, quantiles or moments): Using parallel measurements, reference series,… combining with corrections as HOM, HOMAD, SPLIDHOM, QM, PM,... **but** from a **breakpoint** detection at **monthly** scale.

## Current limitations

- Detection of inhomogeneities in daily time-series is partially saved.
- Alternatives should be explored

# Contents

# Methodology



## Definitions

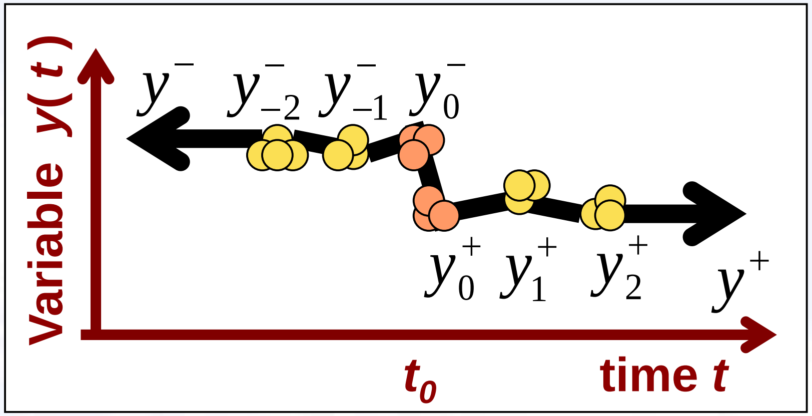
$y_i$ ≡ set of daily data

$t_0$ ≡ inhomogeneity candidate

$t^-$ ≡ time values to the left

$t^+$ ≡ time values to the right

PV ≡ p-value of KS test




### Measure of the dissimilitude between two sets (e.g. 365 days)

$$LPV_j^i \equiv \log_{10}\left(PV(y_i, y_j)\right)$$

$$PV_j^i \in (0,\ 1) \longrightarrow LPV_j^i \in (-\infty,\ 0)$$

High ↔ low dissimilitude

### Measure of possible Inhomogeneity jumps: Dissimilitude between 2 contiguous sets

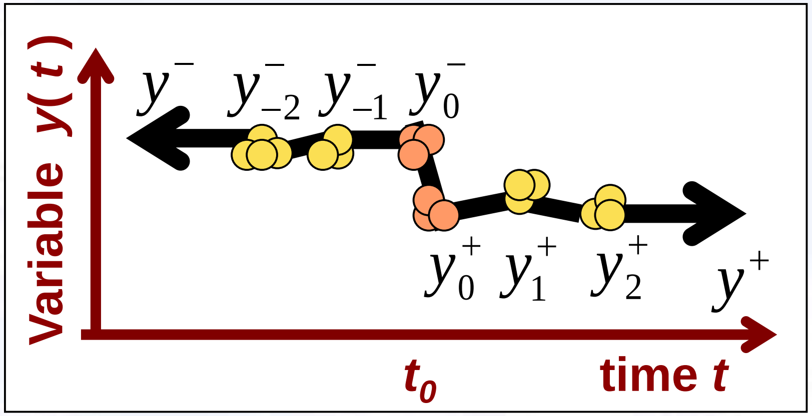
$$LPV_{i+1}^i \equiv \log_{10}\left(PV(y_i, y_{i+1})\right)$$

# Methodology



## Control Analysis

**Introducing** an **artifitial inhomogeneity** to each set of data $y_i \rightarrow \tilde{y}_i$



# Methodology

## Control Analysis

**Introducing** an **artifitial inhomogeneity** to each set of data $y_i \rightarrow \tilde{y}_i$

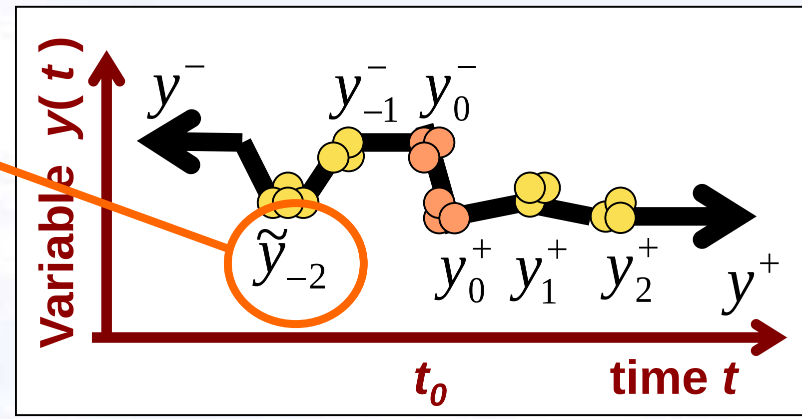
$$LPV^{i}_{\{i+1\}} \equiv \log_{10}\left(PV\left(y_i, \tilde{y}_{i+1}\right)\right)$$

| ***Variable*** | ***a*** | ***b*** |
|---|---|---|
| Temperature | 1 | 2 |
| Precipitation | 3 | 0 |

EXAMPLE

$$\tilde{y}_i \equiv a^{\pm 1}\cdot y_i \pm b$$

**Artificial inhomogeneity**




**Reference LPV** is defined from the **average value**: inhomogeneity of control

$$LPV_{Inh} \equiv \frac{1}{N-1}\sum_{i=1}^{N-1} LPV^{i}_{\{i+1\}}$$

$$LPV^{i}_{i+1} \le LPV_{Inh} \rightarrow \begin{cases} y^-_0 = y_i \\ y^+_0 = y_{i+1} \end{cases}$$

# Methodology



## Inhomogeneity detection

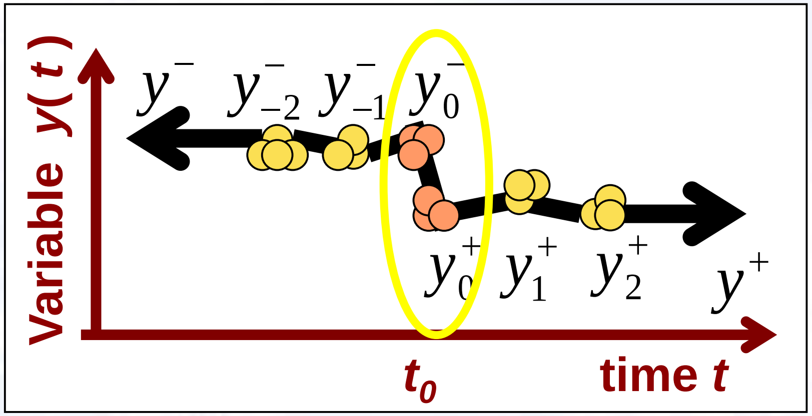
Similarity between the **candidates** $y_0$ and the other **populations** that are on the left (-) and right (+)

$$LPV_-^- \equiv \log_{10}\left(PV(y^-, y_0^-)\right)$$ **Similar**

$$LPV_-^+ \equiv \log_{10}\left(PV(y^+, y_0^-)\right)$$ **different**

$$LPV_+^- \equiv \log_{10}\left(PV(y^-, y_0^+)\right)$$ **different**

$$LPV_+^+ \equiv \log_{10}\left(PV(y^+, y_0^+)\right)$$ **Similar**



# Methodology

## Inhomogeneity detection

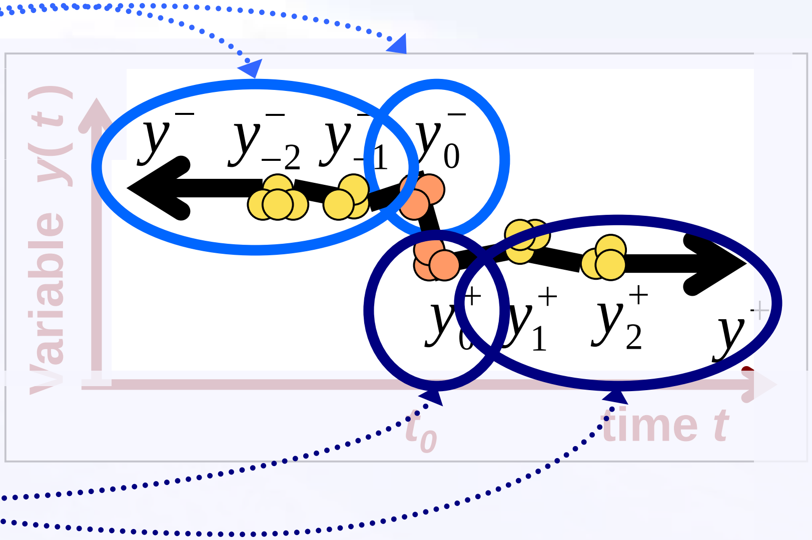
Similarity between the **candidates** $y_0$ and the other **populations** that are on the left (-) and right (+)

$$LPV_-^- \equiv \log_{10}\left(PV(y^-, y_0^-)\right)$$ **Similar**

$$LPV_-^+ \equiv \log_{10}\left(PV(y^+, y_0^-)\right)$$ **different**

$$LPV_+^- \equiv \log_{10}\left(PV(y^-, y_0^+)\right)$$ **different**

$$LPV_+^+ \equiv \log_{10}\left(PV(y^+, y_0^+)\right)$$ **Similar**



# Methodology



## Inhomogeneity detection

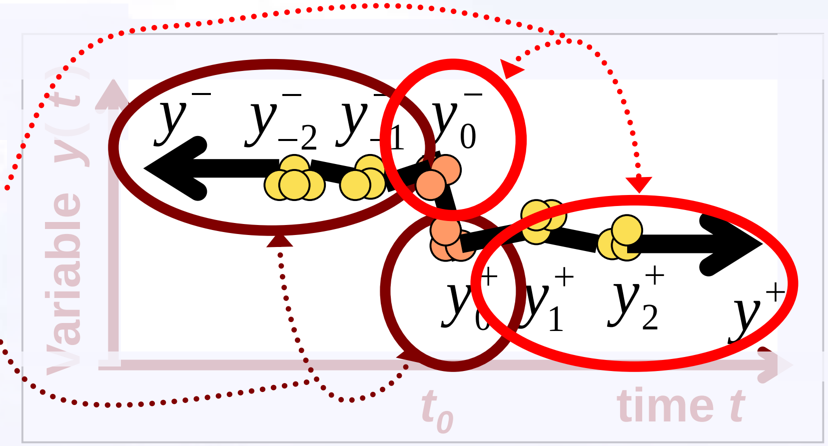
Similarity between the **candidates** $y_0$ and the other **populations** that are on the left (-) and right (+)

$$LPV_{-}^{-} \equiv \log_{10}\left(PV\left(y^{-}, y_{0}^{-}\right)\right)$$ Similar

$$LPV_{-}^{+} \equiv \log_{10}\left(PV\left(y^{+}, y_{0}^{-}\right)\right)$$ different

$$LPV_{+}^{-} \equiv \log_{10}\left(PV\left(y^{-}, y_{0}^{+}\right)\right)$$ different

$$LPV_{+}^{+} \equiv \log_{10}\left(PV\left(y^{+}, y_{0}^{+}\right)\right)$$ Similar

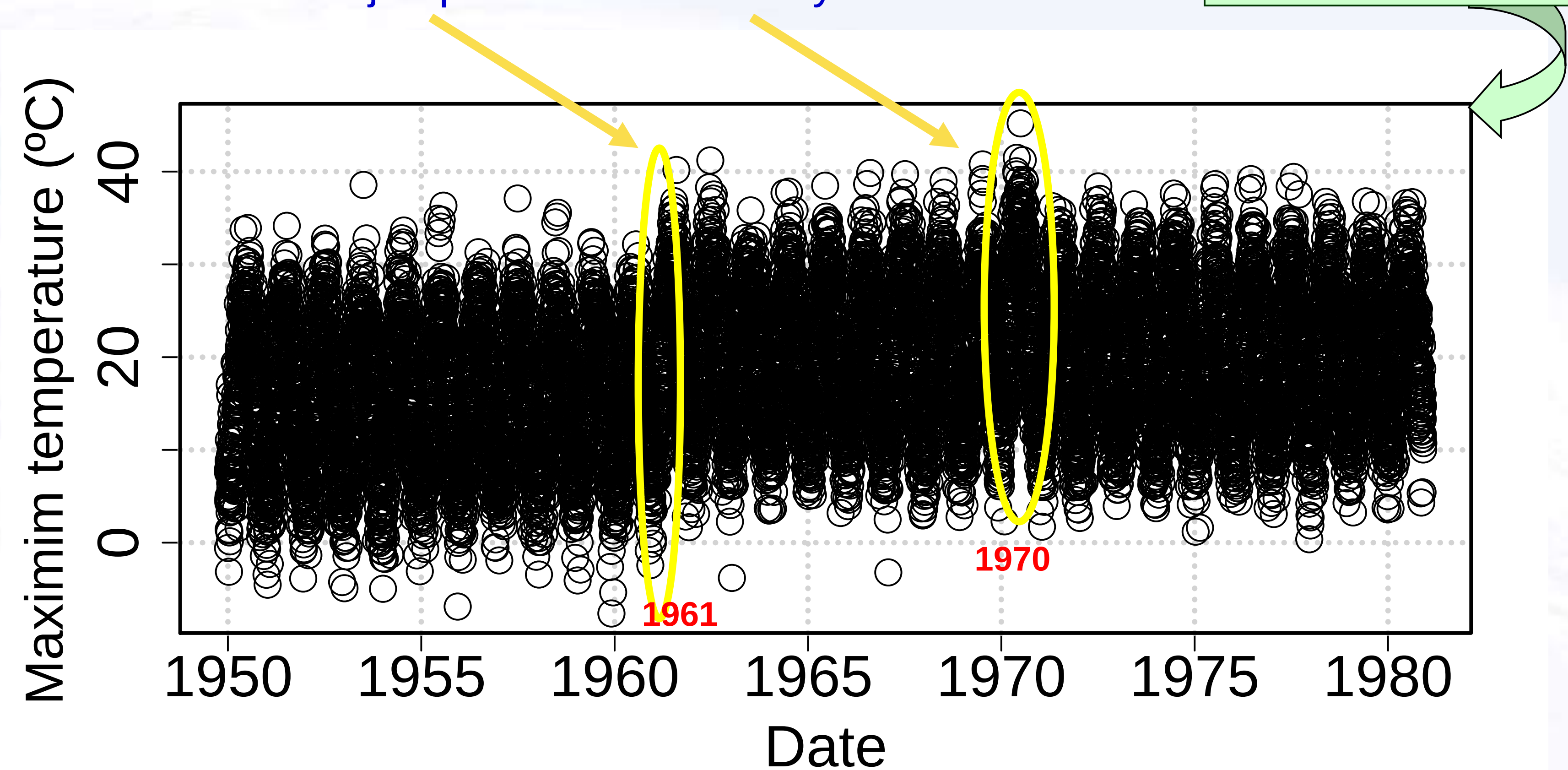


# Methodology

## Inhomogeneity detection: EXAMPLE 1

**Ideal case** with a jump and one unusual year

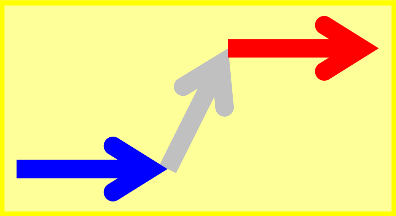
Theoretical case built from normal distributions



# Methodology

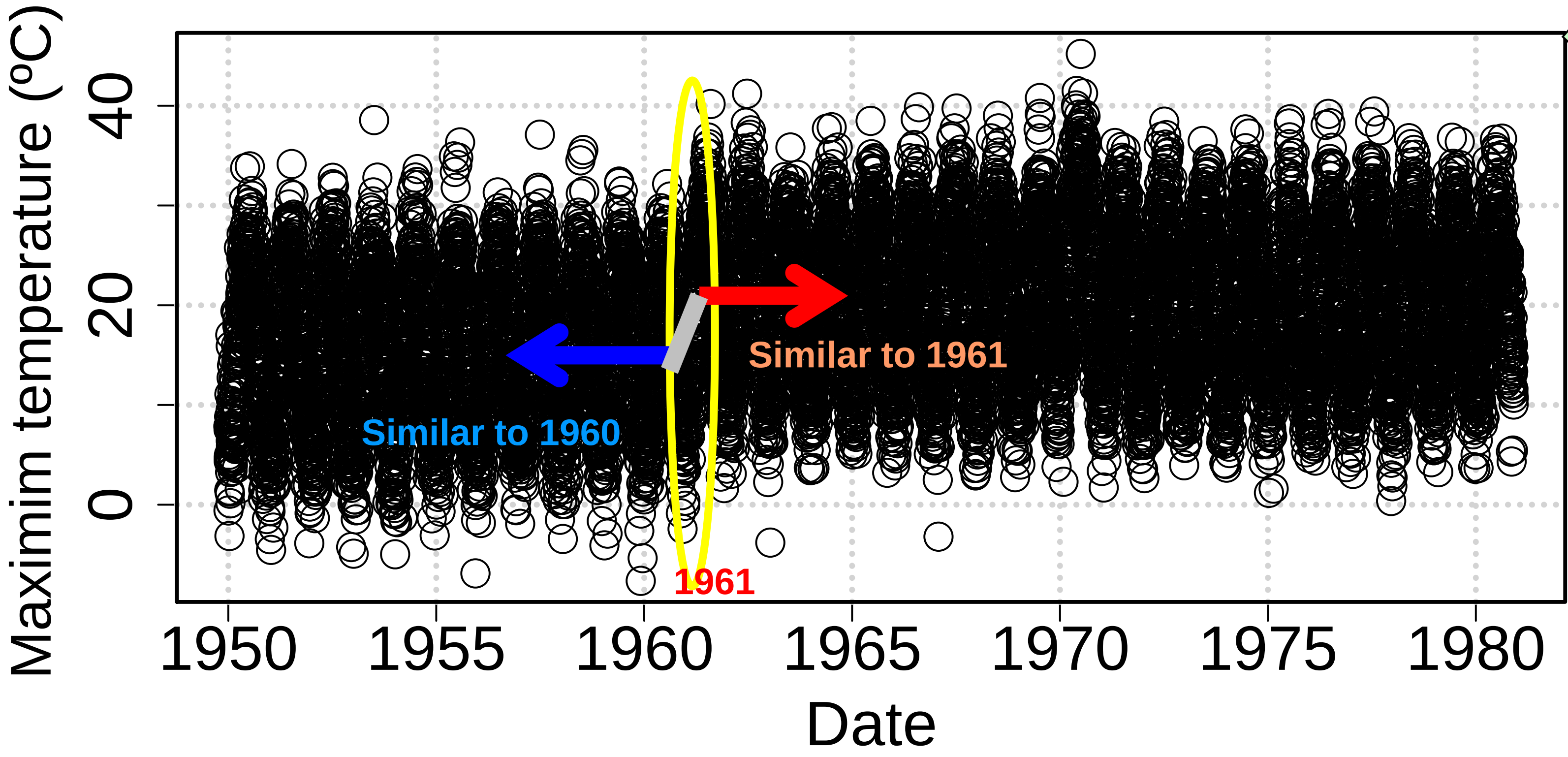


## Inhomogeneity de… …MPLE 1



**Ideal case** of the **jump**

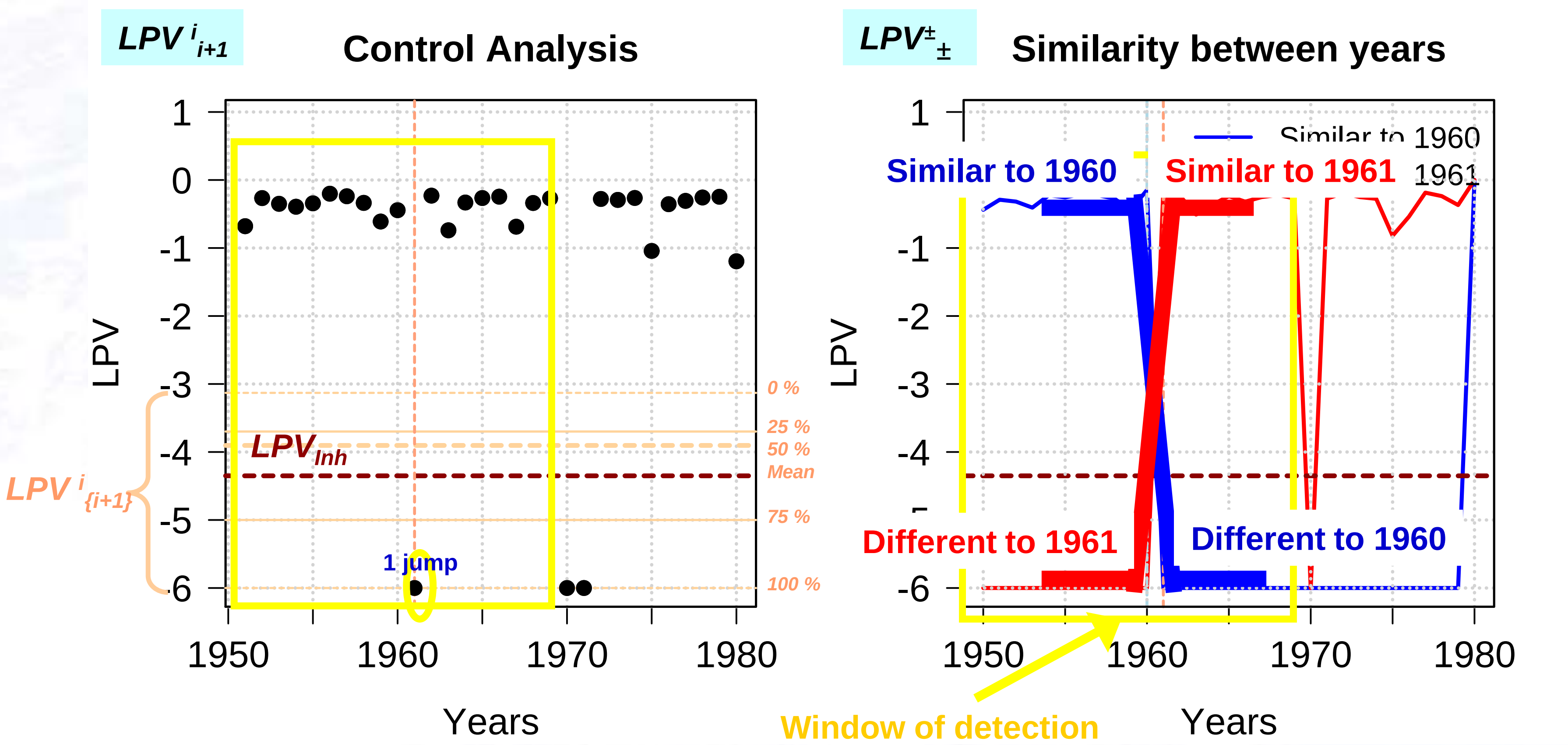
Theoretical case built from normal distributions



# Methodology

## Inhomogeneity detection: EXAMPLE 1

LPV diagrams for the ideal case of **jump**



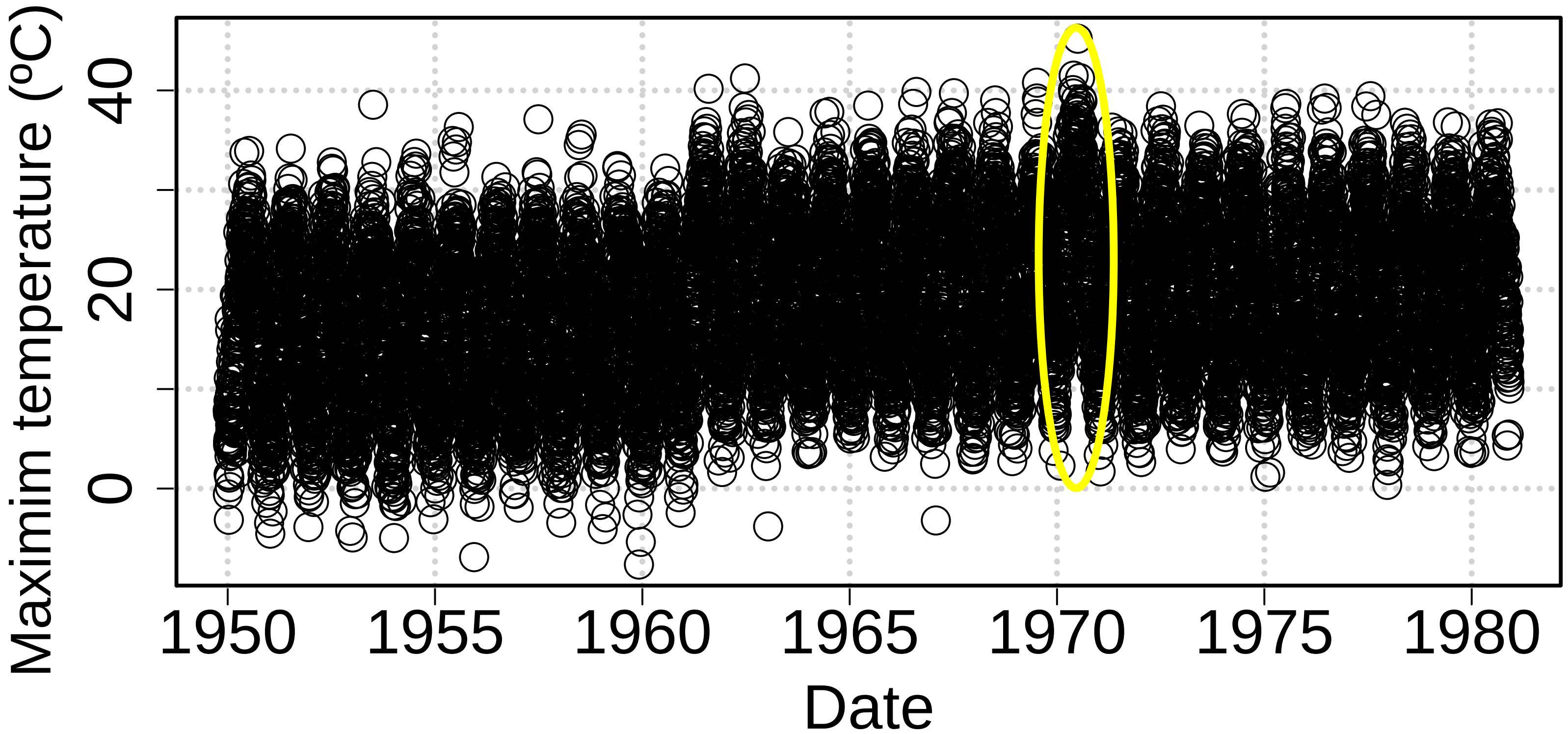
# Methodology



## Inhomogeneity detection: EXAMPLE 1

Theoretical case built from normal distributions

Ideal case of **unusual year** (i.e., 2 jumps)

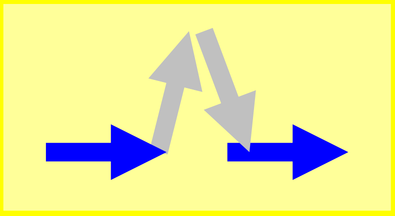


# Methodology

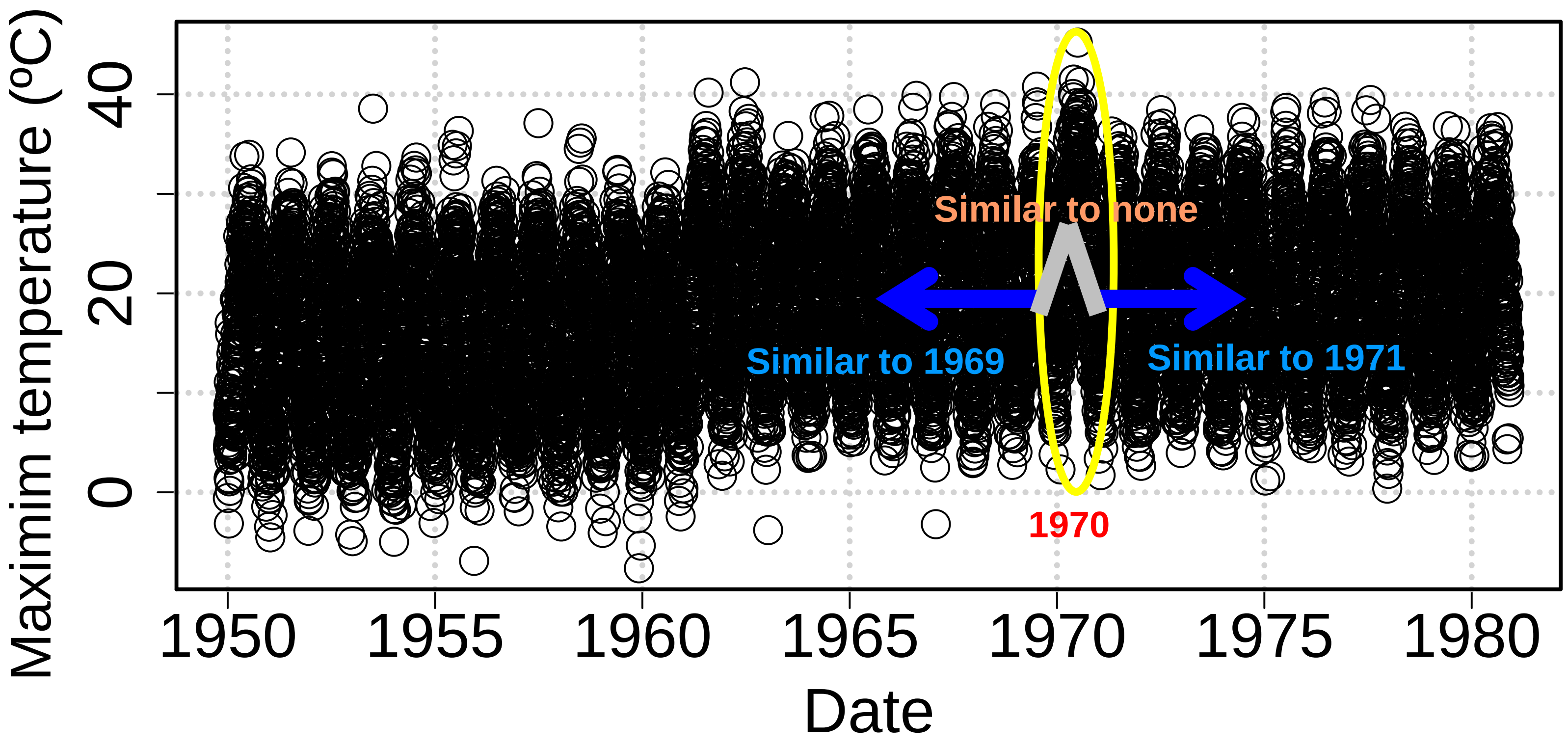


## Inhomogeneity detection: EX

Ideal case of **unusual year** (i.e., 2 jumps)



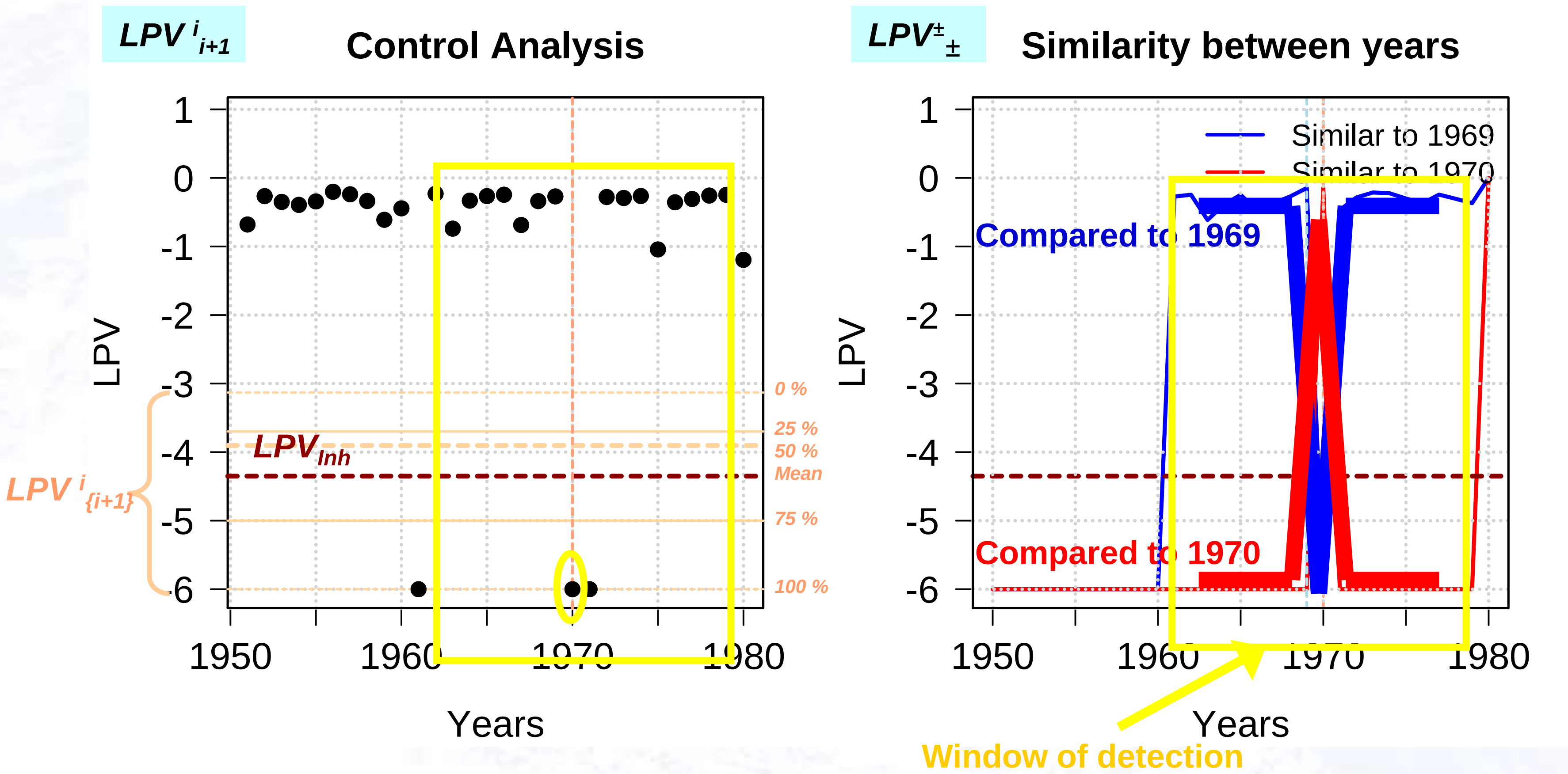
Theoretical case built from normal distributions



# Methodology

## Inhomogeneity detection: EXAMPLE 1

LPV diagrams for the ideal case of **unusual year** (2 jumps)



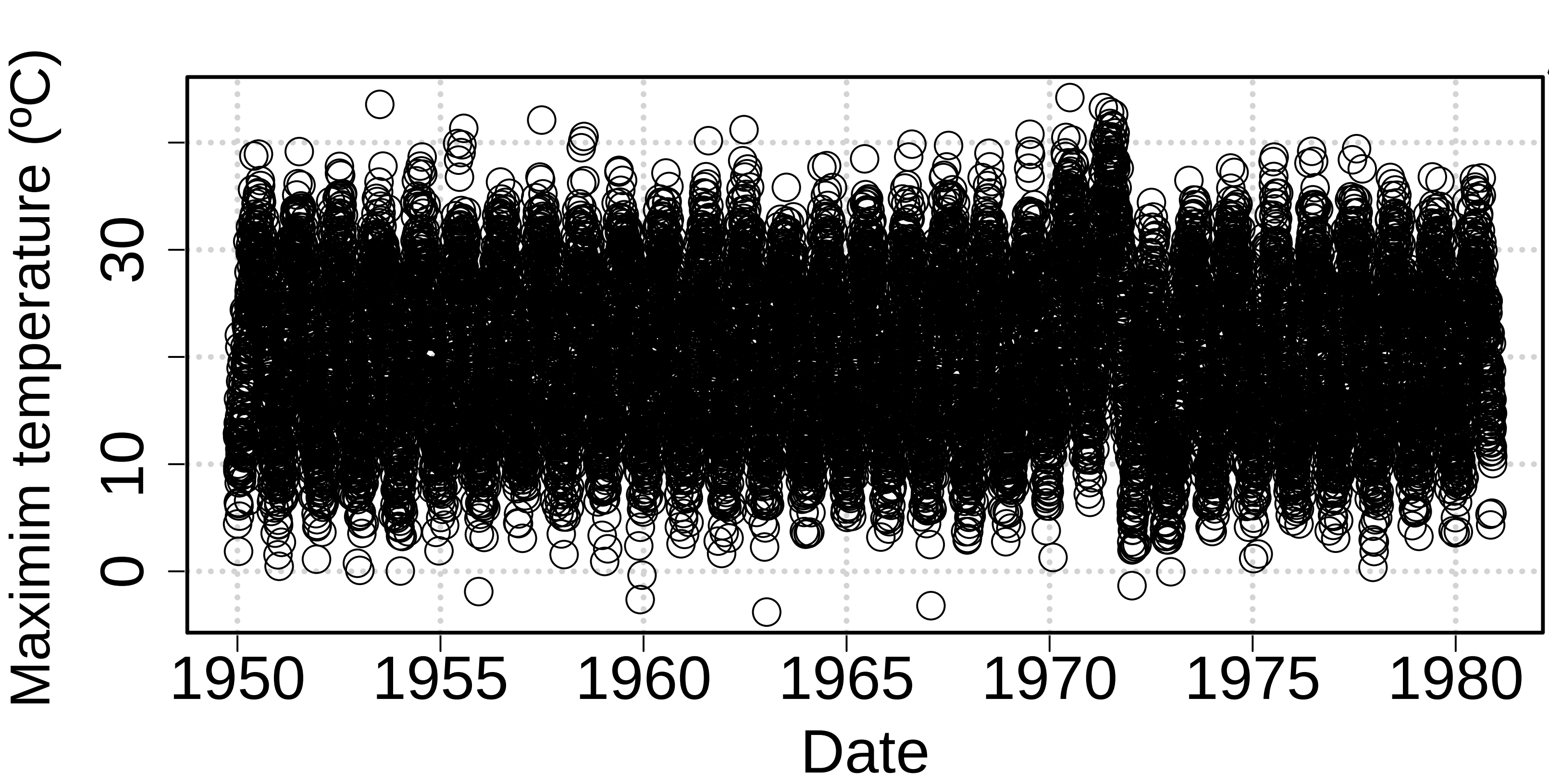
# Methodology



## Inhomogeneity detection: EXAMPLE 2

Theoretical case built from normal distributions

Ideal case of inhomogeneity of **second order** (i.e., more of 2 jumps)



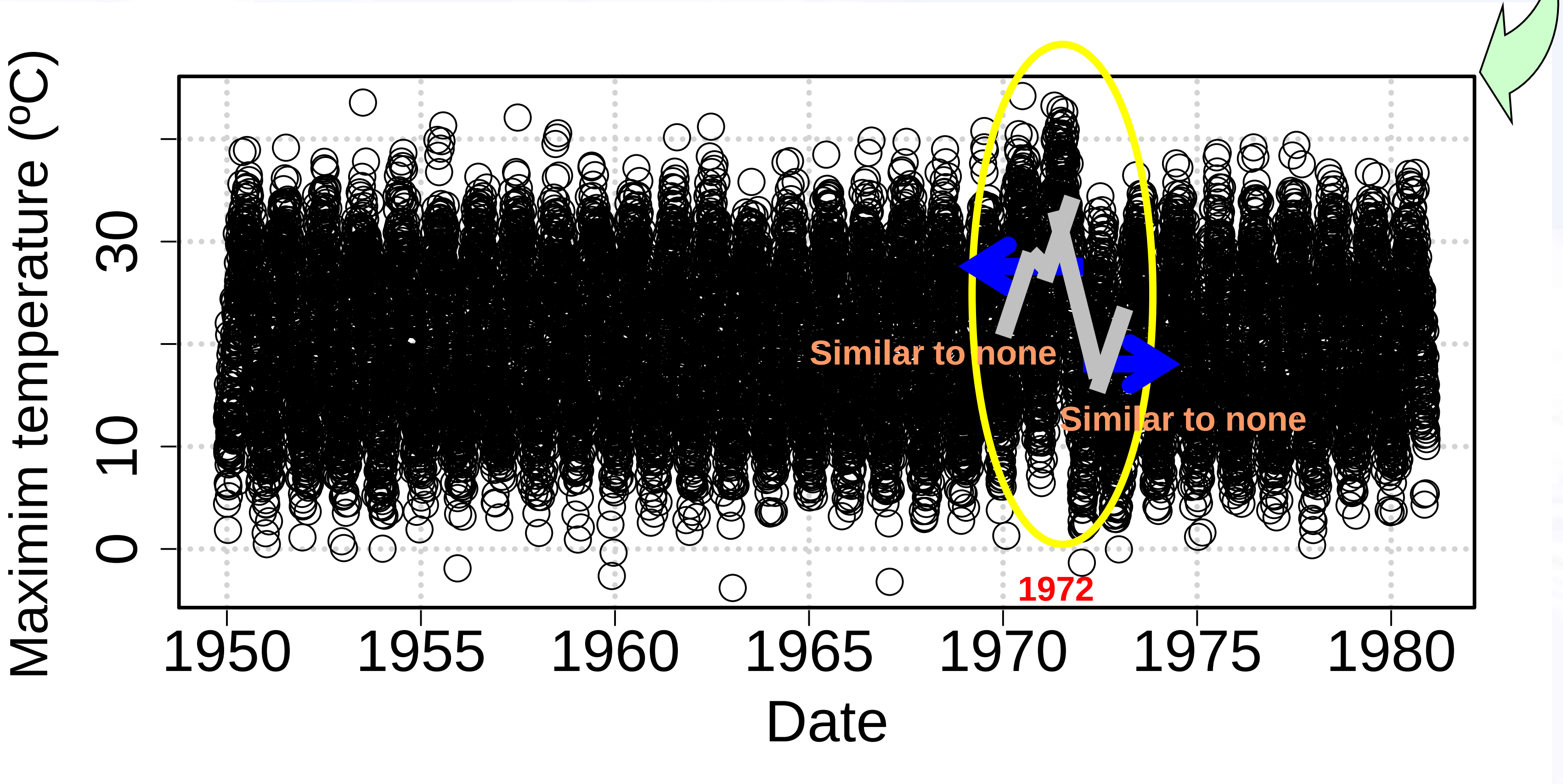
# Methodology



## Inhomogeneity detection: EXAMPLE 2

Theoretical case built from normal distributions

Ideal case of inhomogeneity of **second order** (i.e., more of 2 jumps)

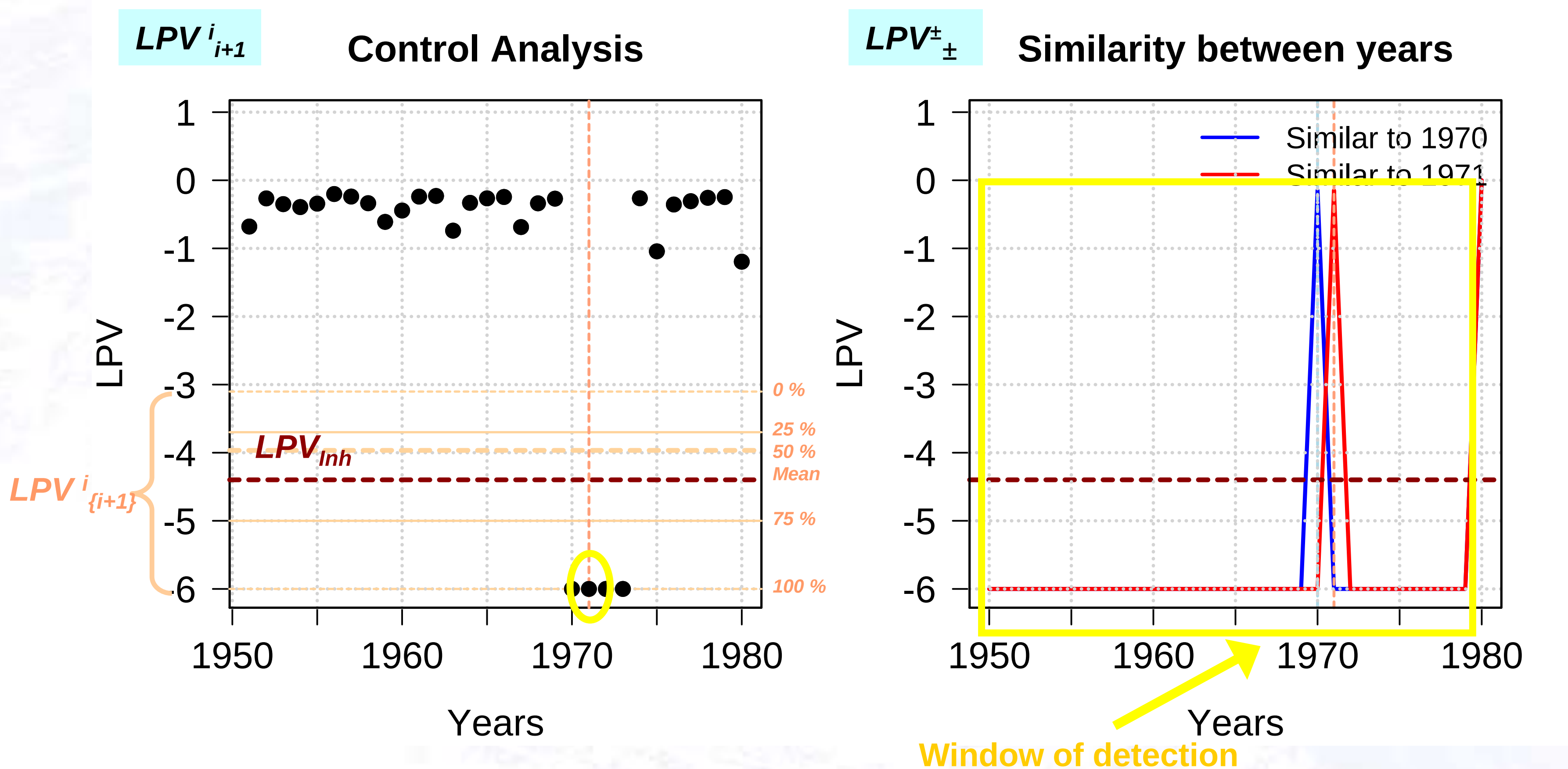


# Methodology



## Inhomogeneity detection: EXAMPLE 2

LPV diagrams for the ideal case of inhomogeneity of **second order**

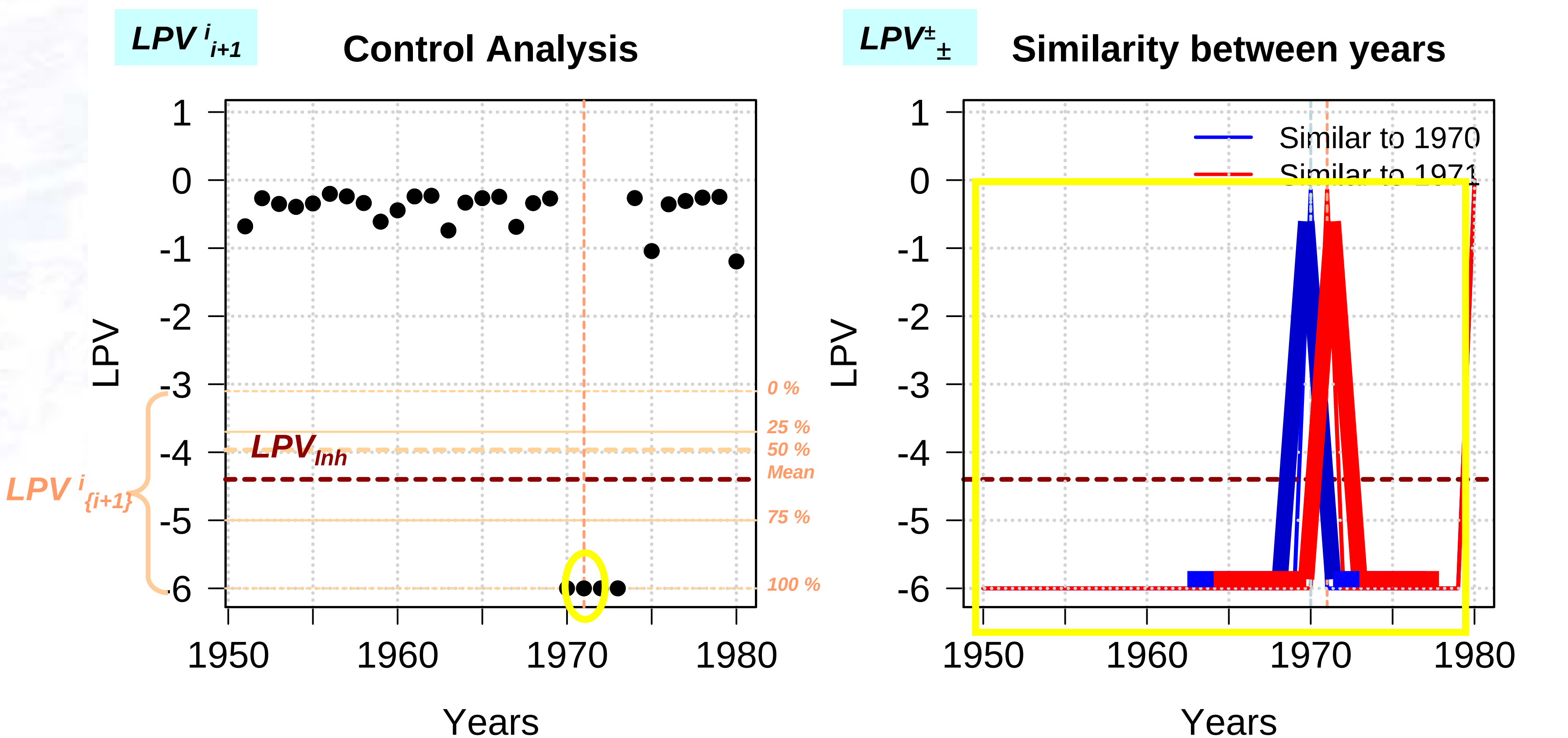


# Methodology



## Inhomogeneity detection: EXAMPLE 2

LPV diagrams for the ideal case of inhomogeneity of **second order**



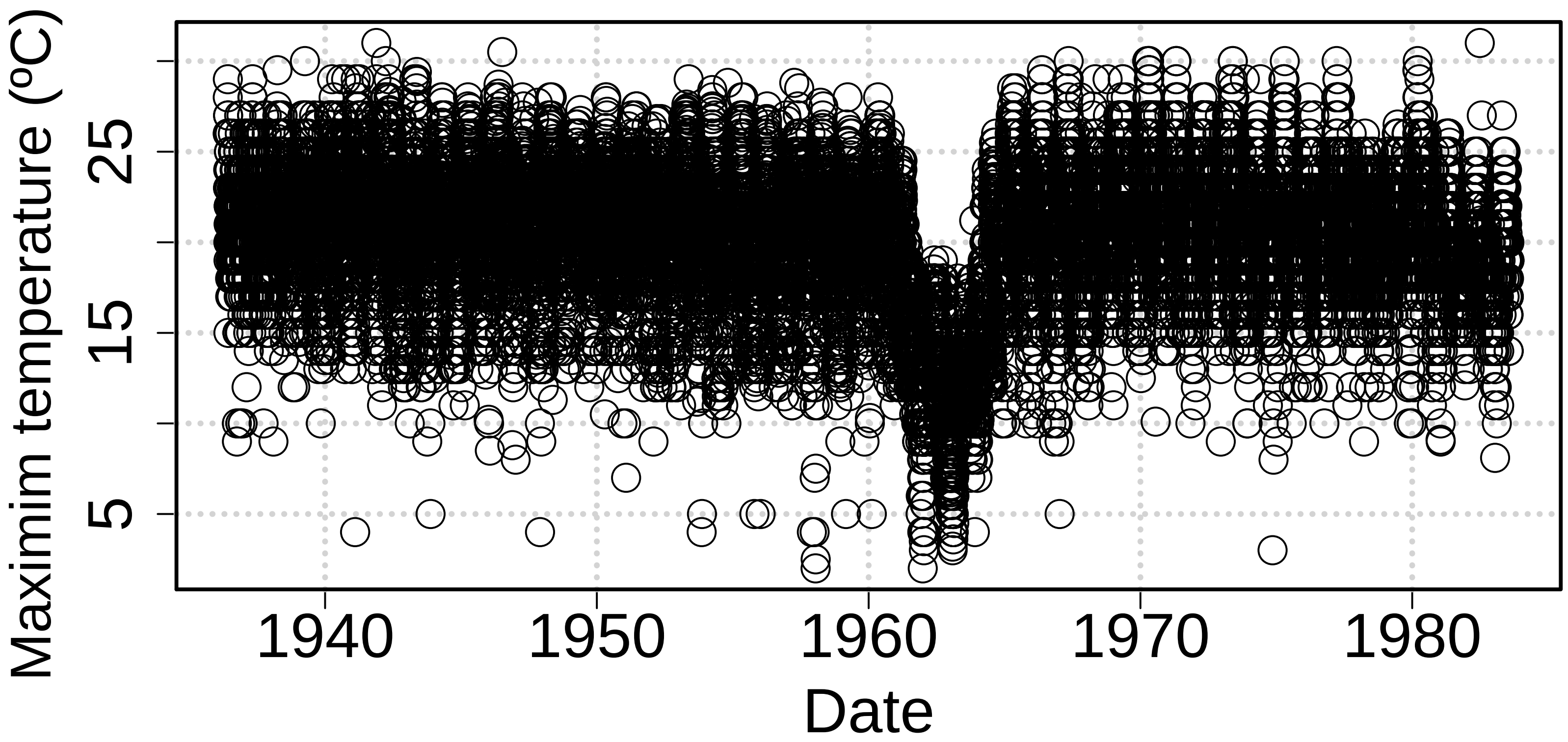
# Contents

# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**

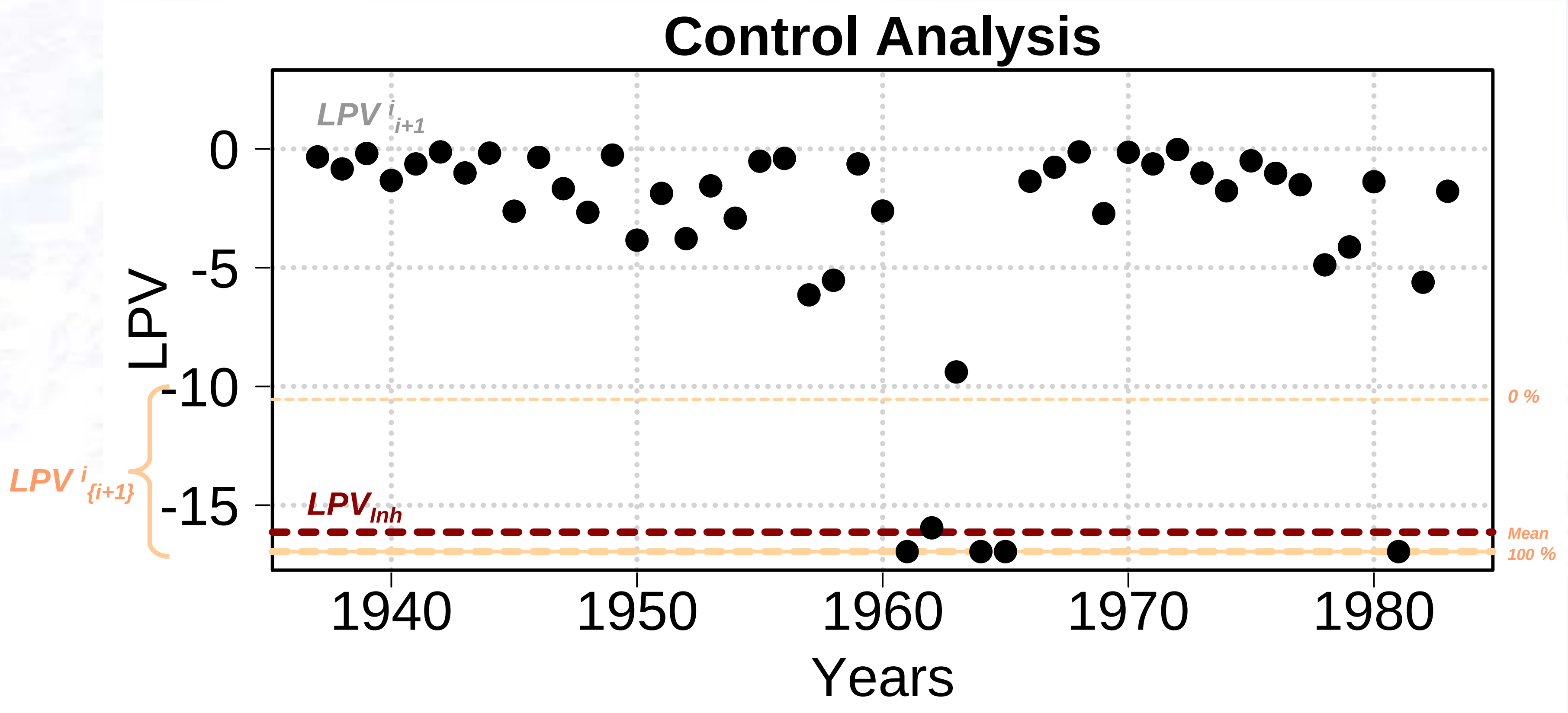


# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**: LPV diagrams

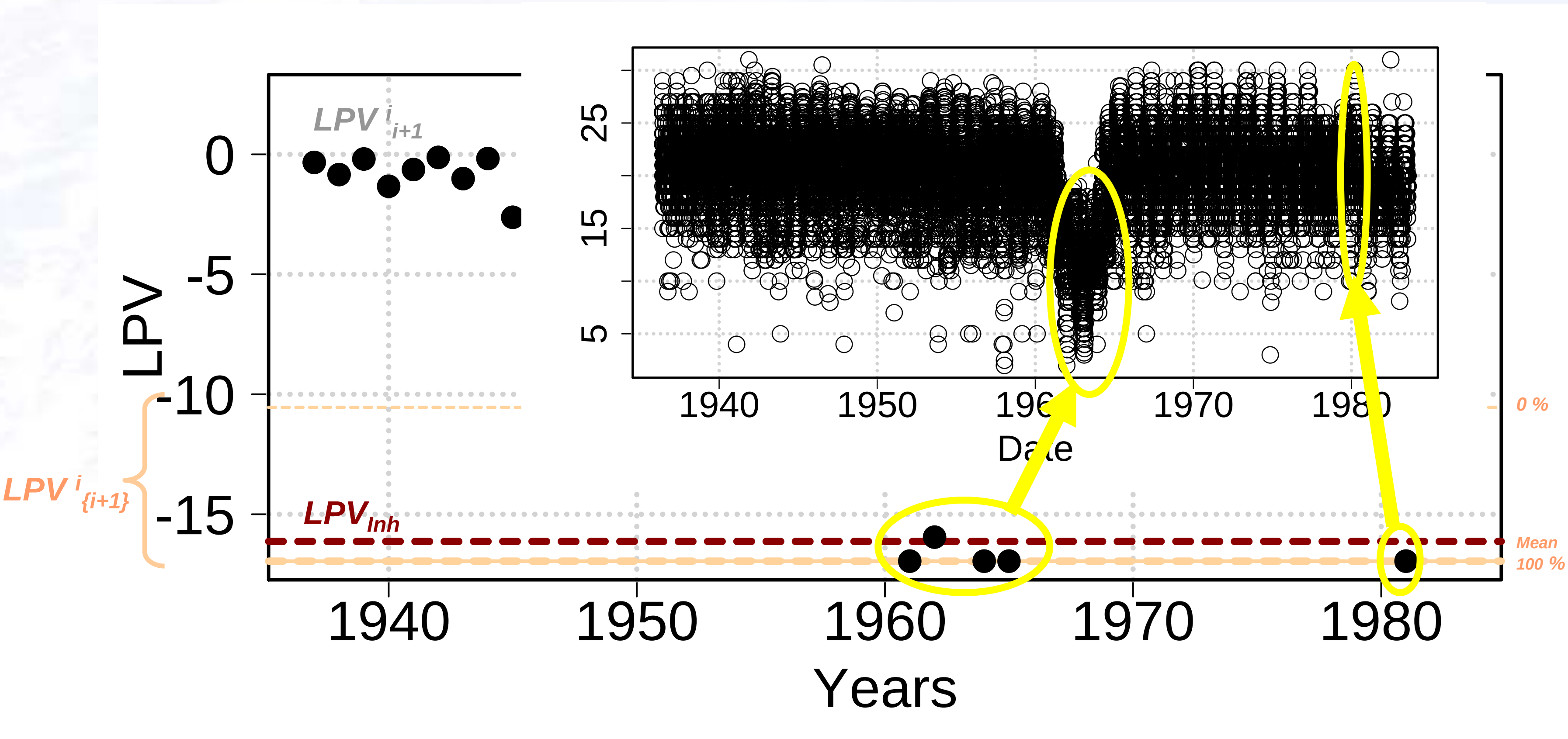


# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**: LPV diagrams

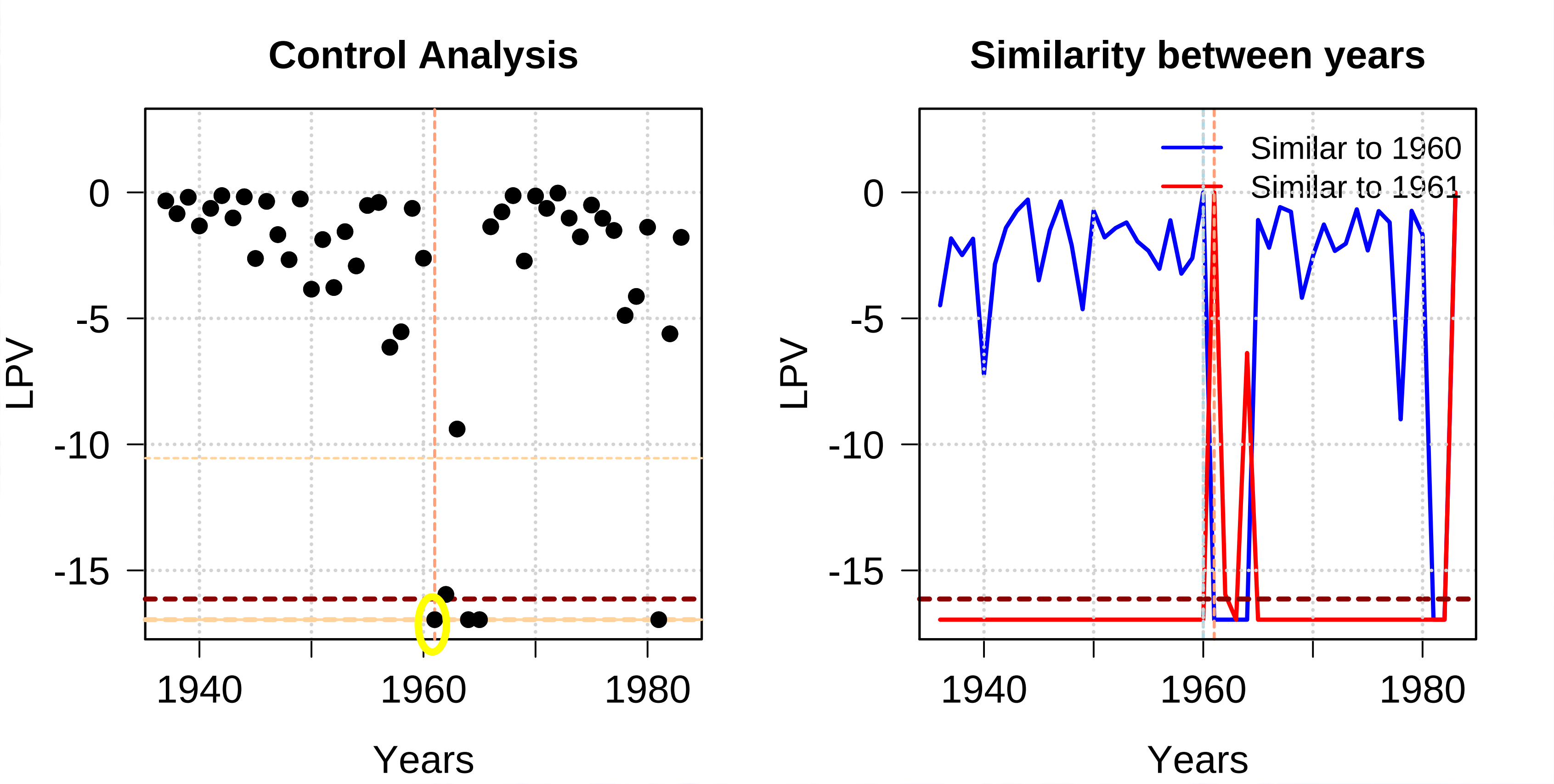


# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**: LPV diagrams

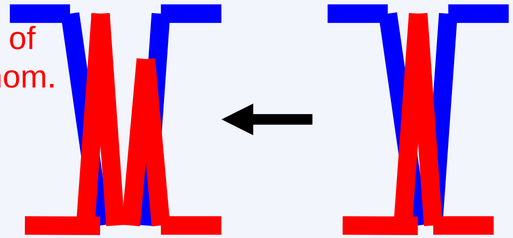


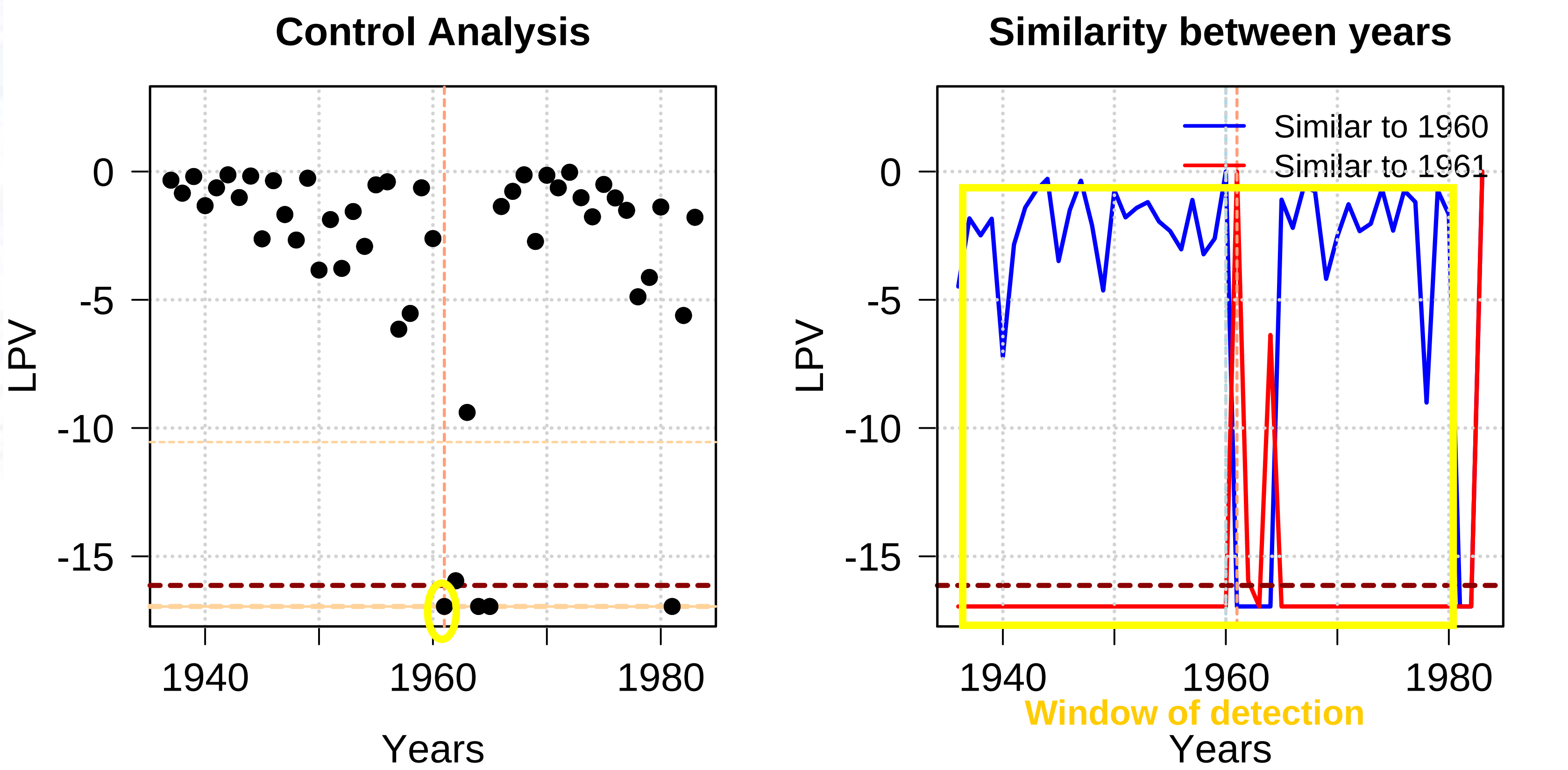
# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**: LPV diagrams



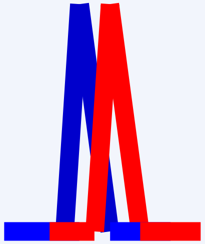



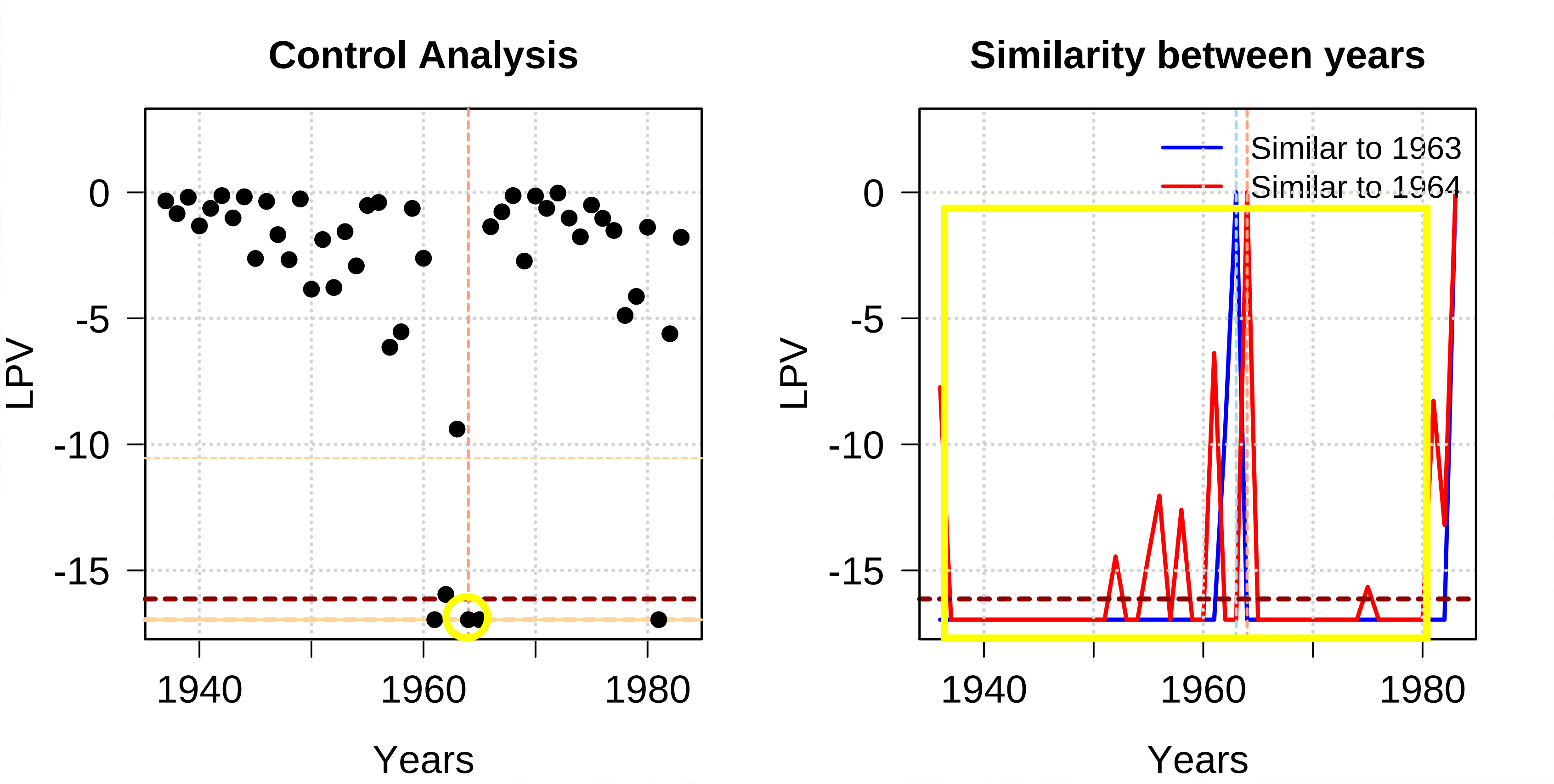
# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**: LPV diagrams



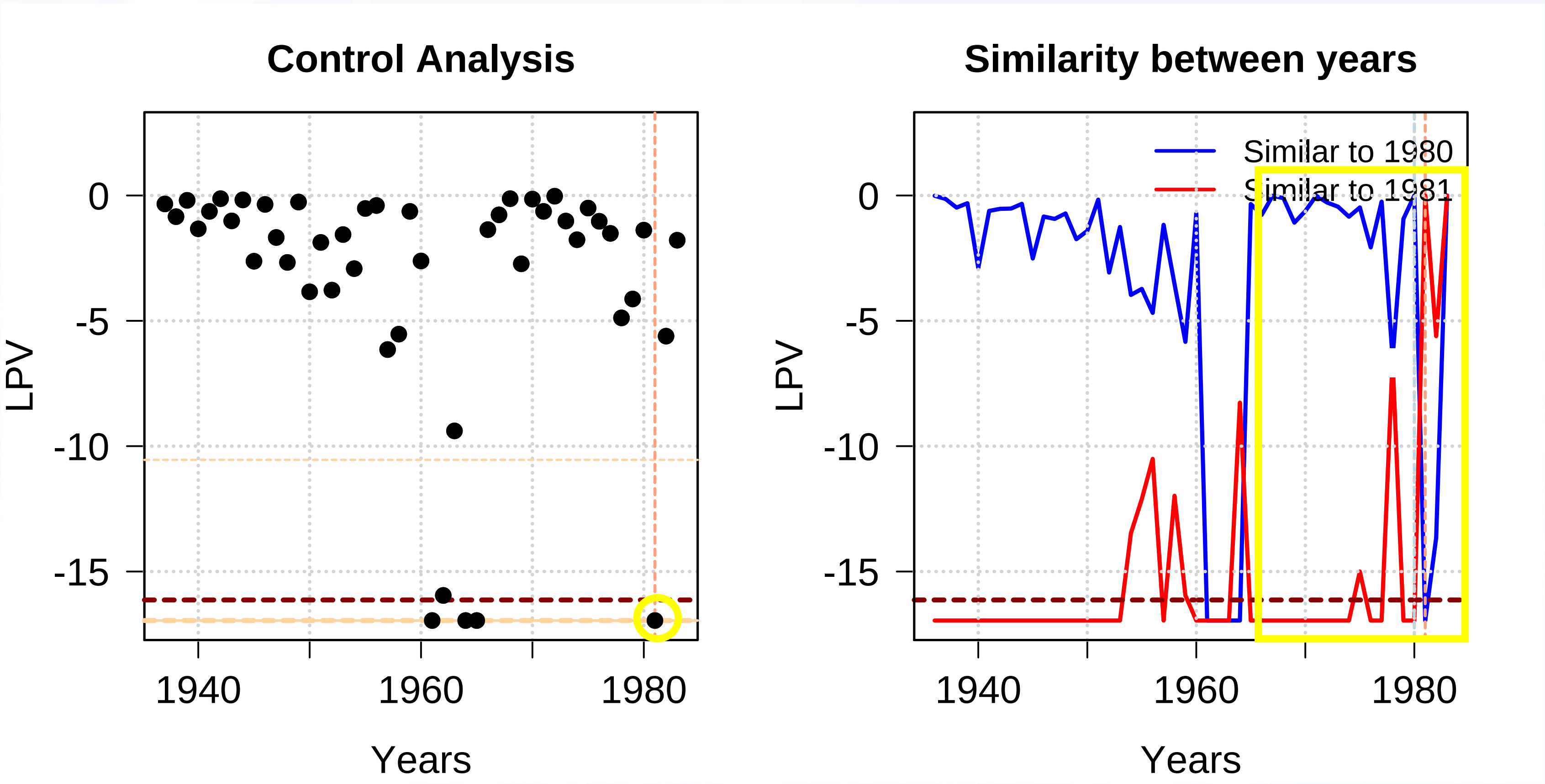


# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**: LPV diagrams

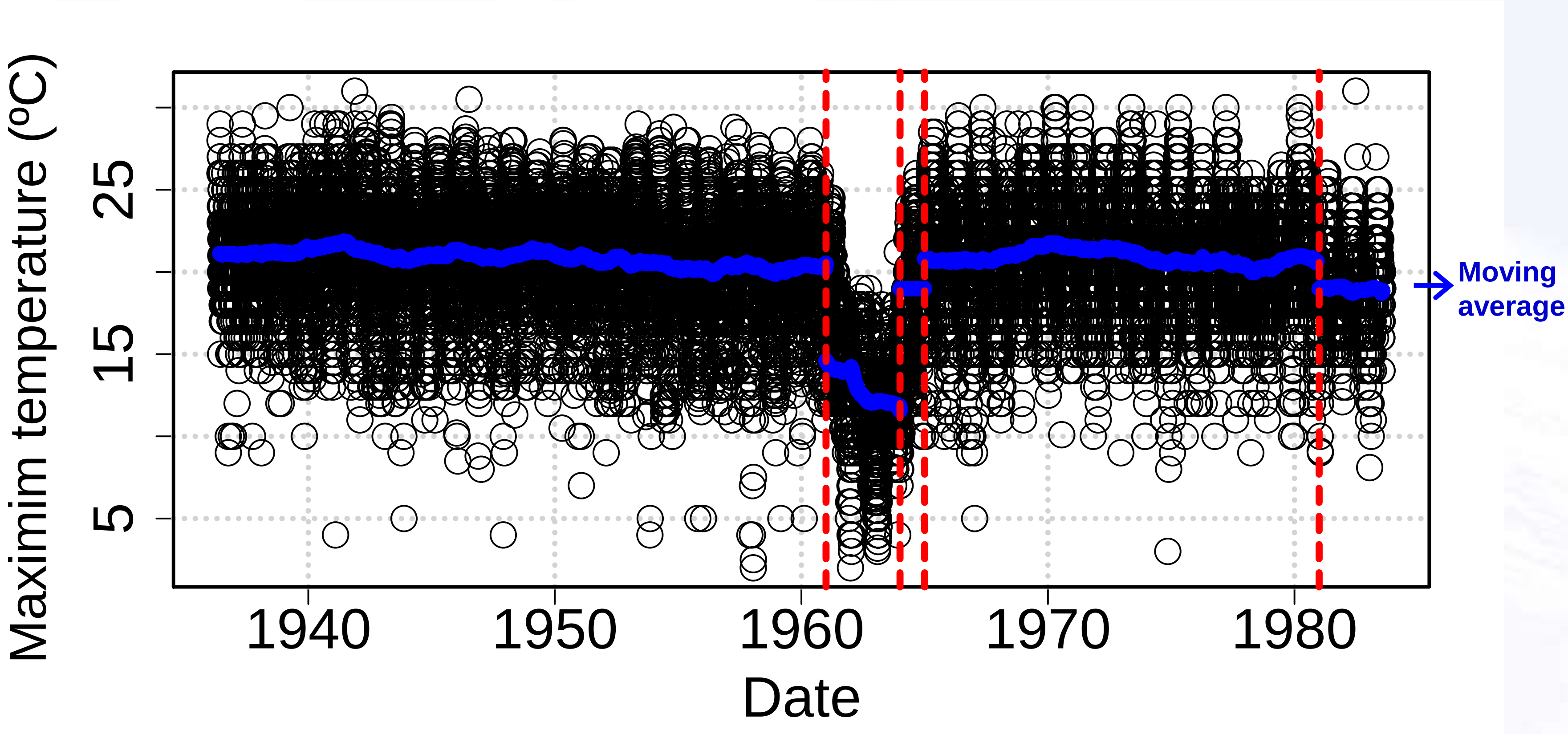


# Real cases



## Temperature: CASE 1

Inhomogeneities in **temperature**: final analysis

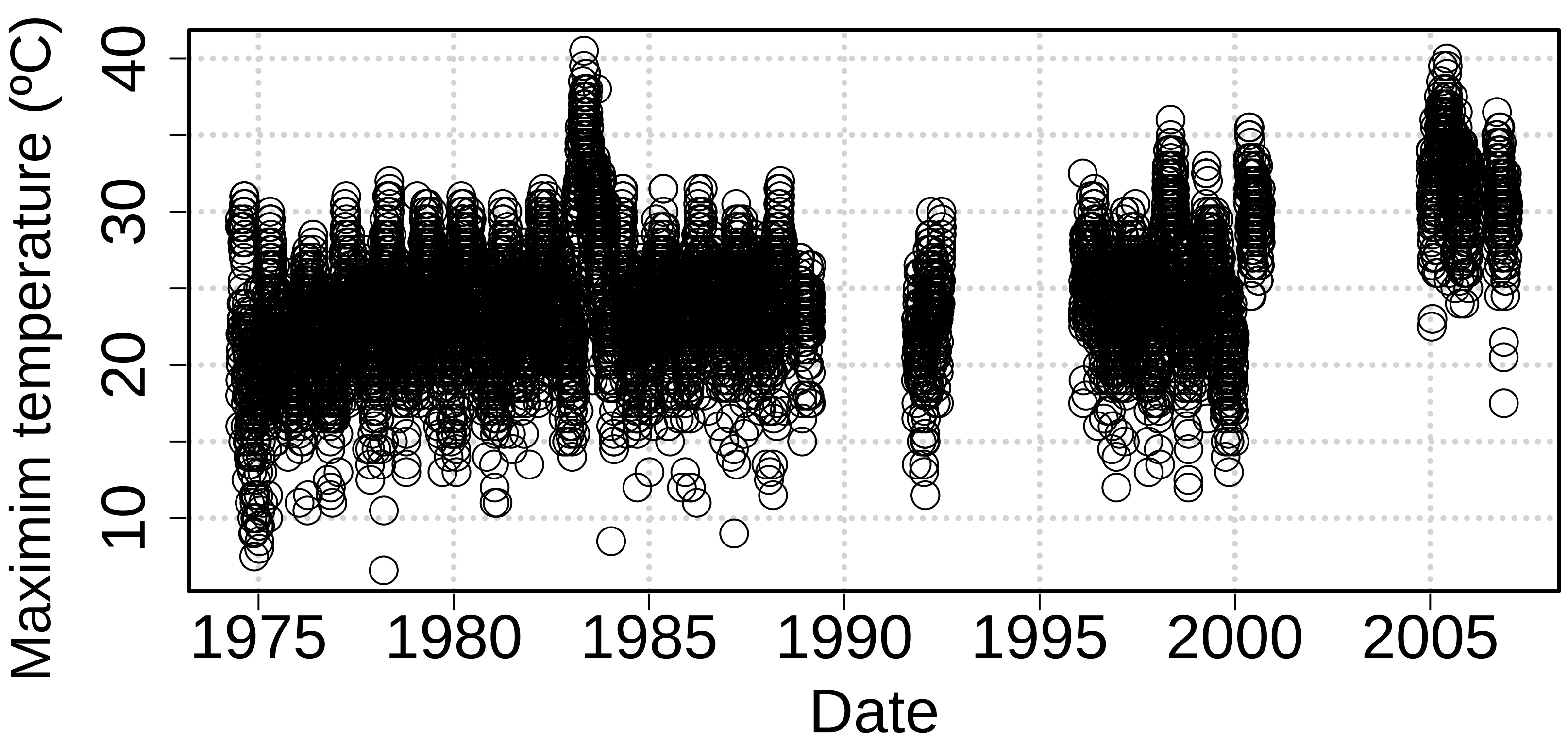


# Real cases



## Temperature: CASE 2

Inhomogeneities in **temperature**

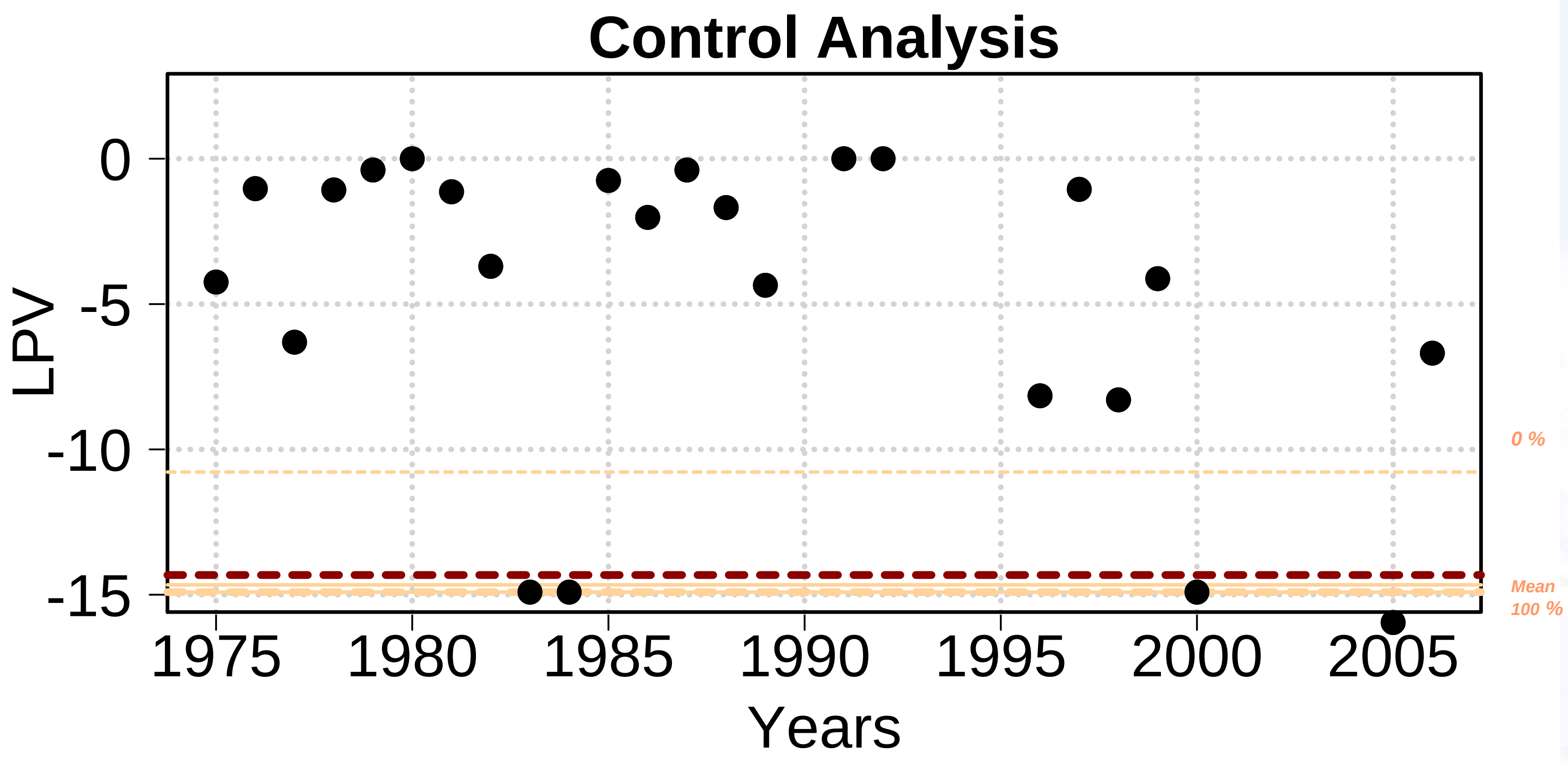


# Real cases



## Temperature: CASE 2

Inhomogeneities in **temperature**: LPV diagrams

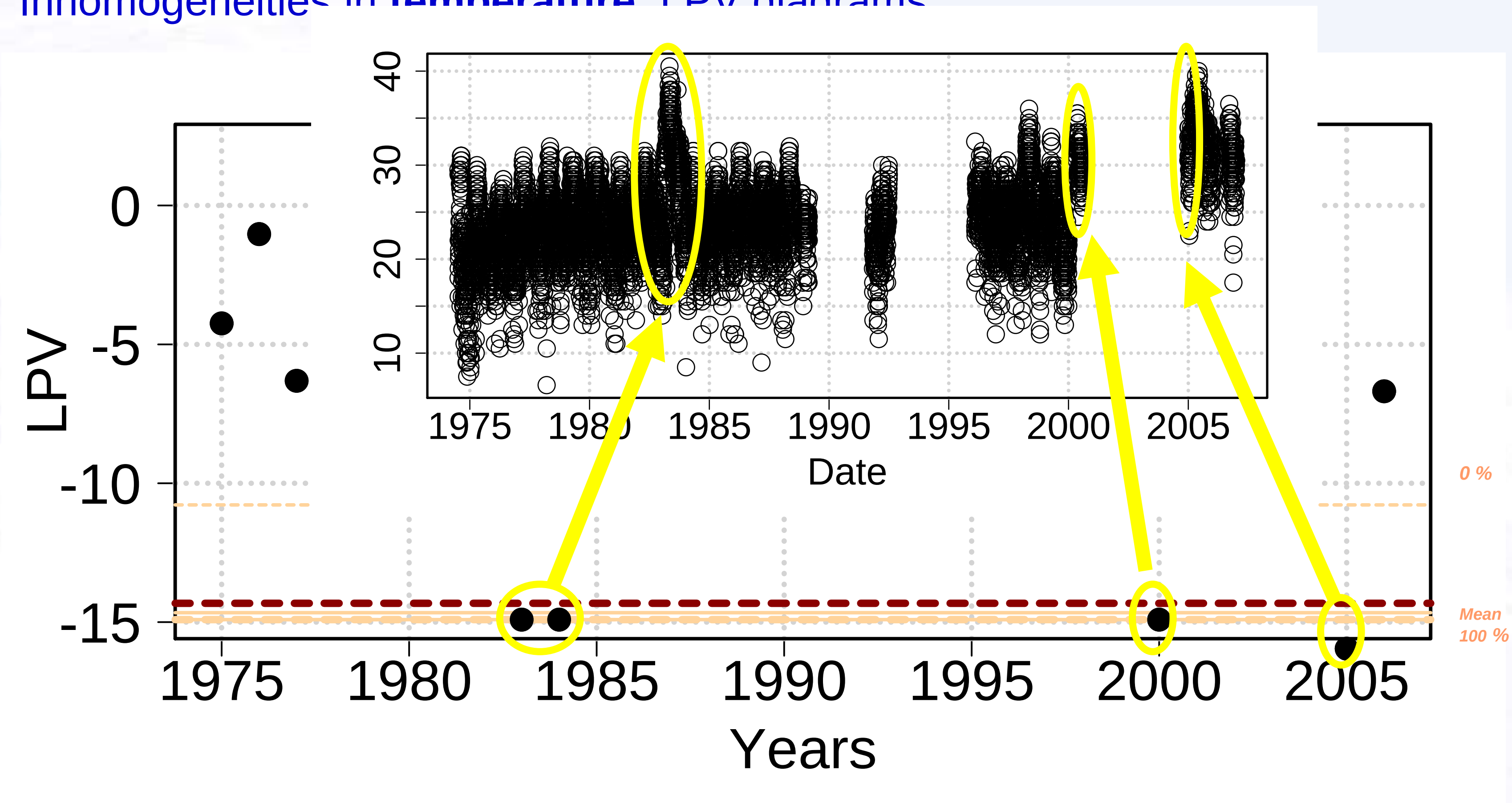


# Real cases



## Temperature: CASE 2

Inhomogeneities in **temperature**: LPV diagrams

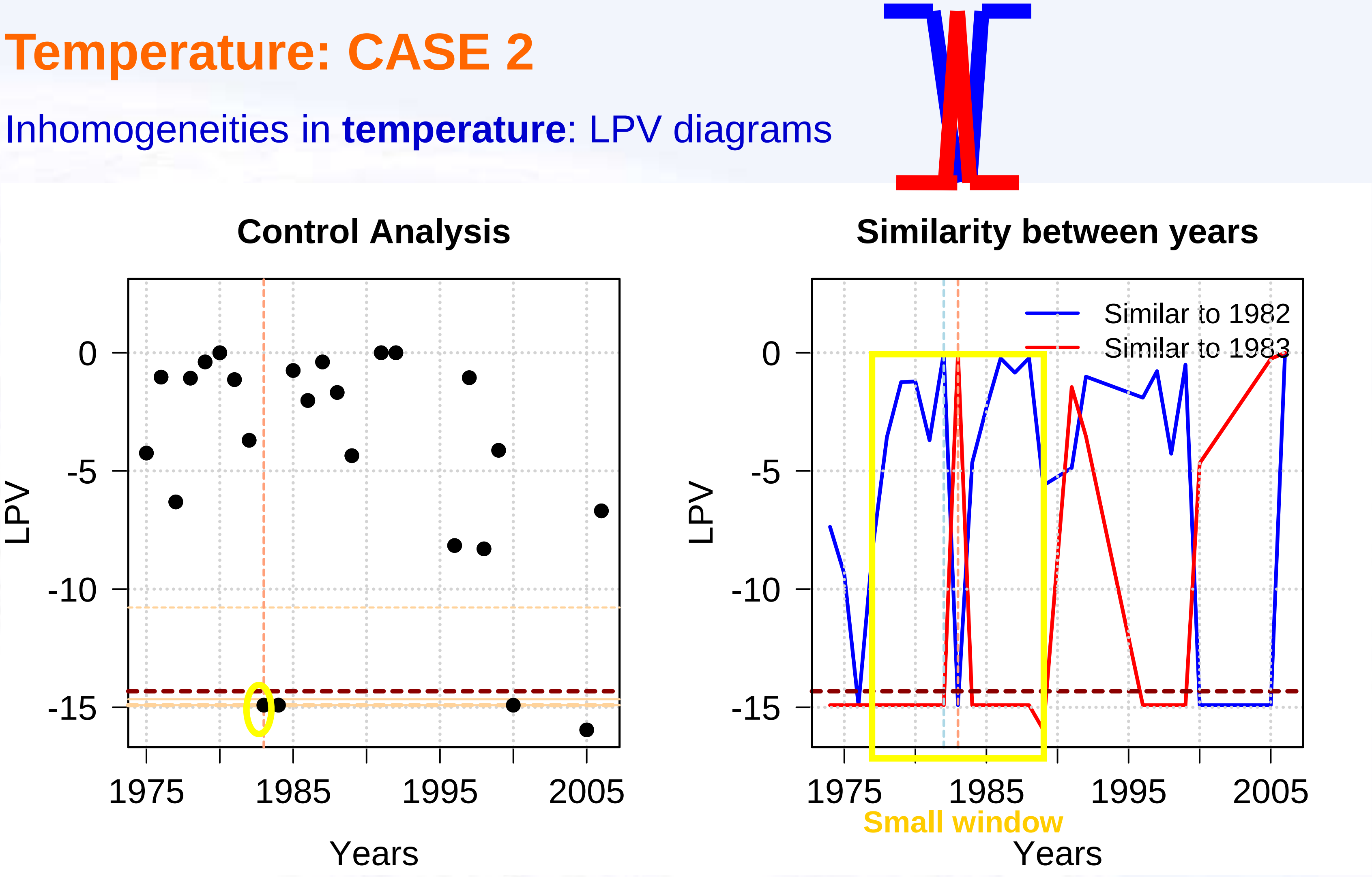


# Real cases



## Temperature: CASE 2

Inhomogeneities in **temperature**: LPV diagrams

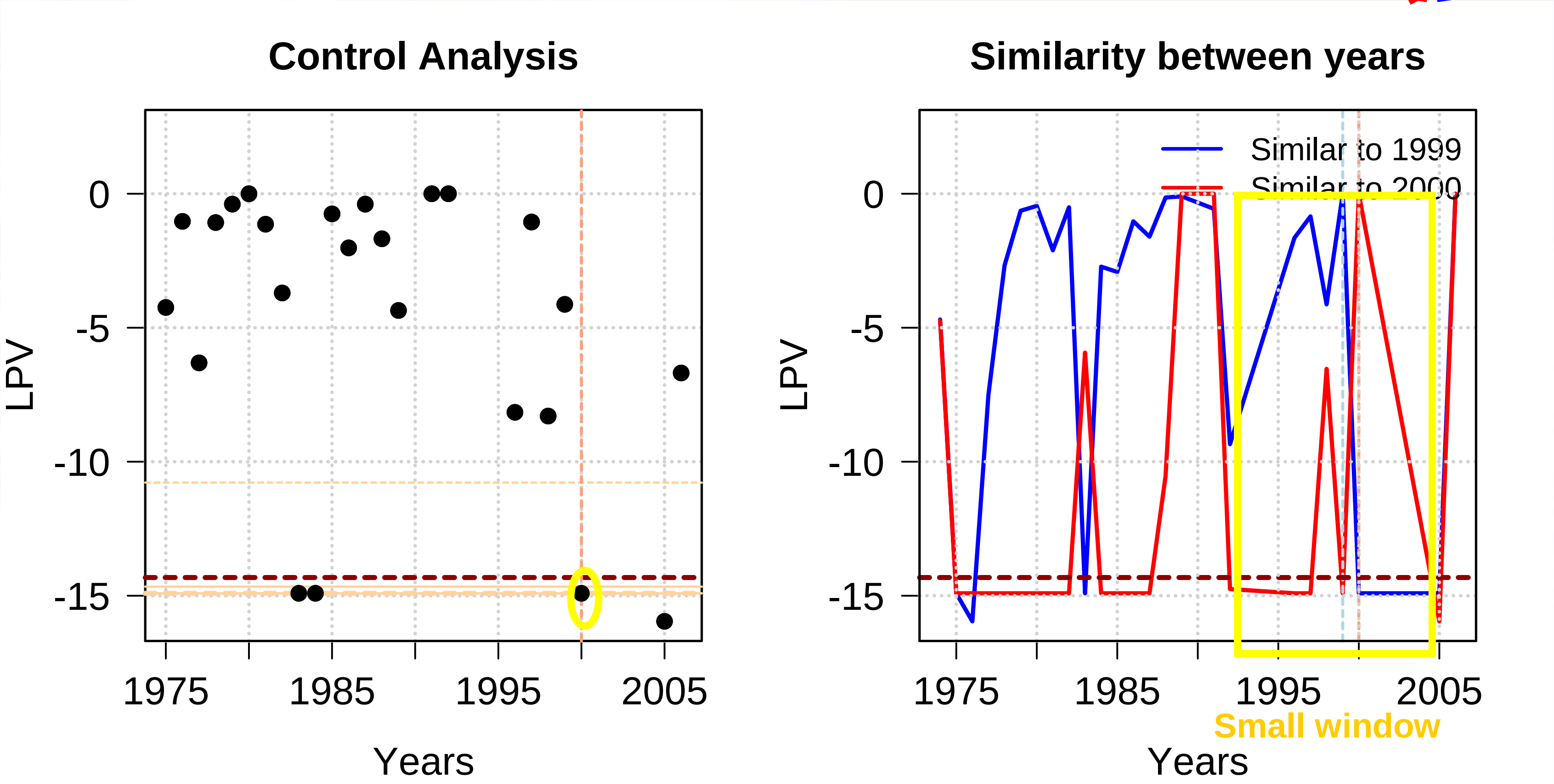


# Real cases



## Temperature: CASE 2

Inhomogeneities in **temperature**: LPV diagrams

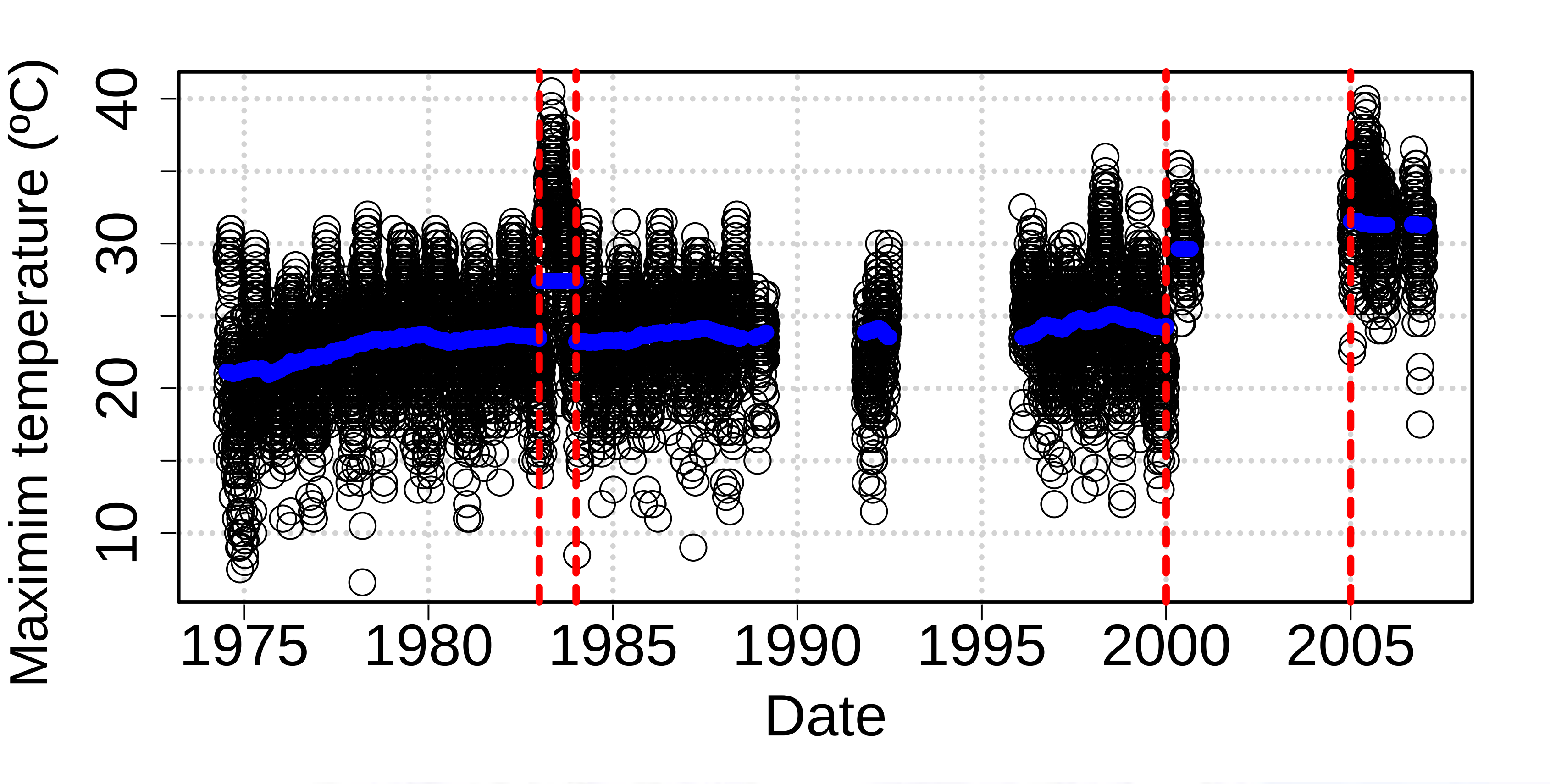


# Real cases



## Temperature: CASE 2

Inhomogeneities in **temperature**: final analysis

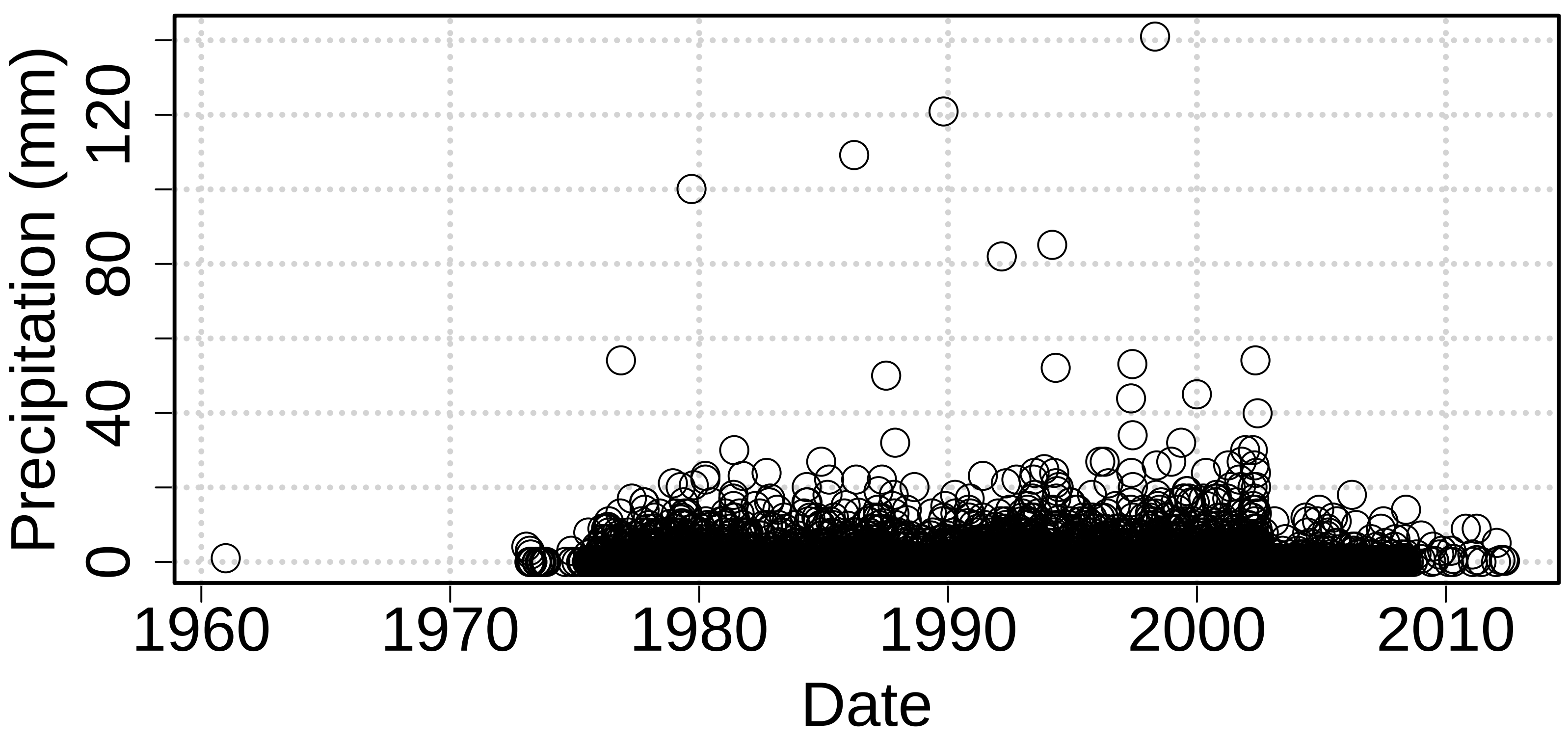


# Real cases



## Precipitation: CASE 1

Inhomogeneities in **precipitation**

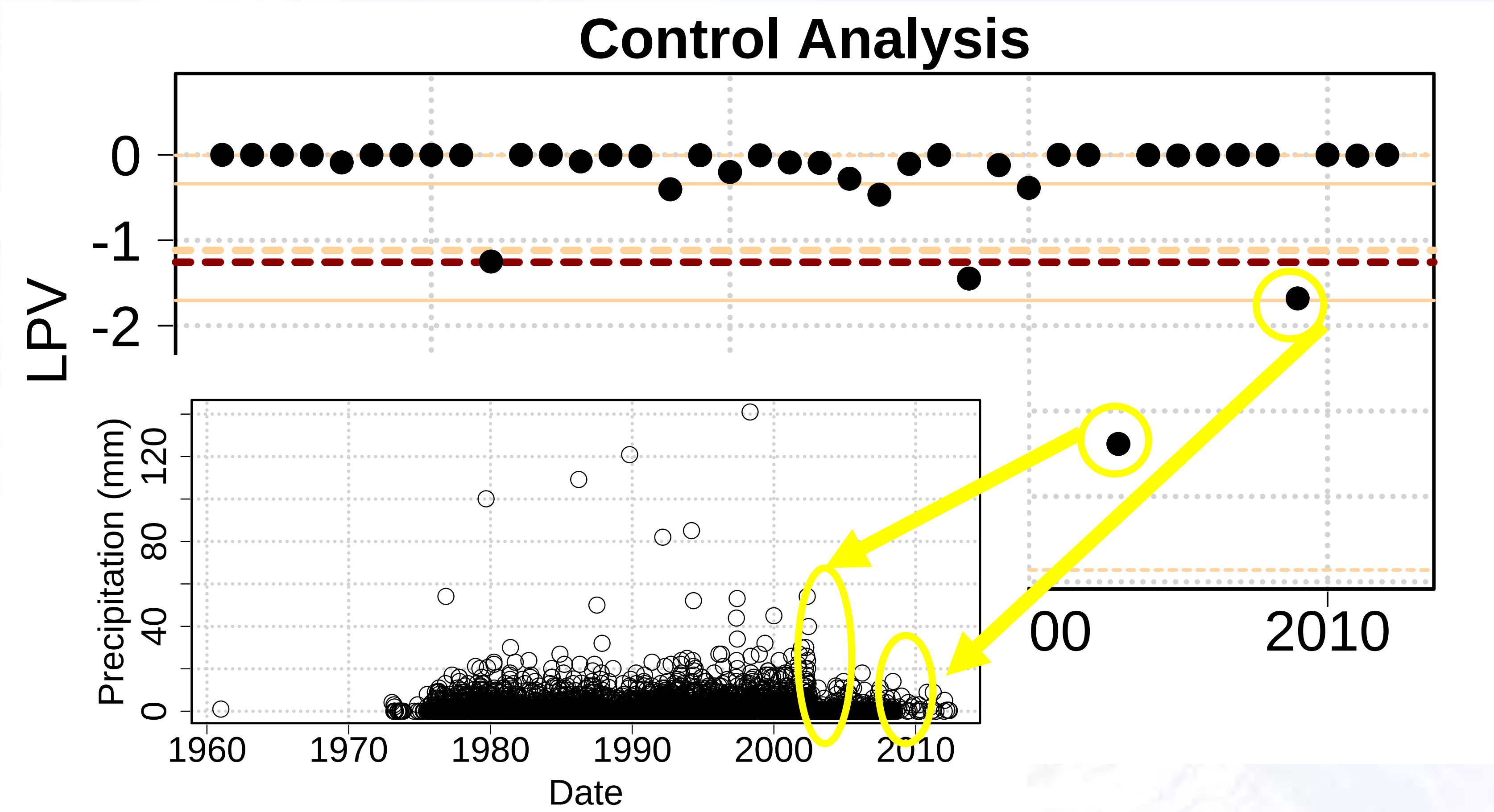


# Real cases



## Precipitation: CASE 1

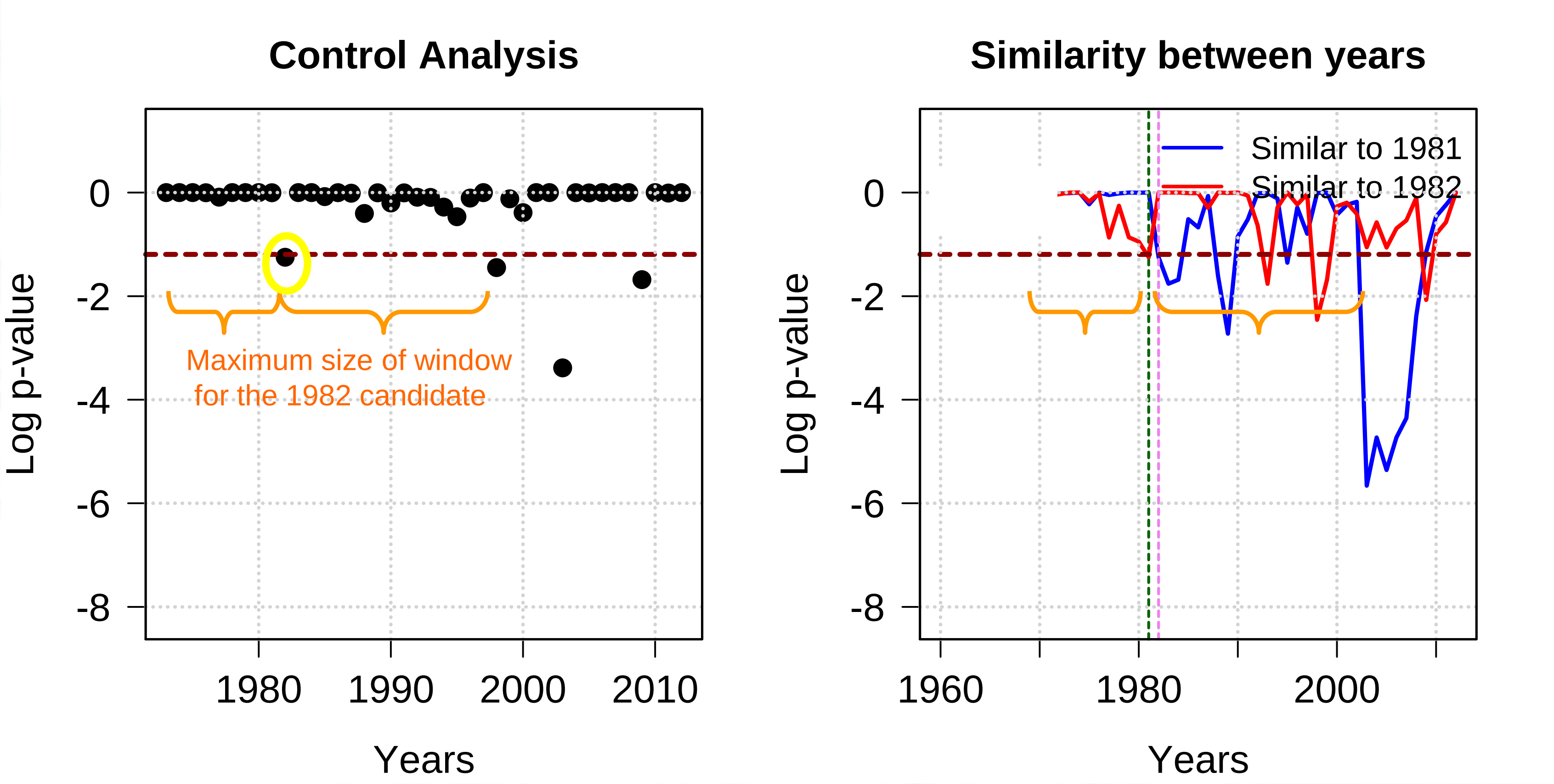
Inhomogeneities in **precipitation**: LPV diagrams



# Real cases

## Precipitation: CASE 1

Inhomogeneities in **precipitation**: LPV diagrams

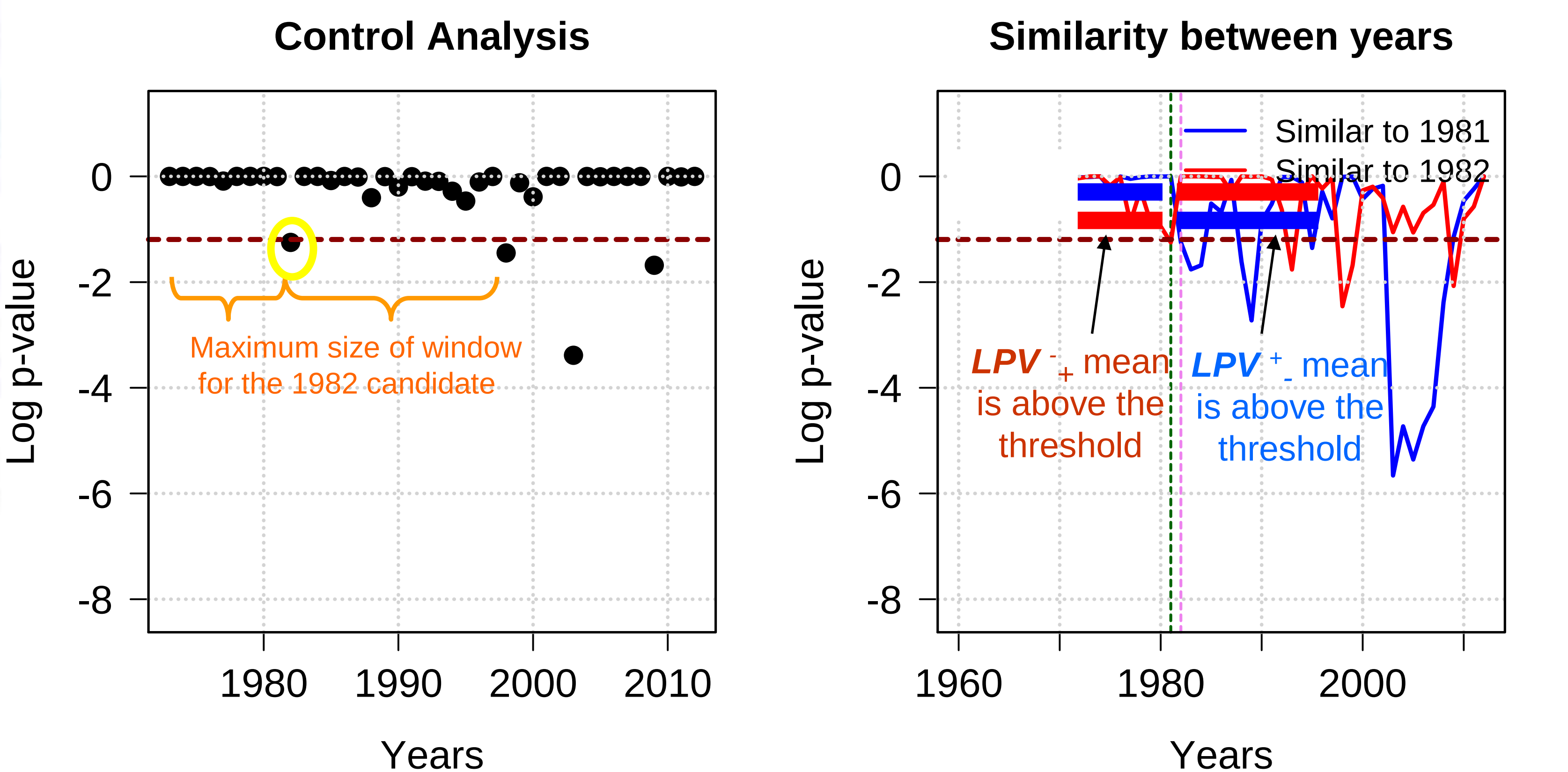


# Real cases



## Precipitation: CASE 1

Inhomogeneities in **precipitation**: LPV diagrams

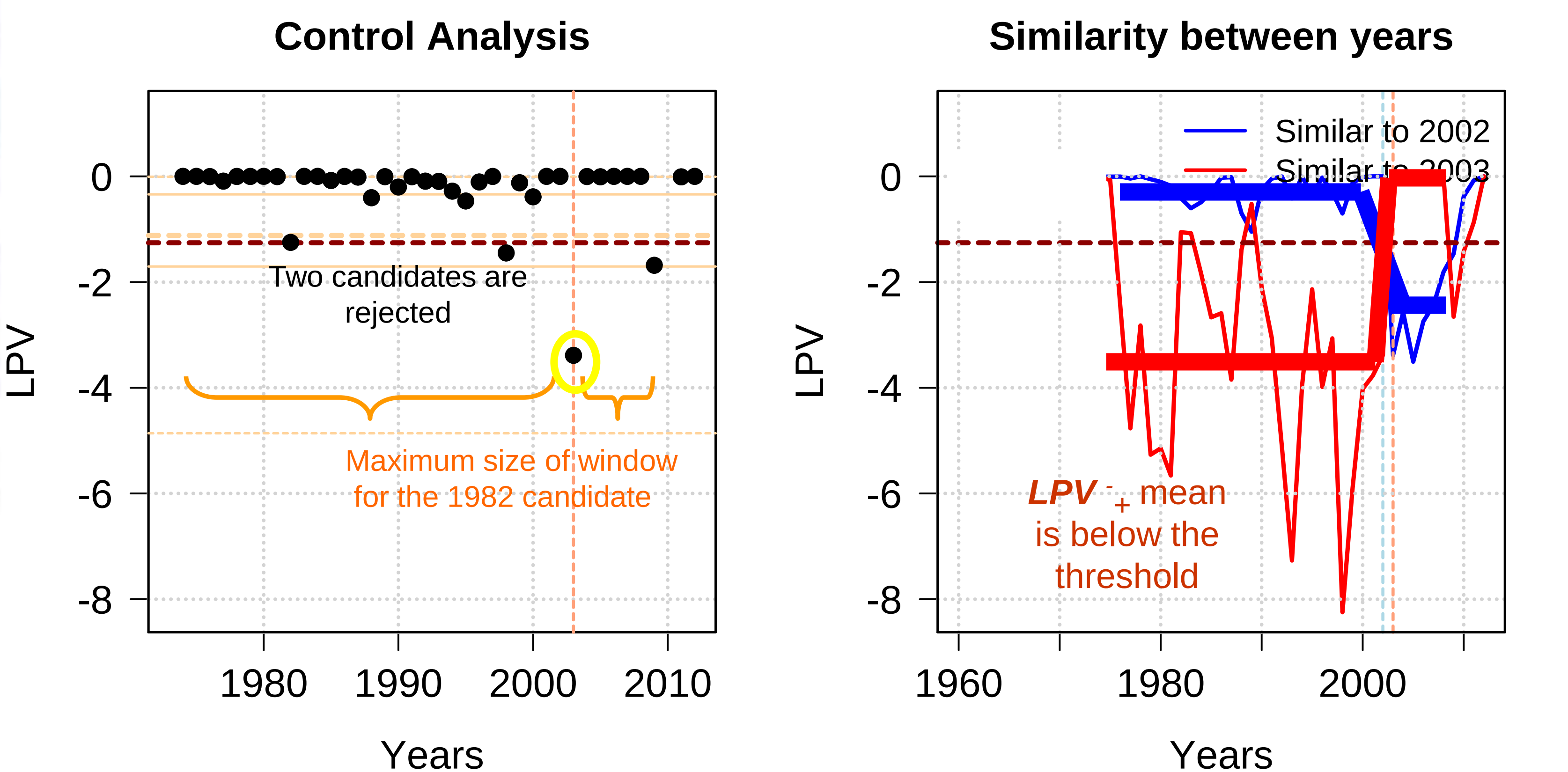


# Real cases



## Precipitation: CASE 1

Inhomogeneities in **precipitation**: LPV diagrams

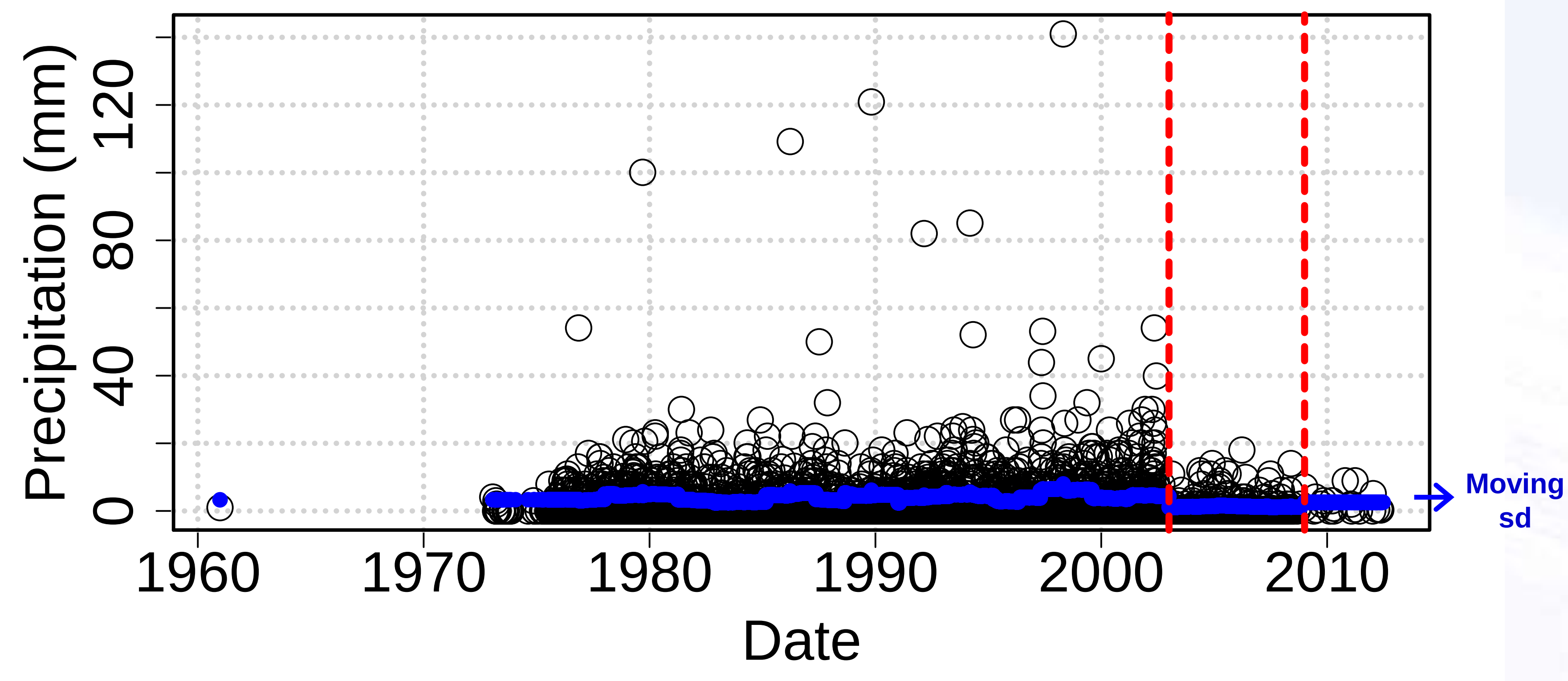


# Real cases



## Precipitation: CASE 1

Inhomogeneities in **precipitation**: final analysis

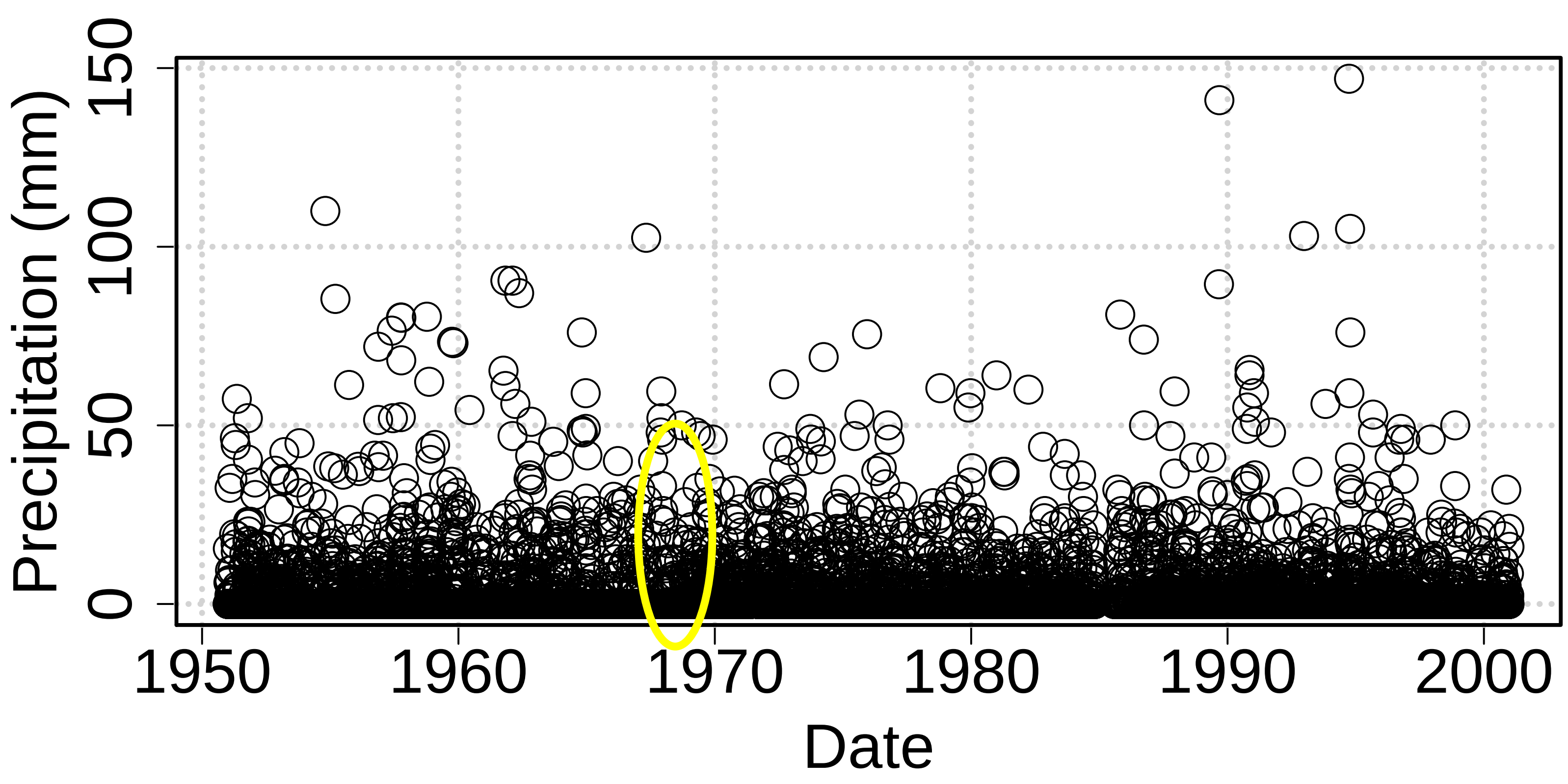


# Real cases



## Precipitation: CASE 2

Inhomogeneities in **precipitation**

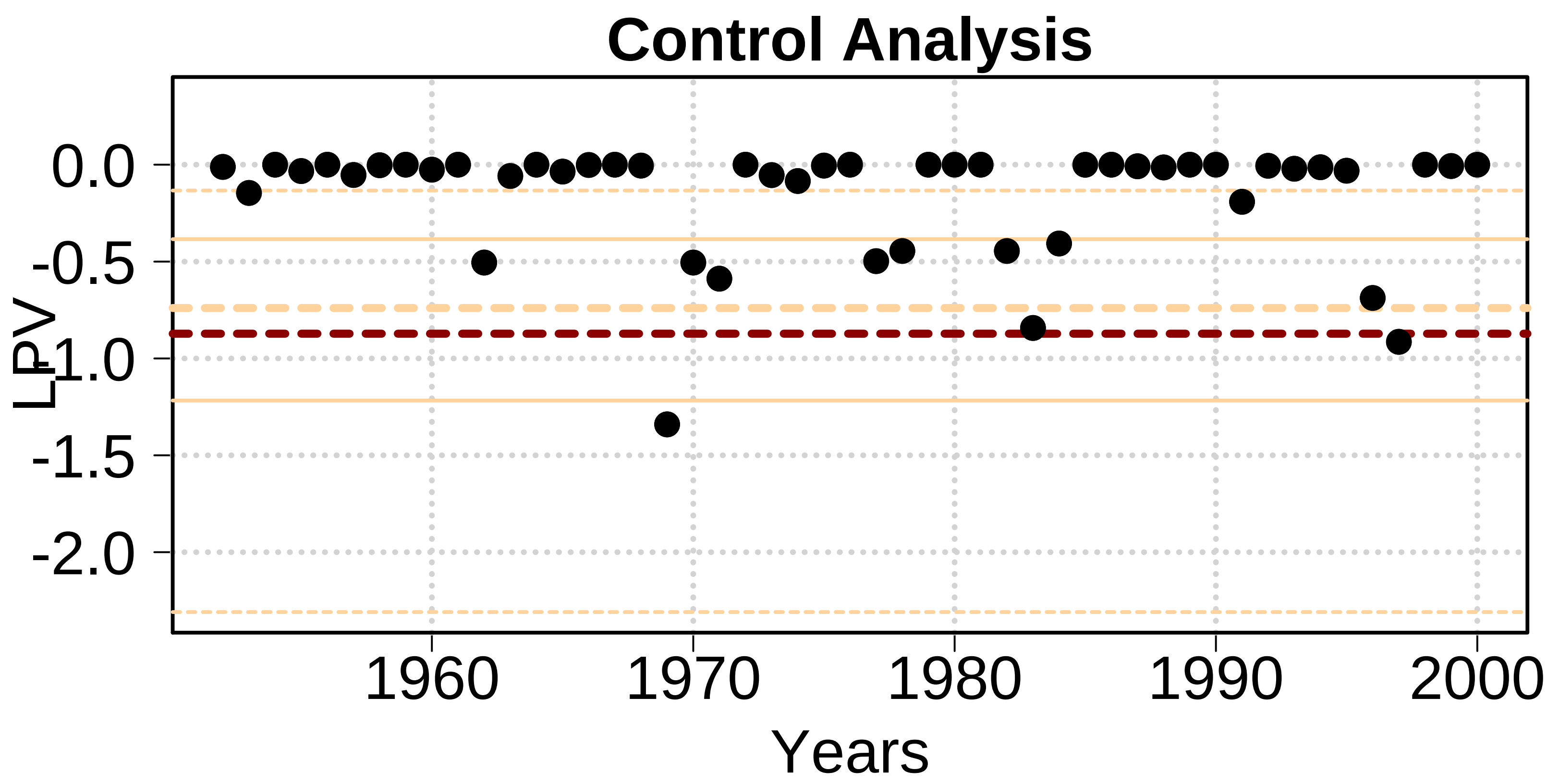


# Real cases



## Precipitation: CASE 2

Inhomogeneities in **precipitation**: LPV diagrams

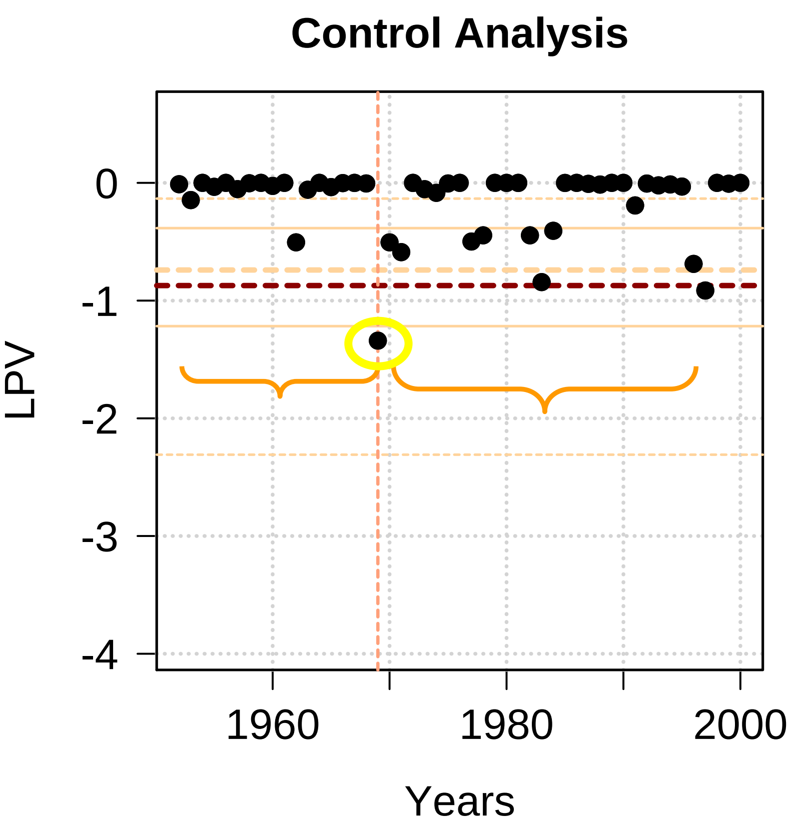


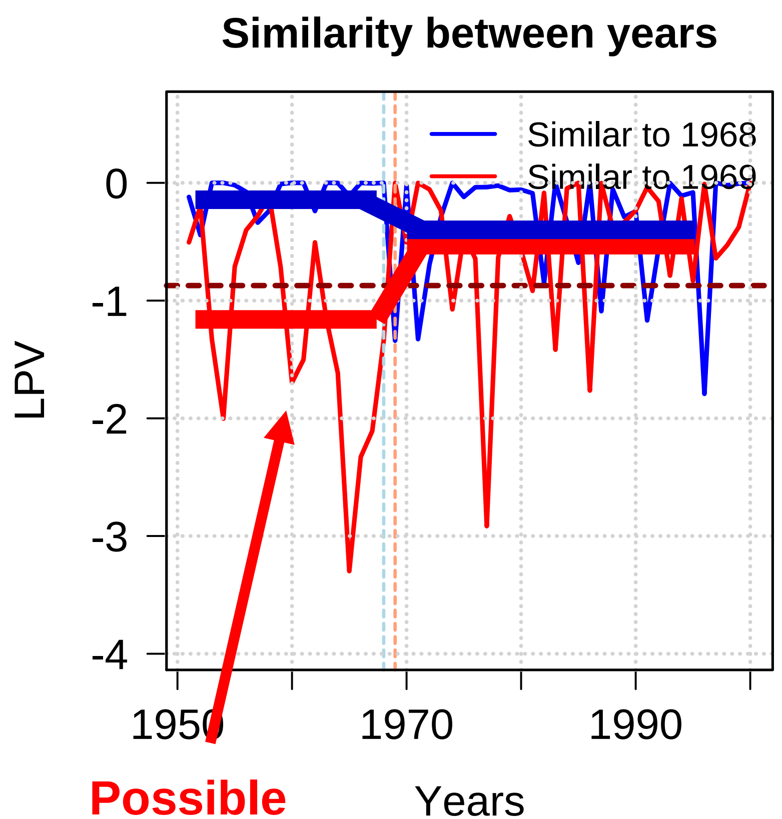
# Real cases



## Precipitation: CASE 2

Inhomogeneities in **precipitation**: LPV diagrams



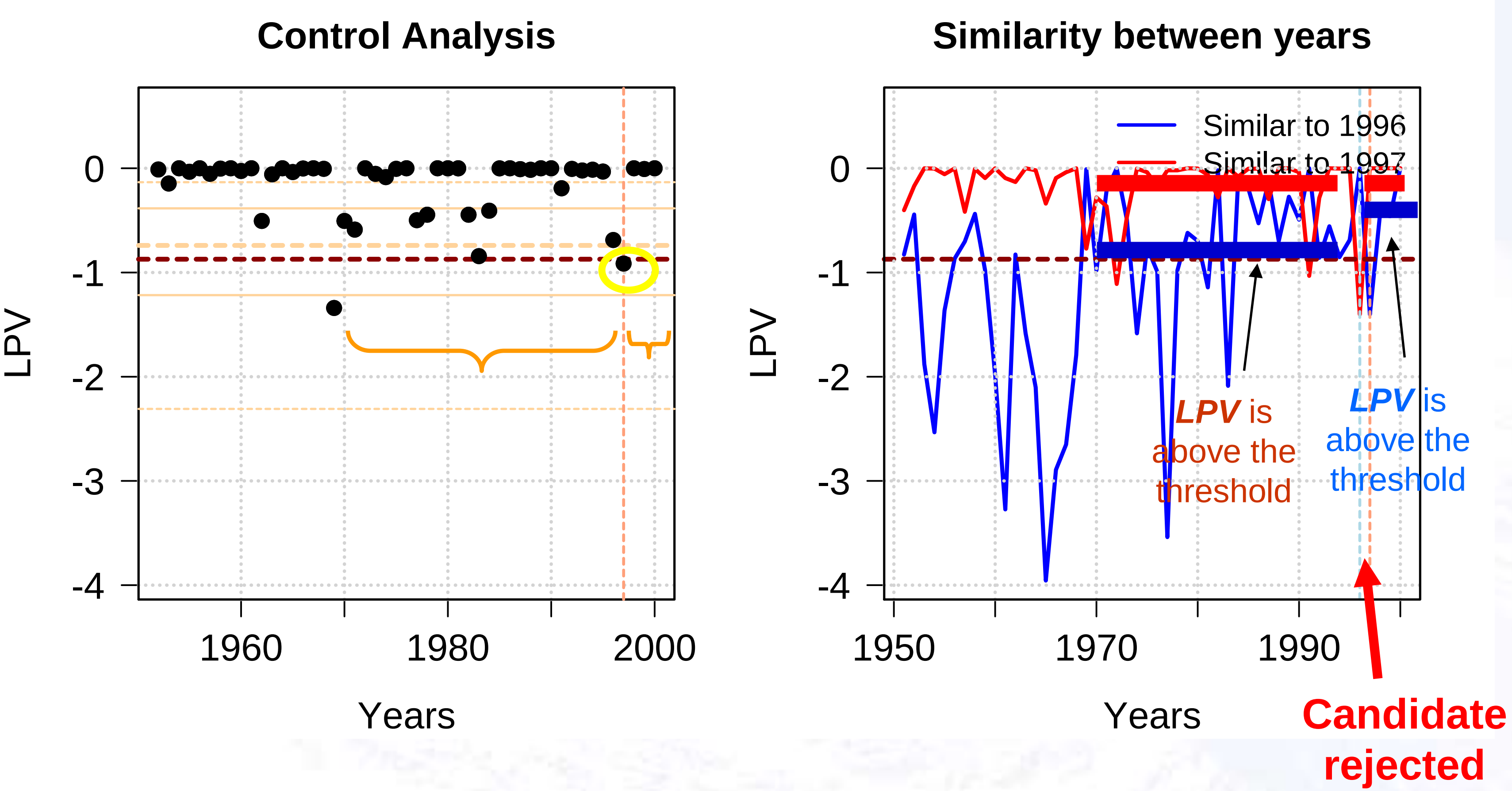



# Real cases



## Precipitation: CASE 2

Inhomogeneities in **precipitation**: LPV diagrams

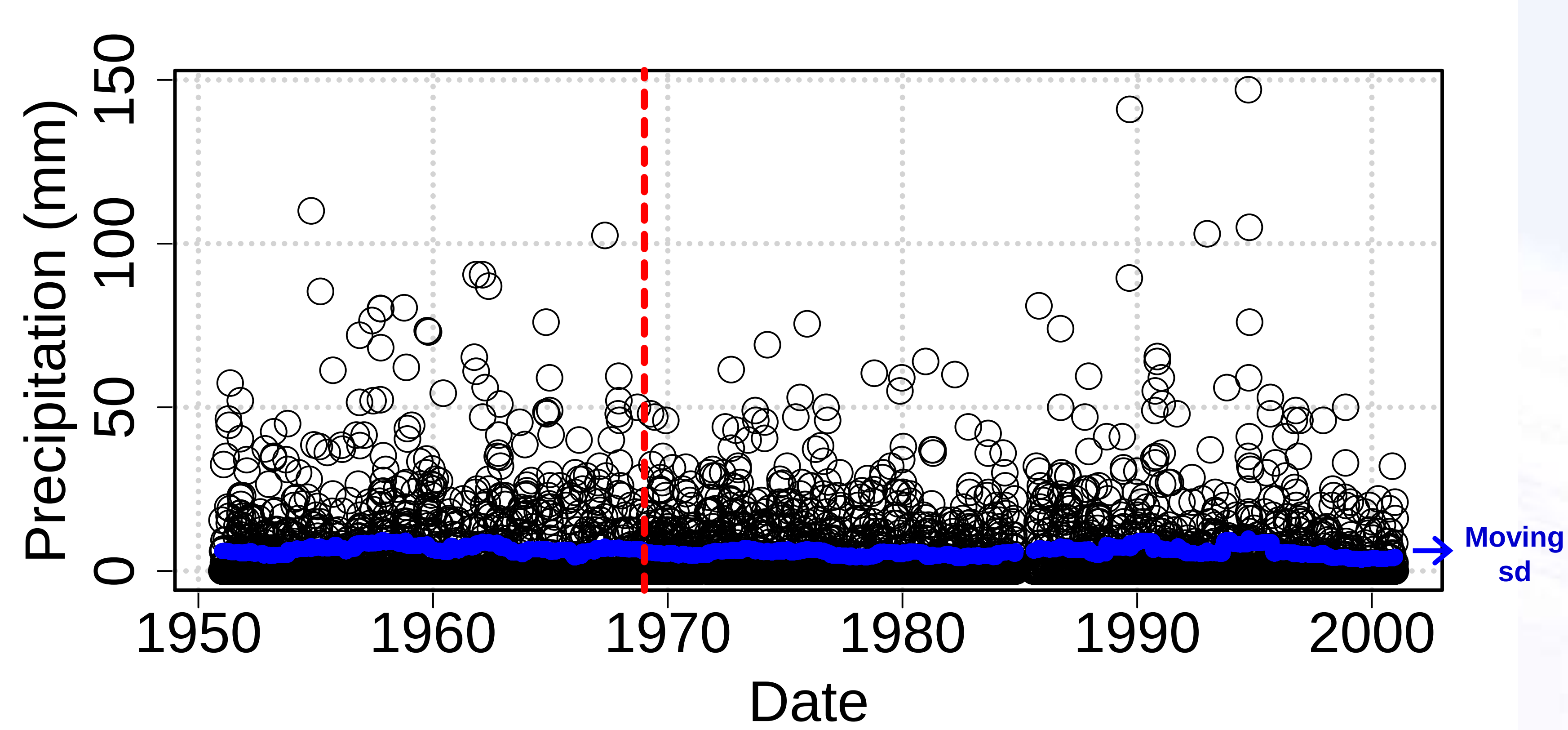


# Real cases



## Precipitation: CASE 2

Inhomogeneities in **precipitation**: final analysis



# Contents

# Discussion



- **Selection of non-parametric test**:
  - Kolmogorov-Smirnov: It is adequate for temperature and precipitation.
  - Other? Anderson-Darling: For precipitation is too much sensitive due to the high natural variability (takes a lot of extremes as "unusual year")
- **Analysis of control**:
  - Introduction of the artificial inhomogeneity: Choice of parameters of the control (a=?, b =?). Our case, *a* = 3 for precipitation, *b* = 2 for temperature
  - Percentile of the LPV of reference. Level of confidence for a candidate? For precipitation can be important due to the high uncertainty.
- **Inhomogeneity detection**:
  - Inhomogeneity (on the left/right) must be greater than the reference jump
  - Size of the detection window is limited by the candidate neighbors.
  - Time-series with trends are a priori considered homogenious because increase is soft (no jumps), except if there is a long gap.

# Contents

# Conclusions



## About the methodology

- Non-parametric test of Kolmogorov-Smirnov can be used to detect inhomogeneity for both temperature and precipitation at daily scale.

- A control analysis is required introducing an artificial inhomogeneity (“reference jump”) for comparing with the original time-series: LPV diagrams.

- For temperature, the recommended “reference jump” is 2ºC and for precipitation is a multiplication factor of 3.

## About its aplication

- Temperature shows a reference LPV with less uncertainty than precipitation.

- Test can detect (sistematic) changes in ECDF, not only in the averages

- Some candidates of precipitation can be due to natural variability.

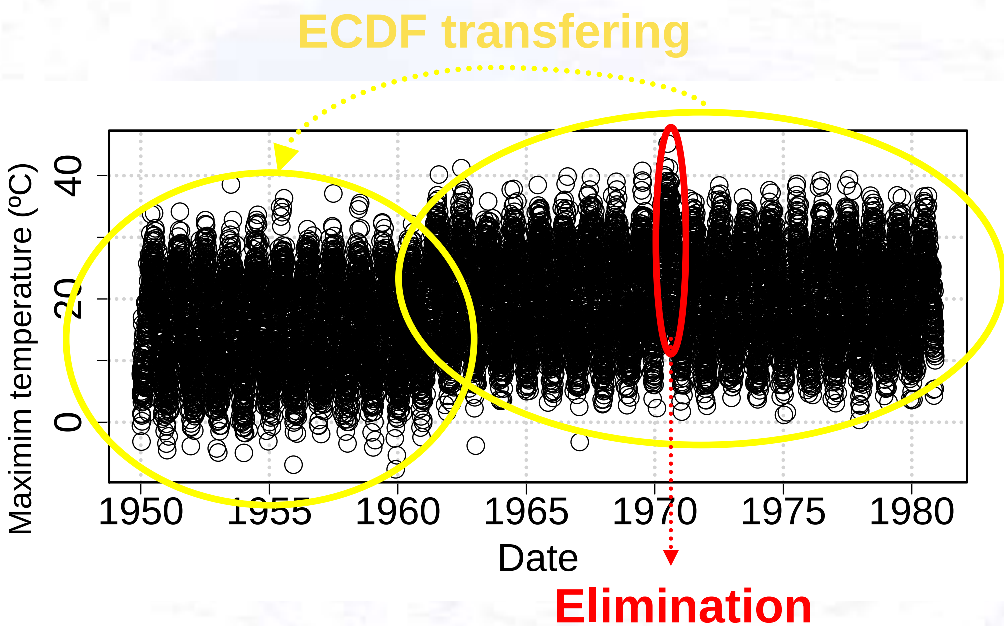
- A more extensive study of real cases could refine the methodology.

# Recommendations

**During inhomogeneity detection:** Automatic or supervised (recommended)

**After inhomogeneity detection:** Possible automatic correction in three ways :

- Use of nearby stations to detect and correct the wrong section of time-series
- Take the most recent or long (homogeneous) section of time series,
- Apply a ECDF function to transfer a climate features from recent to older section



# Thanks

**Thank you for your attention**

**and**
**for getting up so early…**