

# Applied Time Series Analysis

## SS 2013 – Week 14

*Marcel Dettling*

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

[marcel.dettling@zhaw.ch](mailto:marcel.dettling@zhaw.ch)

<http://stat.ethz.ch/~dettling>

ETH Zürich, May 27, 2013

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Multivariate Time Series Analysis*

**Idea:** Infer the relation between two time series

$$X_1 = (X_{1,t}) \text{ and } X_2 = (X_{2,t}).$$

#### **What is the difference to time series regression?**

- Here, the two series arise „on an equal footing“, and we are interested in the correlation between them.
- In time series regression, the two (or more) series are causally related and we are interested in inferring that relation. There is an independent and several dependent variables.
- The difference is comparable to the difference between correlation and regression.

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Example: Permafrost Boreholes*



A collaboration between the Swiss Institute for Snow and Avalanche Research with the Zurich University of Applied Sciences:

Evelyn Zenklusen Mutter & Marcel Dettling

# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Example: Permafrost Boreholes***

- given is a bivariate time series with  $2 \cdot 92$  observations
- 2 measurements were made everyday in summer 2006
- series 1: air temperature at Platthorn 3345m
- series 2: soil temperature at Hörnli hut 3295m

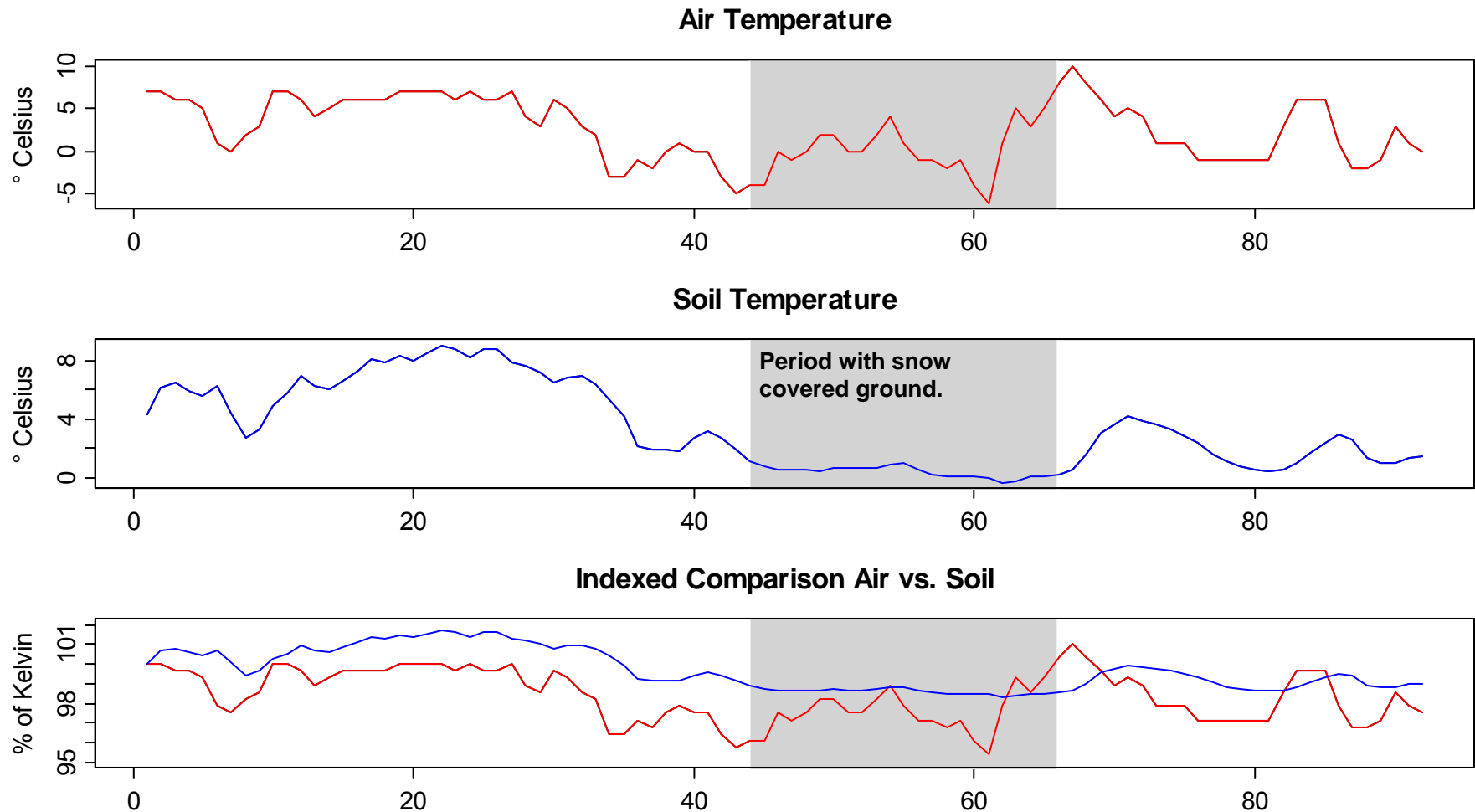
#### **Goal of the analysis:**

- 1) Answer whether changes in the air temperature are correlated with changes in the soil temperature.
- 2) If a correlation is present, what is the delay?

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Air & Soil Temperature Comparison*

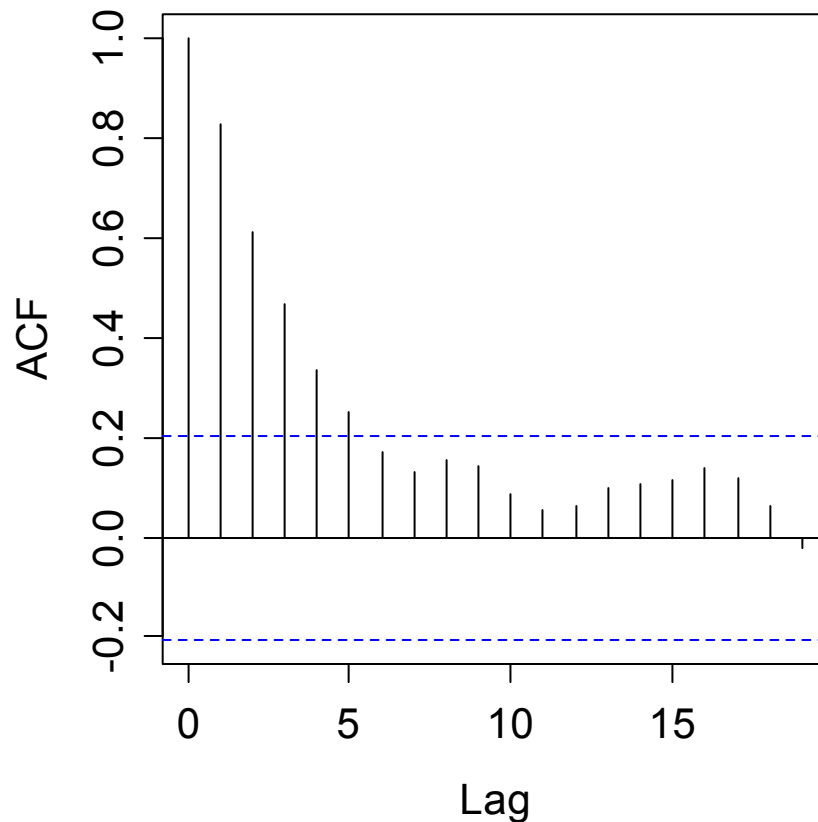


# Applied Time Series Analysis

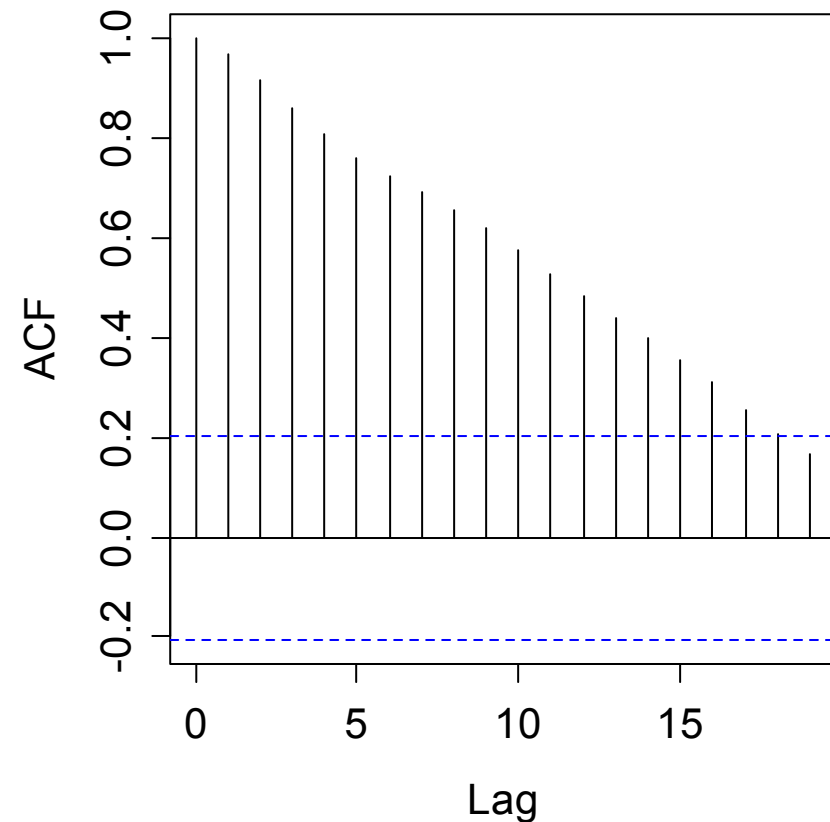
## SS 2013 – Week 14

### *Are the Series Stationary?*

ACF of Air Temperature



ACF of Soil Temperature



# Applied Time Series Analysis

## SS 2013 – Week 14

### ***How to Proceed?***

- 1) The series seem to have „long memory“
  - 2) Pure AR/MA/ARMA do not fit the data well
- Differencing may help with this

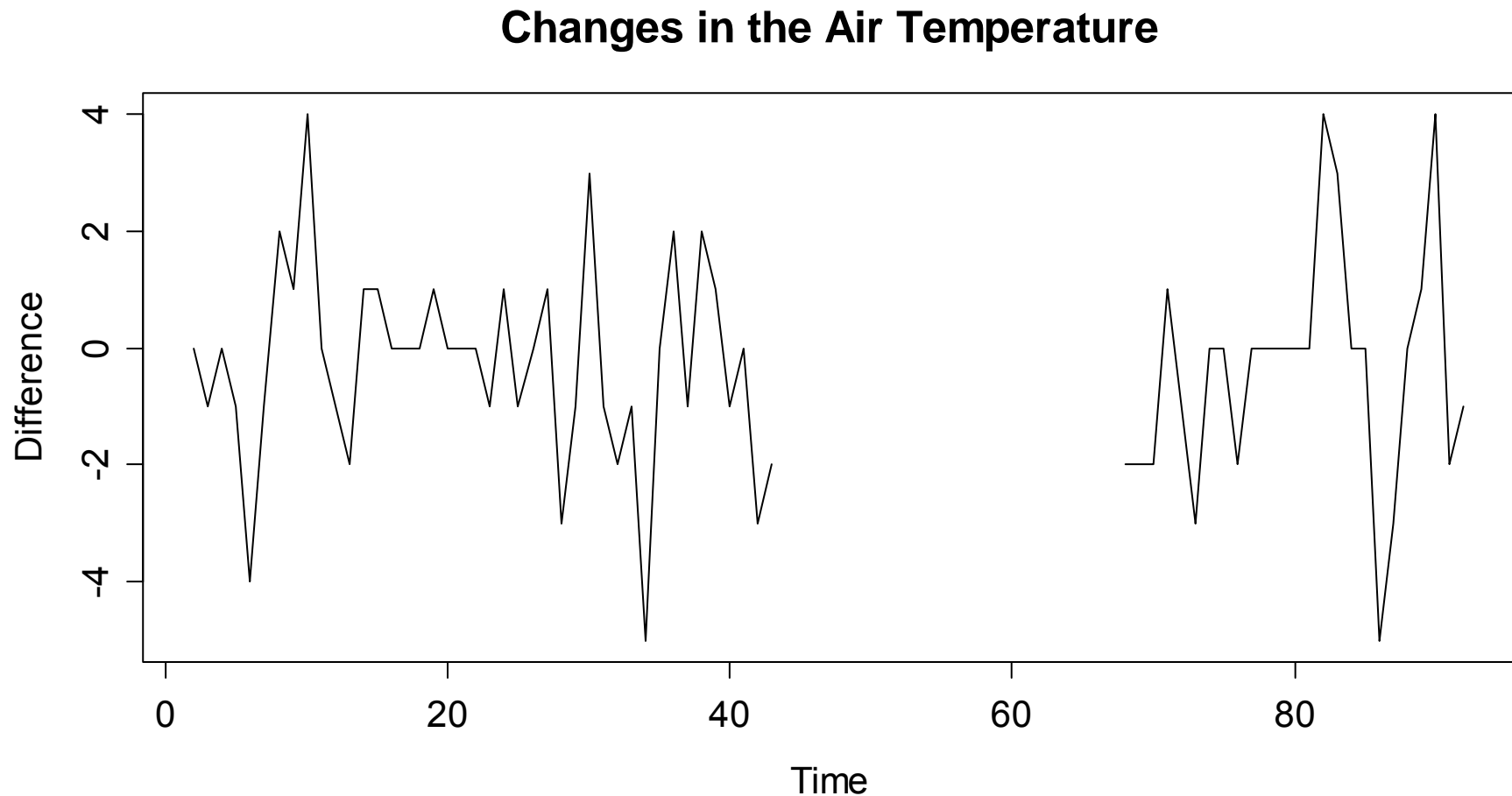
### **Another advantage of taking differences:**

- we infer, whether there is a relation between the changes in the air temperatures, and the changes in the soil temperatures.

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Changes in the Air Temperature*



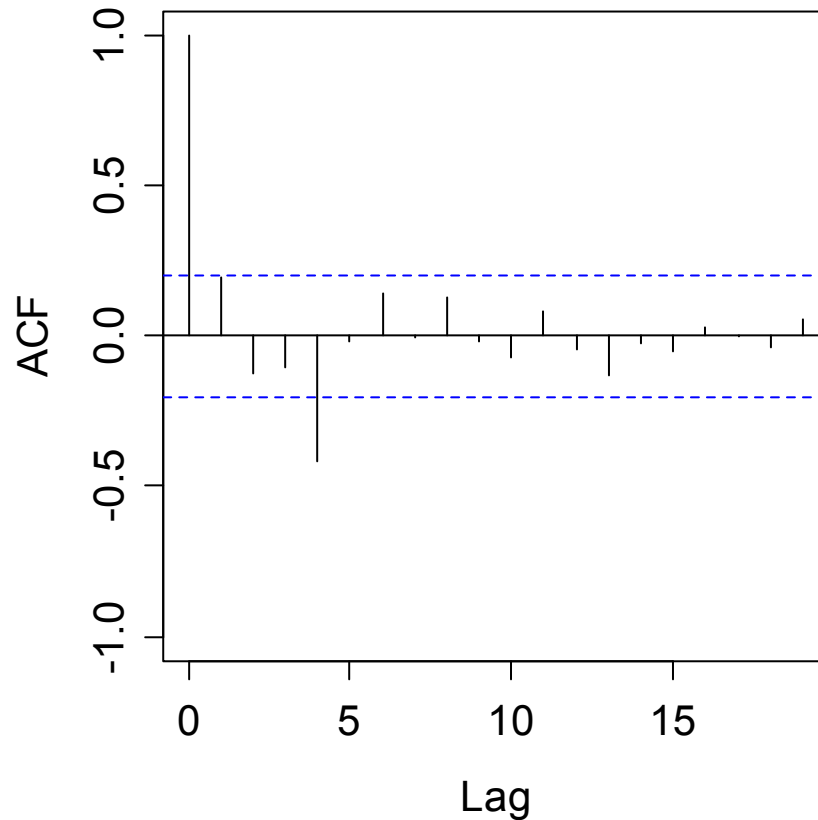


# Applied Time Series Analysis

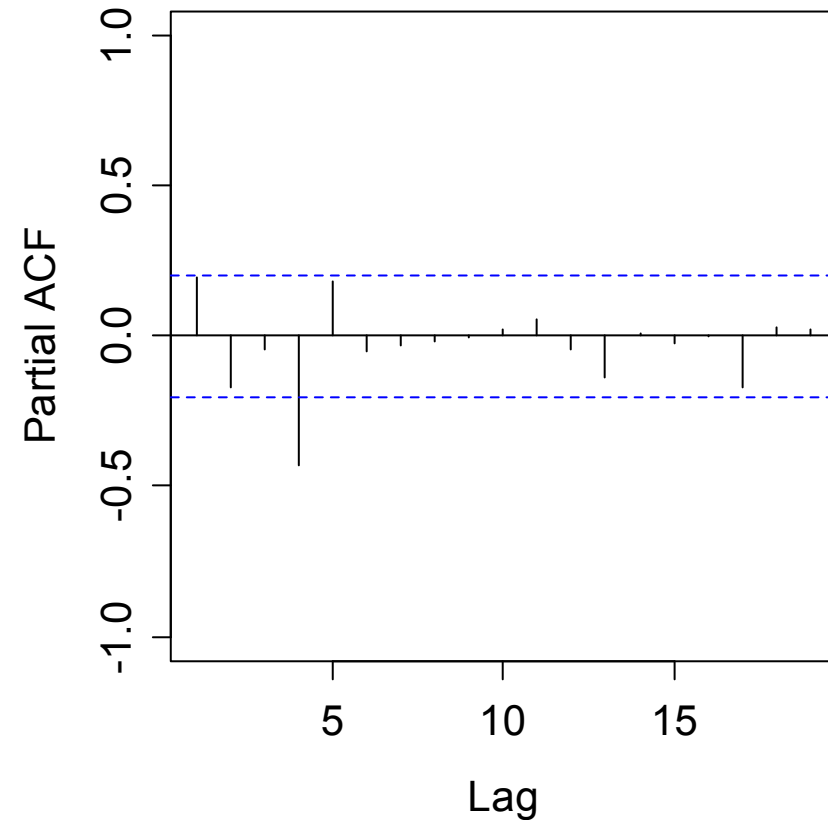
## SS 2013 – Week 14

### *ACF/PACF for Air Temperature Changes*

ACF



PACF

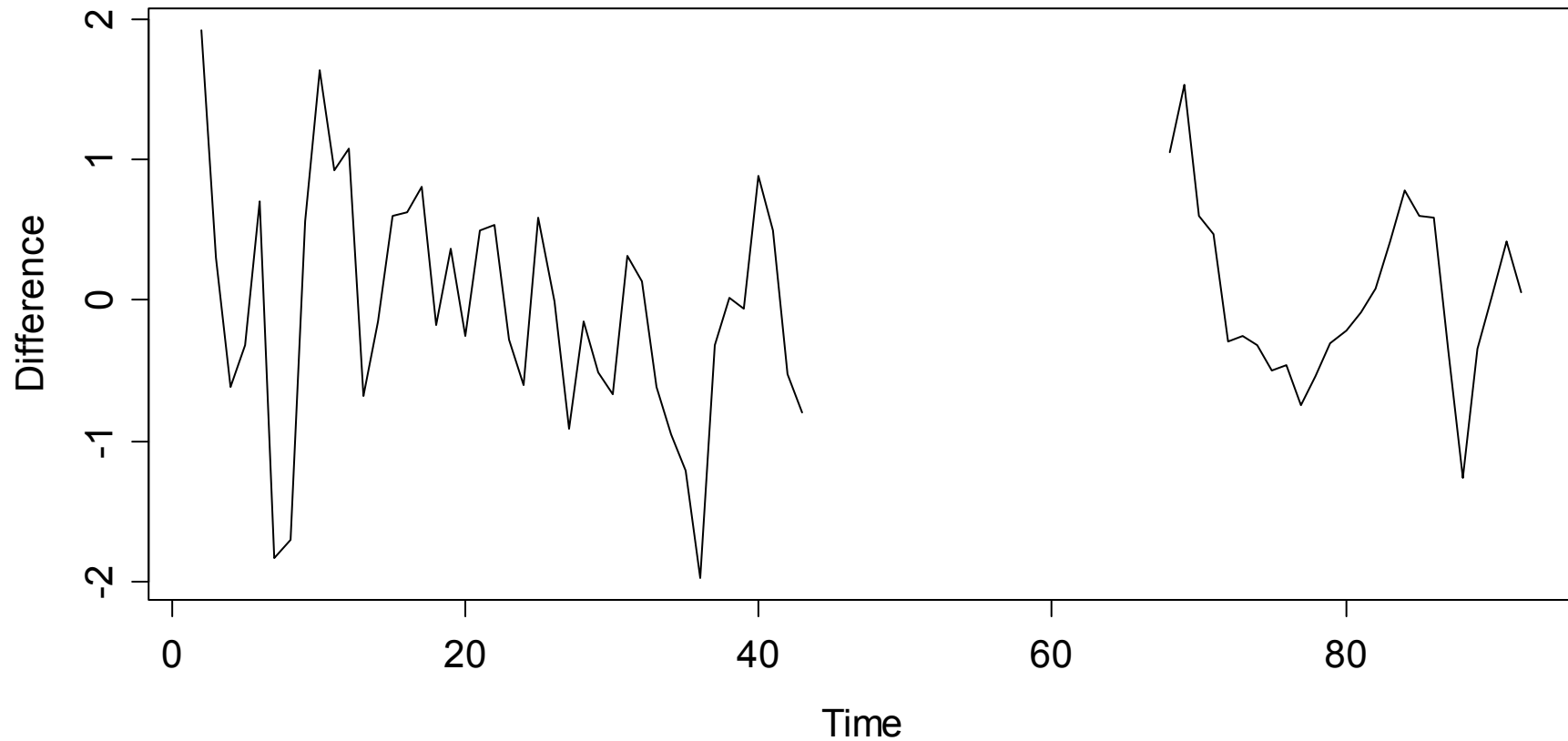


# Applied Time Series Analysis

## SS 2013 – Week 14

### *Changes in the Soil Temperature*

Changes in the Soil Temperature

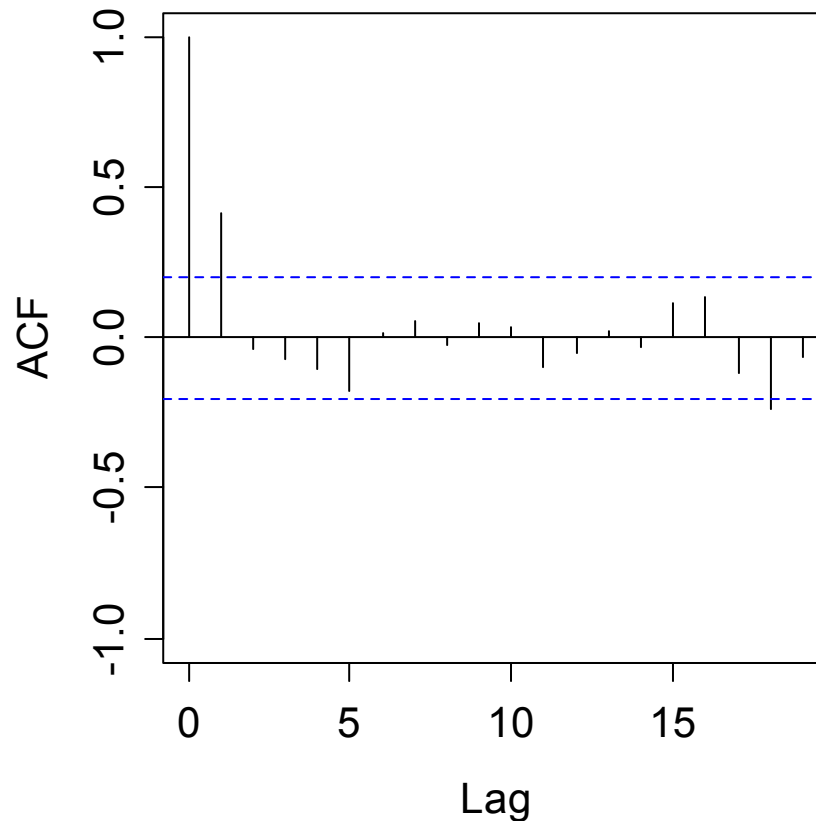


# Applied Time Series Analysis

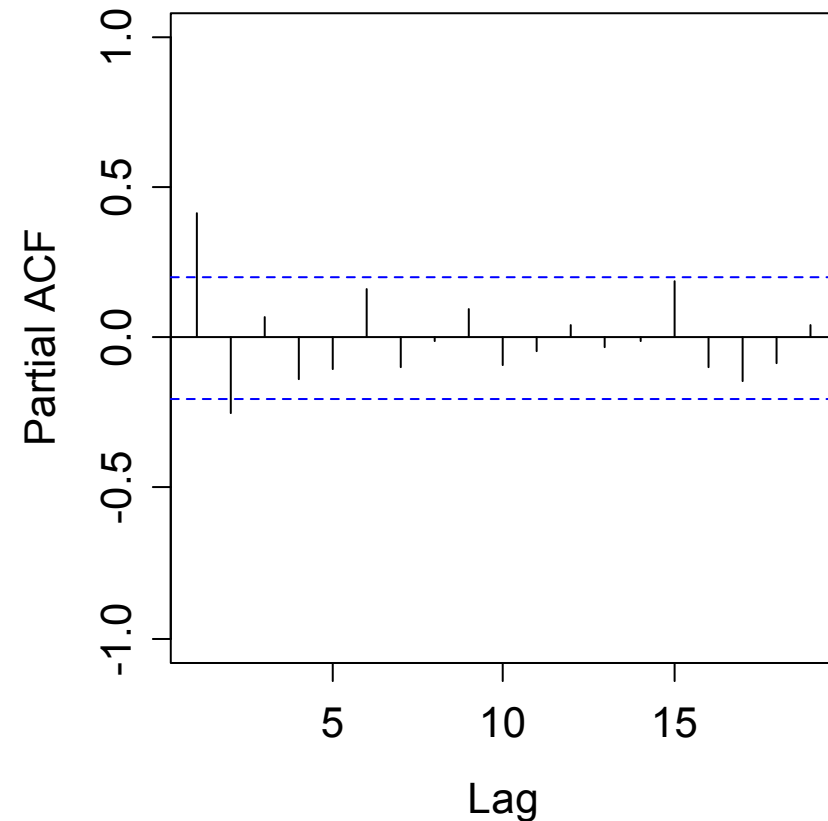
## SS 2013 – Week 14

### *ACF/PACF for Soil Temperature Changes*

ACF



PACF



# Applied Time Series Analysis

## SS 2013 – Week 14

### Cross Covariance

The cross correlations describe the relation between two time series. However, note that the interpretation is quite tricky!

$$\left. \begin{aligned} \gamma_{11}(k) &= \text{Cov}(X_{1,t+k}, X_{1,t}) \\ \gamma_{22}(k) &= \text{Cov}(X_{2,t+k}, X_{2,t}) \end{aligned} \right\} \begin{array}{l} \text{usual „within series“} \\ \text{covariance} \end{array}$$

$$\left. \begin{aligned} \gamma_{12}(k) &= \text{Cov}(X_{1,t+k}, X_{2,t}) \\ \gamma_{21}(k) &= \text{Cov}(X_{2,t+k}, X_{1,t}) \end{aligned} \right\} \begin{array}{l} \text{cross covariance,} \\ \text{independent from } t \end{array}$$

Also, we have:  $\gamma_{12}(-k) = \text{Cov}(X_{1,t-k}, X_{2,t}) = \text{Cov}(X_{2,t+k}, X_{1,t}) = \gamma_{21}(k)$

# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Cross Correlations***

It suffices to analyze  $\gamma_{12}(k)$ , and neglect  $\gamma_{21}(k)$ , but we have to regard both positive and negative lags  $k$ .

We again prefer to work with correlations:

$$\rho_{12}(k) = \frac{\gamma_{12}(k)}{\sqrt{\gamma_{11}(0)\gamma_{22}(0)}}$$

which describe the linear relation between two values of  $X_1$  and  $X_2$ , when the series  $X_1$  is  $k$  time units ahead.

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Estimation*

Cross covariances and correlations are estimated as follows:

$$\hat{\gamma}_{12}(k) = \frac{1}{n} \sum_t (x_{1,t+k} - \bar{x}_1)(x_{2,t} - \bar{x}_2)$$

and

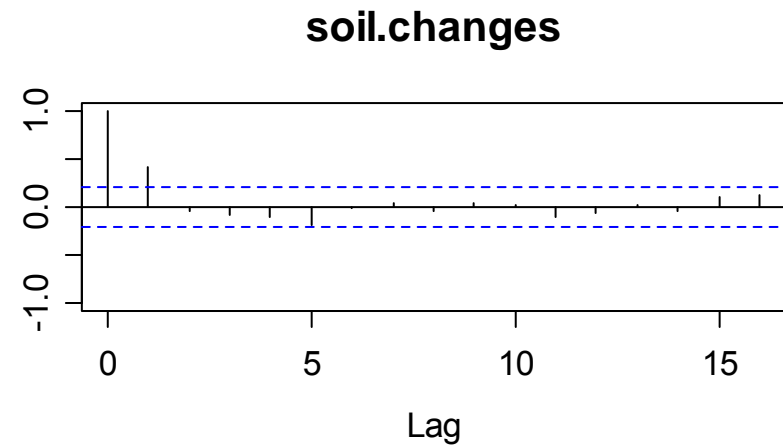
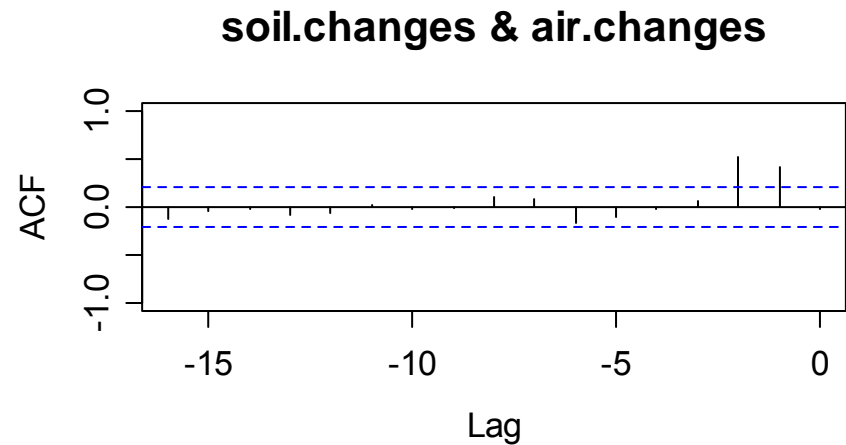
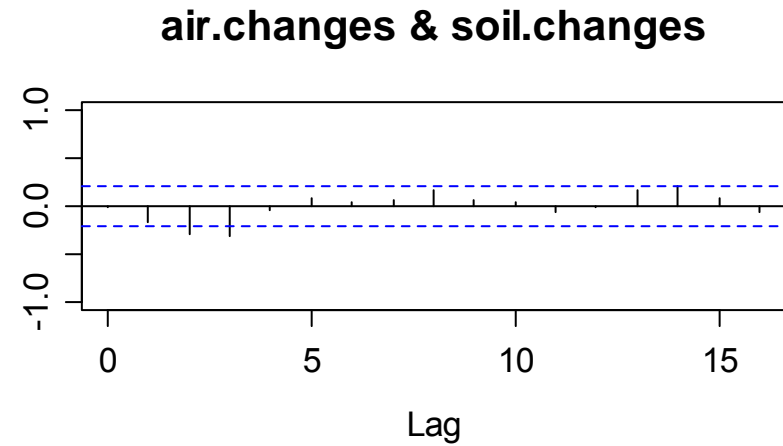
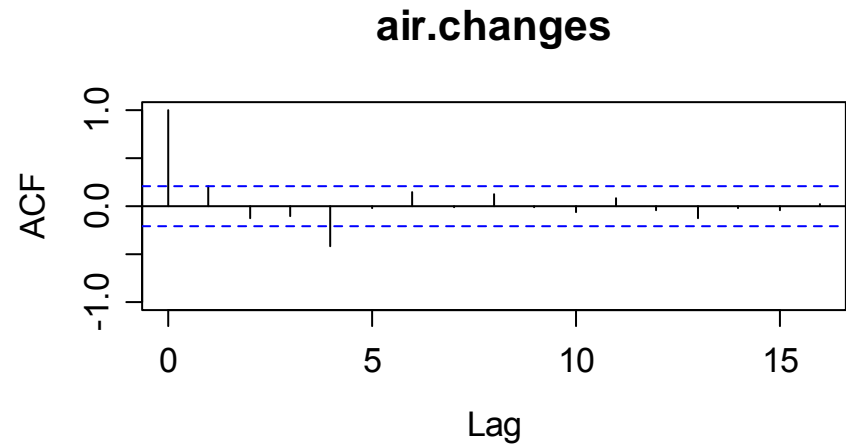
$$\hat{\rho}_{12}(k) = \frac{\hat{\gamma}_{12}(k)}{\sqrt{\hat{\gamma}_{11}(0)\hat{\gamma}_{22}(0)}}, \quad \text{respectively.}$$

The plot of  $\hat{\rho}_{12}(k)$  versus the lag  $k$  is called the cross correlogram. It has to be inspected for both  $+$  and  $-k$ .

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Sample Cross Correlation*



# Applied Time Series Analysis

## SS 2013 – Week 14

### *Interpreting the Sample Cross Correlation*

The confidence bounds in the sample cross correlation are only valid in some special cases, i.e. if there is no cross correlation and at least one of the series is uncorrelated.

**Important: the confidence bounds are often too small!**

For computing them, we need:  $Var(\hat{\rho}_{12}(k))$

This is a difficult problem. We are going to discuss a few special cases and then show how the problem can be circumvented.



# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Special Case 1***

We assume that there is no cross correlation for large lags  $k$ :

If  $\rho_{12}(j) = 0$  for  $|j| \geq m$ , we have for  $|k| \geq m$ :

$$\text{Var}(\hat{\rho}_{12}(k)) \approx \frac{1}{n} \sum_{j=-\infty}^{\infty} (\rho_{11}(j)\rho_{22}(j) + \rho_{12}(j+k)\rho_{12}(j-k))$$

This goes to zero for large  $k$  and we thus have consistency.

For giving statements about the confidence bounds, we would have to know more about the cross correlations, though.

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Special Case 2*

There is no cross correlation, but  $X_1$  and  $X_2$  are time series that show correlation „within“:

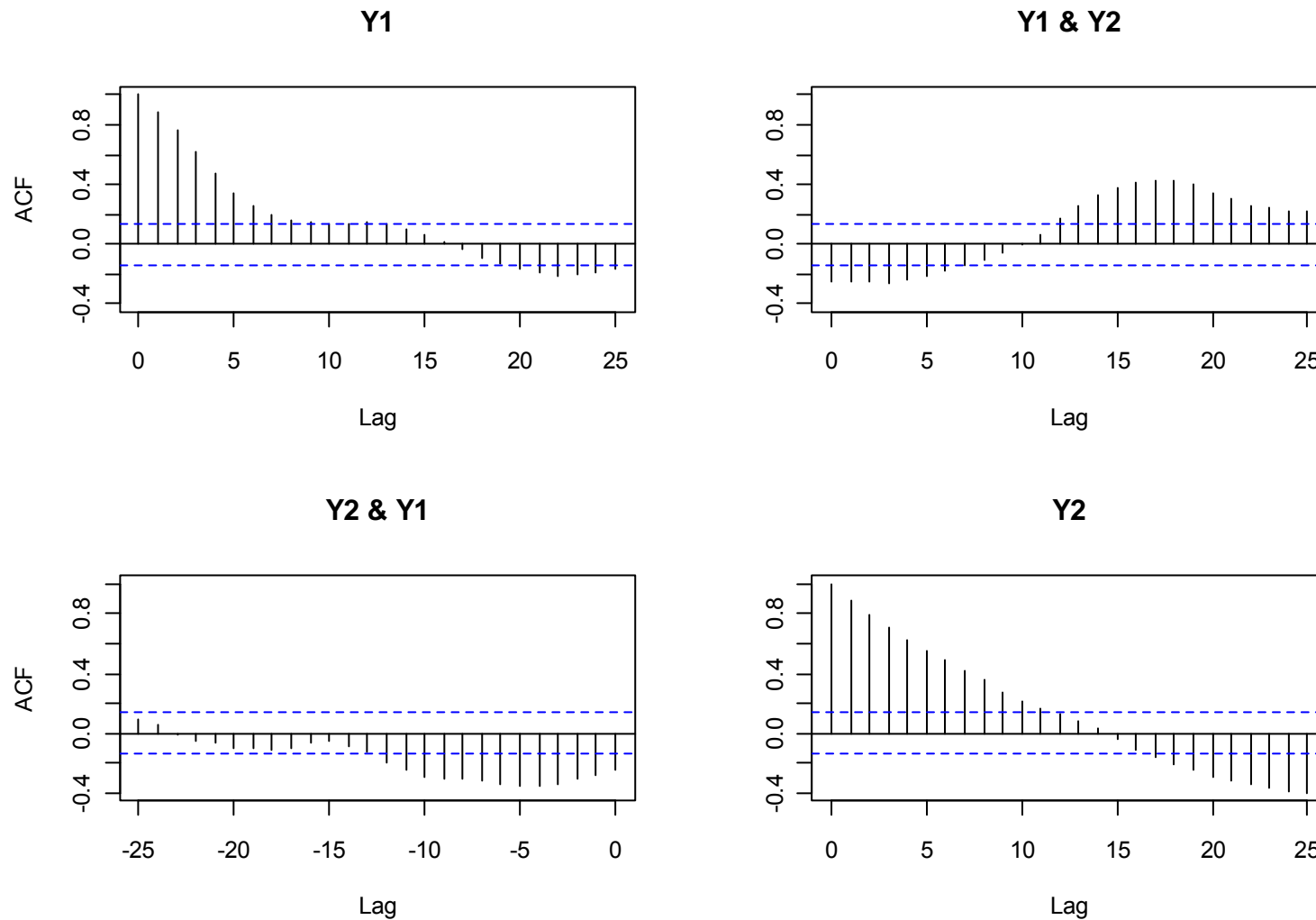
$$\text{Var}(\hat{\rho}_{12}(k)) \approx \frac{1}{n} \sum_{j=-\infty}^{\infty} \rho_{11}(j) \rho_{22}(j)$$

**See the blackboard...** for the important example showing that the cross correlation estimations can be arbitrarily bad!

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Special Case 2: Simulation Example*



# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Special Case 3***

There is no cross correlation, and  $X_1$  is a white noise series that is independent from  $X_2$ . Then, the estimation variance simplifies to:

$$\text{Var}(\hat{\rho}_{12}(k)) \approx \frac{1}{n}$$

Thus, the confidence bounds are valid in this case.

However, we introduced the concept of cross correlation to infer the relation between correlated series. The trick of the so-called „prewhitening“ helps.

# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Prewhitening***

Prewhitening means that the time series is transformed such that it becomes a white noise process, i.e. is uncorrelated.

We assume that both stationary processes  $X_1$  and  $X_2$  can be rewritten as follows:

$$U_t = \sum_{i=0}^{\infty} a_i X_{1,t-i} \quad \text{and} \quad V_t = \sum_{i=0}^{\infty} b_i X_{2,t-i},$$

with uncorrelated  $U_t$  and  $V_t$ . Note that this is possible for ARMA(p,q) processes by writing them as an AR( $\infty$ ). The left hand side of the equation then is the innovation.

# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Cross Correlation of Prewhitened Series***

The cross correlation between  $U_t$  and  $V_t$  can be derived from the one between  $X_1$  and  $X_2$ :

$$\rho_{UV}(k) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_i b_j \rho_{X_1 X_2}(k + i - j)$$

**Thus we have:**

$$\rho_{UV}(k) = 0 \quad \text{for all } k \Leftrightarrow \rho_{X_1 X_2}(k) = 0 \quad \text{for all } k$$

**Now:** generate  $U_t, V_t$ ; estimate cross correlations; and, by using the confidence bands, check whether they are significant

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Simulation Example*

Since we are dealing with simulated series, we know that:

$$X_{i,t} = 0.9 \cdot X_{i,t-1} + E_t, \text{ thus } E_t = X_{i,t} - 0.9 \cdot X_{i,t-1}$$

In practice, we don't know the AR-coefficients, but plug-in the respective estimates:

$$U_t = X_{1,t} - \hat{\alpha}_{1,1} X_{1,t-1} \quad \text{with } \hat{\alpha}_{1,1} = 0.911$$

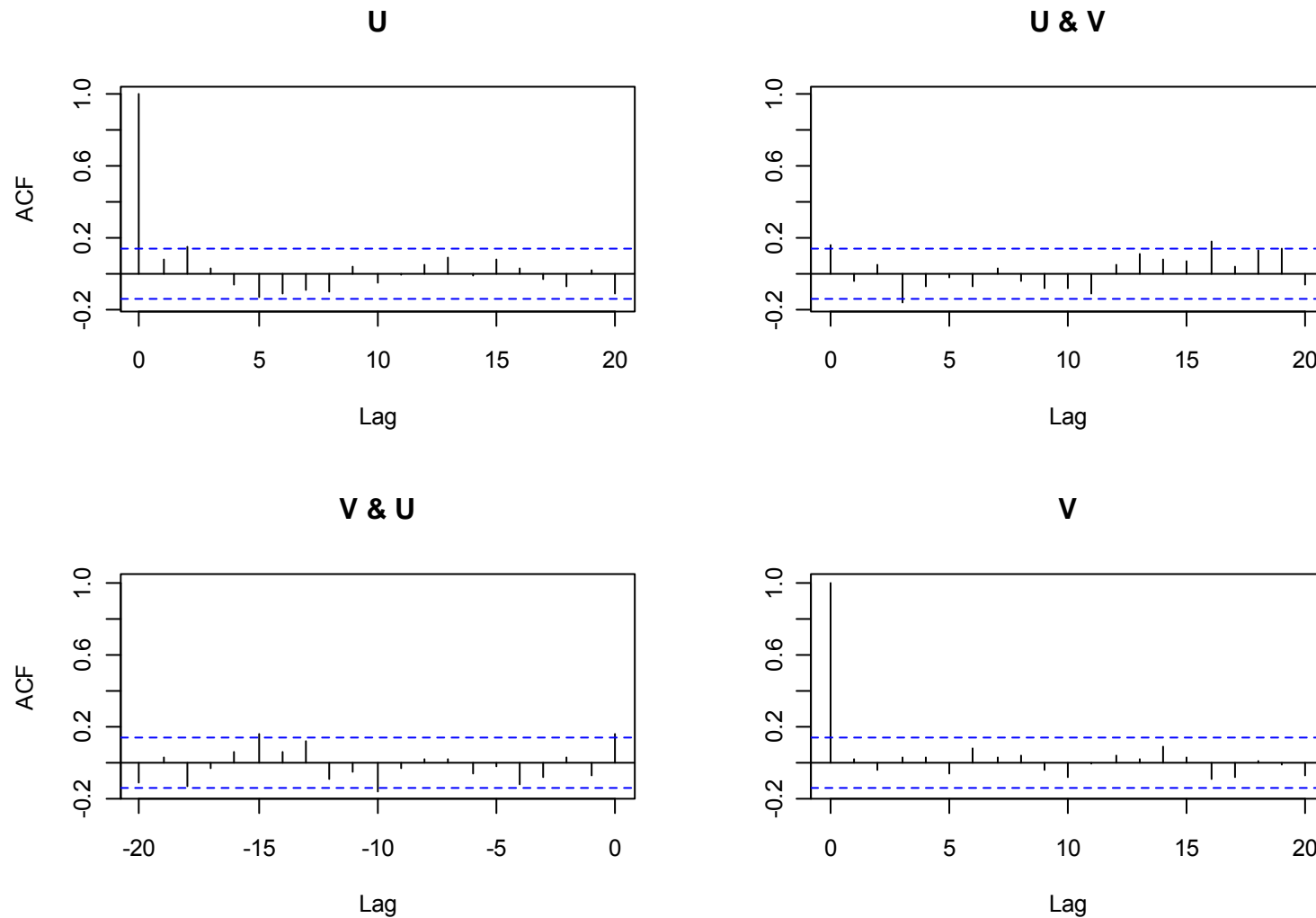
$$V_t = X_{2,t} - \hat{\alpha}_{2,1} X_{2,t-1} \quad \text{with } \hat{\alpha}_{2,1} = 0.822$$

We will now analyse the sample cross correlation of  $U_t$  and  $V_t$ , which will also allow to draw conclusions about  $X_1$  and  $X_2$ .

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Cross Correlation in the Simulation Example*





# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Cross Correlation in the Simulation Example***

**We observe that:**

- $U_t$  and  $V_t$  are white noise processes
- There are no (strongly) significant cross correlations

**We conjecture that:**

- $X_1$  and  $X_2$  are not cross correlated either.

→ This matches our „expectations“, or better, true process.

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Prewhitening the Borehole Data*

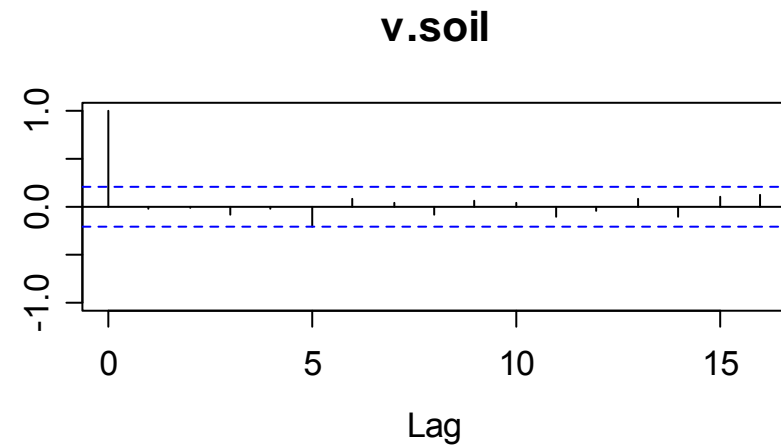
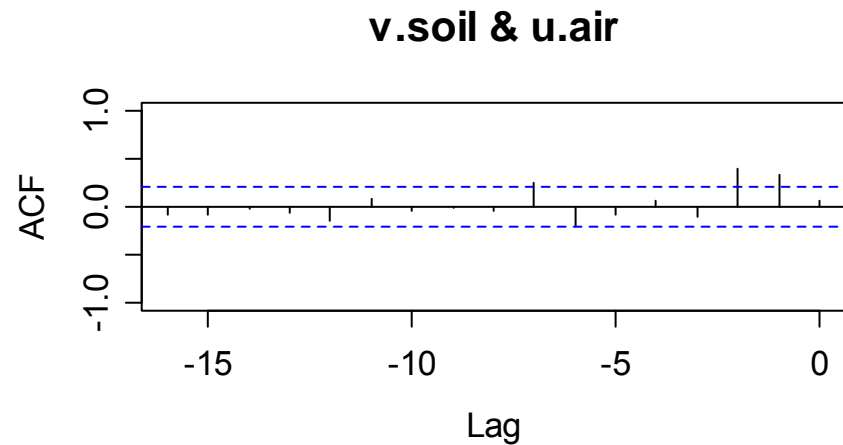
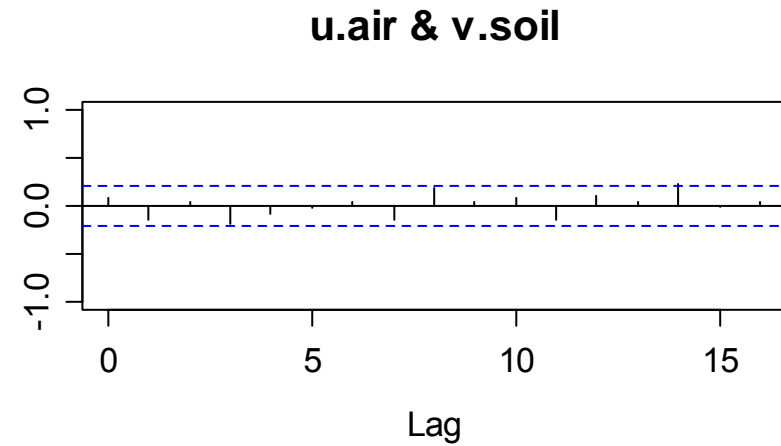
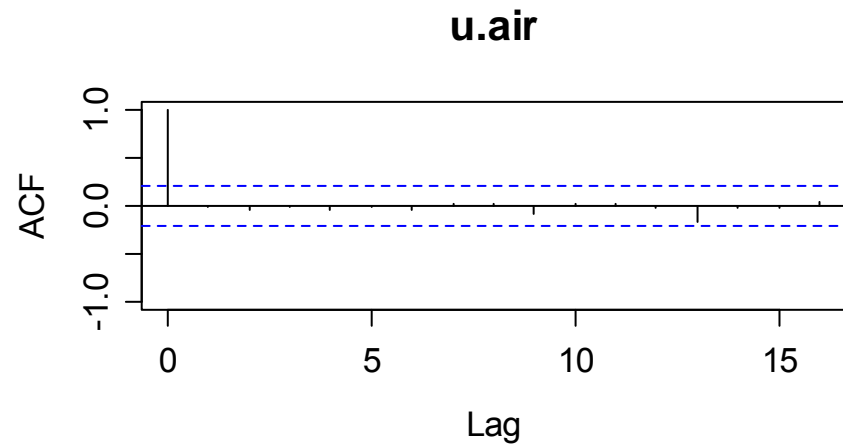
#### What to do:

- ARMA(p,q)-models are fitted to the differenced series
- Best choice: AR(5) for the air temperature differences  
MA(1) for the soil temperature differences
- The residual time series are  $U_t$  and  $V_t$ , white noise
- Check the sample cross correlation (see next slide)
- Model the output as a linear combination of past input values: transfer function model.

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Prewhitening the Borehole Data*



# Applied Time Series Analysis

## SS 2013 – Week 14

### *Transfer Function Models*

#### Properties:

- Transfer function models are an option to describe the dependency between two time series.
- The first (input) series influences the second (output) one, but there is no feedback from output to input.
- The influence from input to output only goes „forward“.

#### The model is:

$$X_{2,t} - \mu_2 = \sum_{j=0}^{\infty} v_j (X_{1,t-j} - \mu_1) + E_t$$

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Transfer Function Models*

The model is:

$$X_{2,t} - \mu_2 = \sum_{j=0}^{\infty} v_j (X_{1,t-j} - \mu_1) + E_t$$

- $E[E_t]=0$ .
- $E_t$  and  $X_{1,s}$  are uncorrelated for all  $t$  and  $s$ .
- $E_t$  and  $E_s$  are usually correlated.
- For simplicity of notation, we here assume that the series have been mean centered.

# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Cross Covariance***

When plugging-in, we obtain for the cross covariance:

$$\gamma_{21}(k) = \text{Cov}(X_{2,t+k}, X_{1,t}) = \text{Cov}\left(\sum_{j=0}^{\infty} v_j X_{1,t+k-j}, X_{1,t}\right) = \sum_{j=0}^{\infty} v_j \gamma_{11}(k-j)$$

- If only finitely many coefficients are different from zero, we could generate a linear equation system, plug-in  $\hat{\gamma}_{11}$  and  $\hat{\gamma}_{21}$  to obtain the estimates  $\hat{v}_j$ .
- **This is not a statistically efficient estimation method.**

# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Special Case: $X_{1,t}$ Uncorrelated***

If  $X_{1,t}$  was an uncorrelated series, we would obtain the coefficients of the transfer function model quite easily:

$$v_k = \frac{\gamma_{21}(k)}{\gamma_{11}(0)}$$

However, this is usually not the case. We can then:

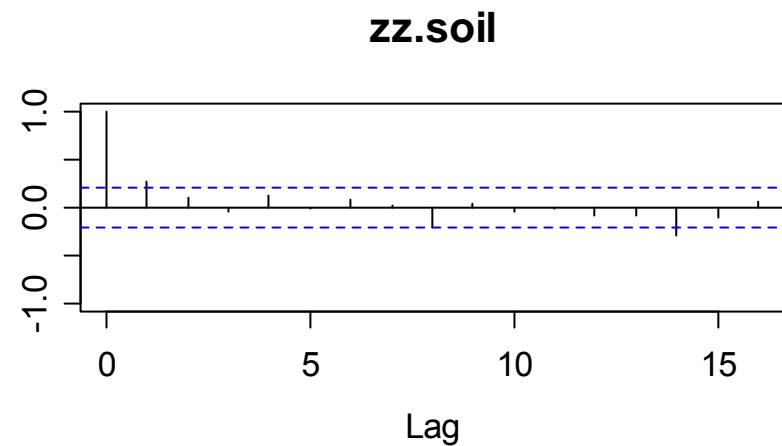
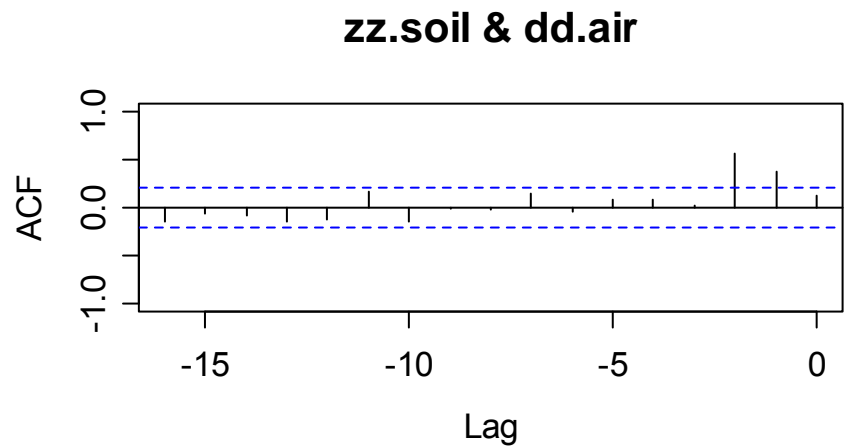
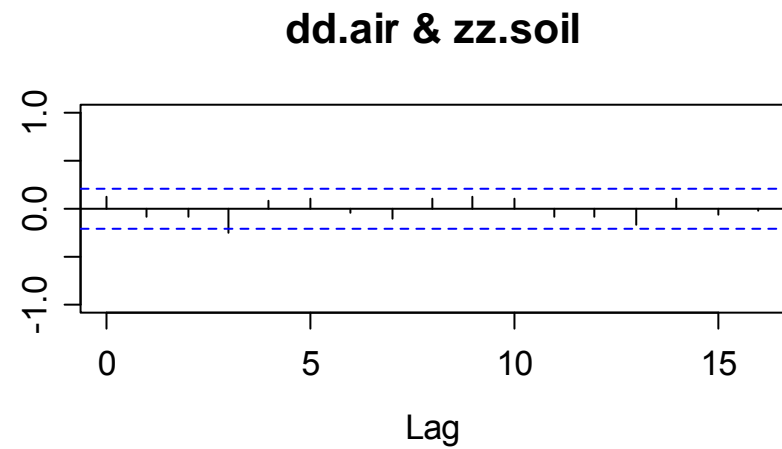
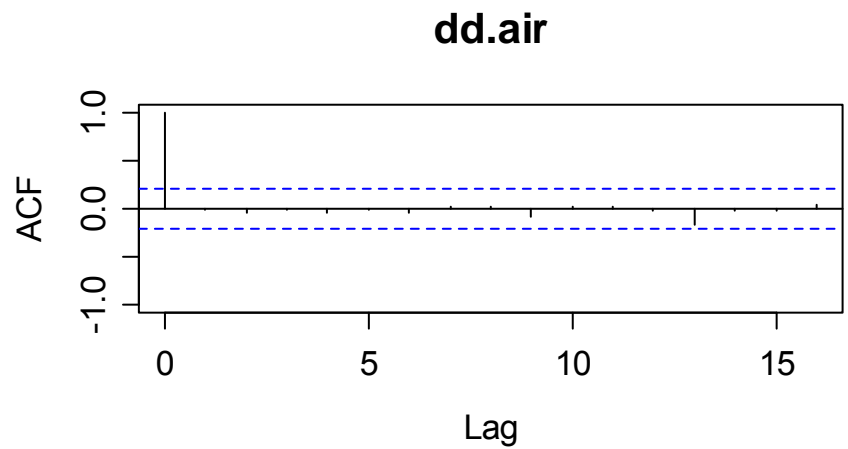
- transform all series in a clever way
- the transfer function model has identical coefficients
- the new, transformed input series is uncorrelated

**→ see blackboard for the transformation**

# Applied Time Series Analysis

## SS 2013 – Week 14

### *Borehole Transformed*





# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Borehole: Final Remarks***

- In the previous slide, we see the empirical cross correlations  $\hat{\rho}_{21}(k)$  of the two series  $D_t$  and  $Z_t$ .
- The coefficients  $\hat{v}_k$  from the transfer function model will be proportional to the empirical cross correlations. We can already now conjecture that the output is delayed by 1-2 days.
- The formula for the transfer function model coefficients is:

$$\hat{v}_k = \frac{\hat{\sigma}_Z}{\hat{\sigma}_D} \hat{\rho}_{21}(k)$$

# Applied Time Series Analysis

## SS 2013 – Week 14

### ***Borehole: R-Code and Results***

```
> dd.air <- resid(fit.air)
> coefs <- coef(fit.air)[1:5]
> zz.soil <- filter(diff(soil.na), c(1, -coefs, sides=1))
> as.int <- ts.intersect(dd.air, zz.soil)
> acf.val <- acf(as.int, na.action=na.pass)
```

### **Transfer Function Model Coefficients:**

```
> multip <- sd(zz.soil, na.rm=..)/sd(dd.air, na.rm=..)
> multip*acf.val$acf[,2,1]
[1] 0.054305137 0.165729551 0.250648114 0.008416697
[5] 0.036091971 0.042582917 -0.014780751 0.065008411
[9] -0.002900099 -0.001487220 -0.062670672 0.073479065
[13] -0.049352348 -0.060899602 -0.032943583 -0.025975790
[17] -0.057824007
```