# A Vector-Autoregressive “co-shiftable” Weathergenerator for Hydrodynamic Modeling of Lakes

Universität Stuttgart
Germany

D. Schlabing, M.Eder, M.Frassl, K.Rinke, A. Bárdossy

Dirk.Schlabing@iws.uni-stuttgart.de, Magdalena.Eder@iws.uni-stuttgart.de, Marieke.Frassl@uni-konstanz.de, karsten.rinke@ufz.de, Andras.Bardossy@iws.uni-stuttgart.de

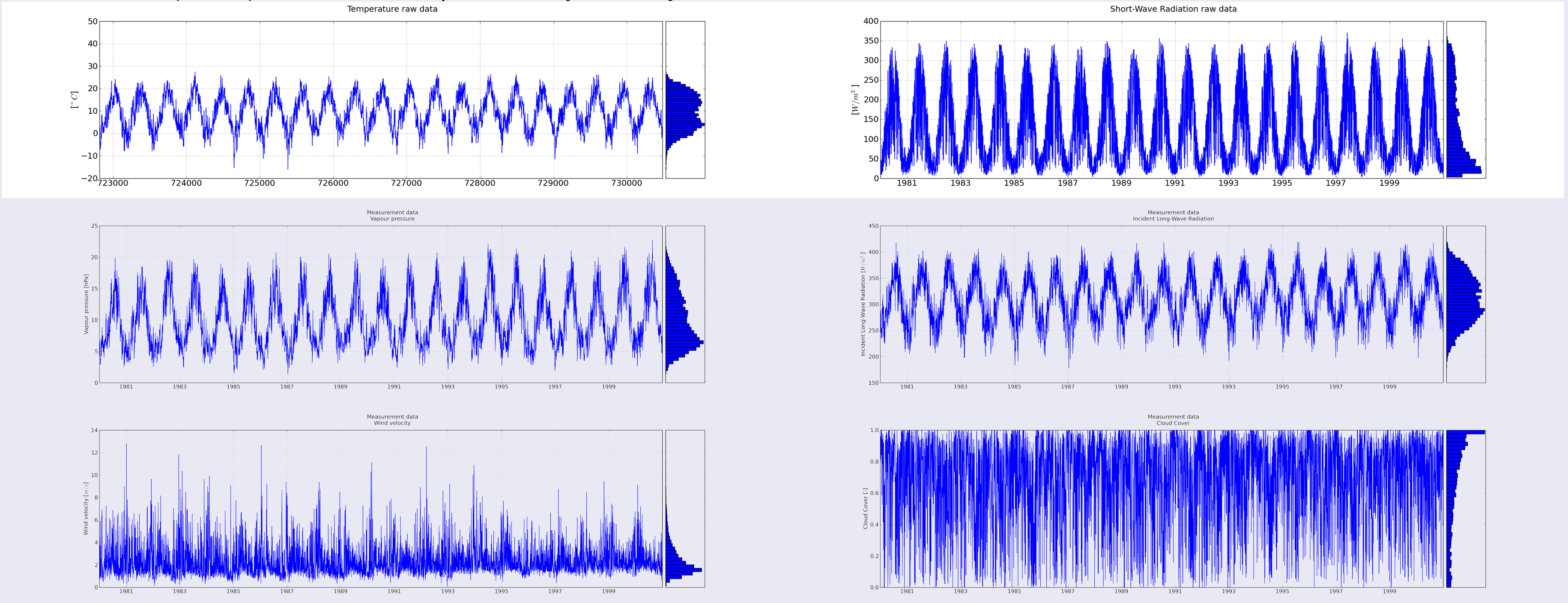
Institute of Hydraulic Engineering
Department of Hydrology and Geohydrology
Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy
Pfaffenwaldring 61, 70569 Stuttgart, Germany www.iws.uni-stuttgart.de

## Introduction

To use simulation as a tool to investigate the impact of possible climatic changes on lakes, input data reflecting these changes is needed. Using vector-autoregressive processes fitted to measured data, one can produce multivariate time series. These can be adjusted to reflect changes in meteorological conditions. However, simply changing values of one variable (e.g. increasing the mean of the temperature) treats the variables as being independent and thus neglects the dependency structure between them.
To overcome such problems, a vector autoregressive moving average weather generator was developed to generate multivariate time series that retain the statistical properties of the original data. The seasonalities, as well as covariances, auto- and cross-covariances are to be reproduced in synthetic time series.
Furthermore, to study the impact of changing climatic conditions on lakes, **first and/or second order moments of the temperature are modified while still trying to maintain the dependency structure**. In this sense, the weather generator is “co-shiftable”, meaning that changes in one variable will cause the other variables to change accordingly to the covariance matrix of the measured data.
The variables in question are air temperature, humidity, long- and shortwave radiation and wind. The Vector-Autoregressive Weathergenerator (VG) is primarily being designed to provide the means to model “What if?”-scenarios. For a work that uses VG and is mainly directed towards process understanding concerning the mixing behaviour of Lake Constance see *“Climate sensibility of a large lake - a scenario study using a 3D hydrodynamic model and a statistical weather generator* by Maria Magdalena Eder in Session *“Lakes and inland seas”* (HS10.2/OS2.3) (Friday, 08 Apr 10:30â12:00).

## Input Data

The available data was measured hourly in Constance, Germany during the period of 1980 to 2001 by the Deutscher Wetterdienst (DWD). In the subsequent analysis daily means were used.




## Modeling overview

1. Convert the input variables into standard-normal distributed variables.
2. Fit a Vector Autoregressive Moving Average Model (VARMA) to the converted variables.
3. Simulate using Multivariate-Gaussian distributed random disturbances with either:
   - no changes
   - increased mean for the temperature
   - increased variance for the temperature
   - increased mean and variance for the temperature
4. Converting back to the fitted marginal distribution.

## Transformation and De-seasonalization

The measured variables are converted into standard-normal distributed variables using a **Quantile-Quantile-Transformation**. To account for the strong seasonalities inherent in the data, the parameters of the theoretical distributions are described using triangular functions of different forms:

$$p_1(doy) = a + b\sin\left(\frac{2\pi doy}{365} + \phi\right) \tag{1}$$

$$p_2(doy) = a_1 - b_2\sin\left(\frac{2\pi doy}{365} + \phi\right)\left(a_2 - b_2\cos\left(\frac{2\pi doy}{365} + \phi\right)\right) \tag{2}$$

Where $p_i$ is any distribution parameter and *doy* is the day of the year. The choice of the form of triangular function depends on the variable and the chosen theoretical distribution. While the temperature can be described with a normal distribution and the mean of the temperature closely follows a simple sinus curve (see Figure **??**), such description is not sufficient for short-wave radiation. The parameters of the triangular functions are then estimated by **Maximum-Likelihood** using **Simulated Annealing**. Note that this is an optimization problem with constraints, as some distribution parameters (e.g. $\alpha$ and $\beta$ in the Beta distribution) are positive. This has been overcome by only drawing feasible solutions when generating a new candidate solution during Simulated Annealing.

## Fitting results

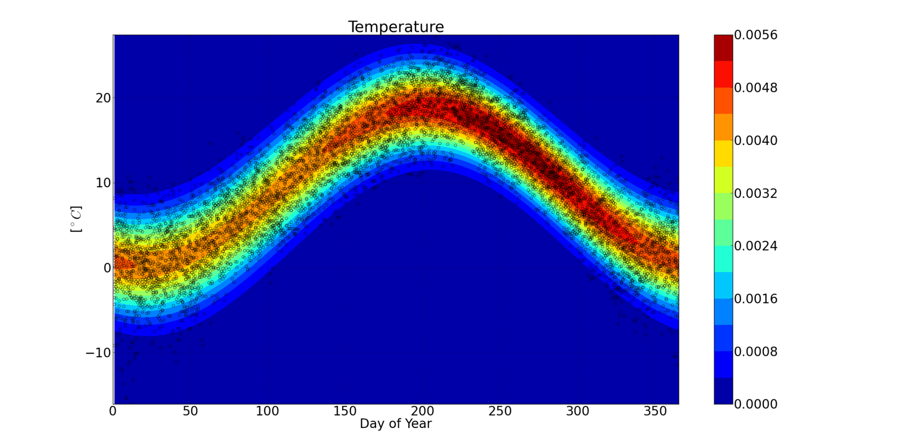



Figure: Normal distribution fitted to daily mean temperatures using equation **??** to describe mean and variance. Circles are the measurements, colours indicate the probability density.

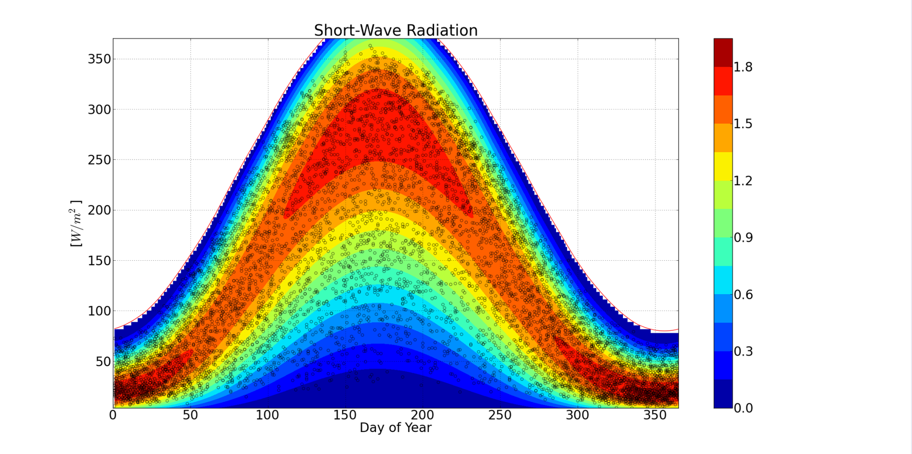



Figure: Kumaraswamy distribution fitted to daily mean short-wave radiations using equation **??** to describe the distribution parameters *a* and *b*. Circles are the measurements, colours indicate the probability density. (The Kumaraswamy distribution is similar to the Beta distribution. Its pdf: $f(x, a, b) = abx^{a-1}(1 - x^a)^{b-1}$ does not contain a transcendental function and is faster to evaluate.)

## The Vector Auto Regressive Moving Average (VARMA) Process

$$y_t = \mu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t + M_1 u_{t-1} + ... + M_q u_{t-q} \tag{3}$$

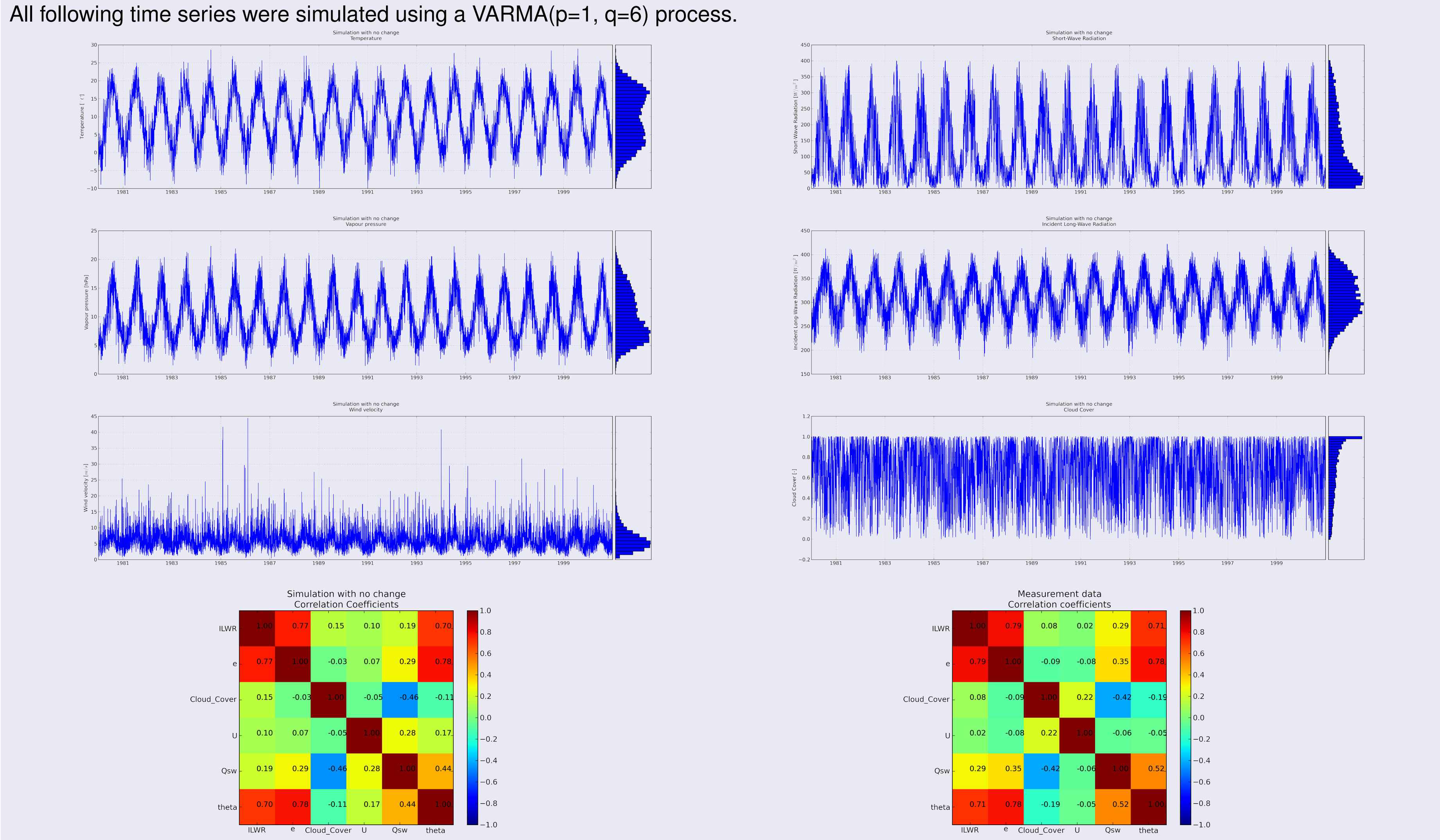
$y_t$ is a *K*-dimensional vector, *K* the number of variables, $\mu$ are the process means, $A_i$ are $(K \times K)$ matrices, *p* is the order of the auto regressive process, $u_t$ are residuals (i.e. the part of the time series the model cannot “explain”), $M_i$ are $(K \times K)$ matrices, *q* is the order of the moving average process.
Under the assumption that the process is Gaussian, a Likelihood function can be formulated. The problem, however, is non-linear and an iterative optimization algorithm (the *scoring algorithm*) is discussed in [**?**]. In this work a preliminary estimator was used that is otherwise the starting solution for the *scoring algorithm*. For obtaining the preliminary estimator one first needs to estimate the residuals $u_t$ with the help of a Vector Auto Regressive (VAR) model of high order $\hat{p}$ (higher than either *p* or *q*. The fitting of the VAR model can be done with the help of a least squares estimator (minimizing the residuals $\hat{u}_t$). The computation of the VARMA parameters is then again a matter of minimizing residuals.

## Simulating a VARMA Process

Generating synthetic time series can be done by setting starting vectors $\tilde{y}_i$ for $i \in \{1, .., p\}$ (e.g. to the process means $\mu$ approximated by the sample means $\overline{y}$) and applying equation **??** recursively. The disturbance vectors $\tilde{u}_t$ can be generated as multivariate Gaussian random numbers with the covariance matrix of the residuals $u_t$, $\Sigma_u$.
After fitting either VAR or VARMA processes, it became apparent that the residuals had a strong autocorrelation. So, when simulating using time-independent disturbance vectors, the simulated time series lacked the autocorrelation of the measured data. This was overcome by **generating disturbance vectors with a memory** by taking the mean of a newly disturbance vector and the *q* previous disturbance vectors. Autocorrelations of measured and simulated time series are shown in figure **??**.

## VARMA Simulation results

All following time series were simulated using a VARMA(p=1, q=6) process.




## Simulating with changed mean and variance

To attain different moments in the simulated time series one can change the disturbance vectors at each iteration step:

$$\tilde{y}_t = \overline{y} + A_1\tilde{y}_{t-1} + ... + A_p\tilde{y}_{t-p} + S\tilde{u}_t + m + M_1\tilde{u}_{t-1} + ... + M_q\tilde{u}_{t-q} \tag{4}$$

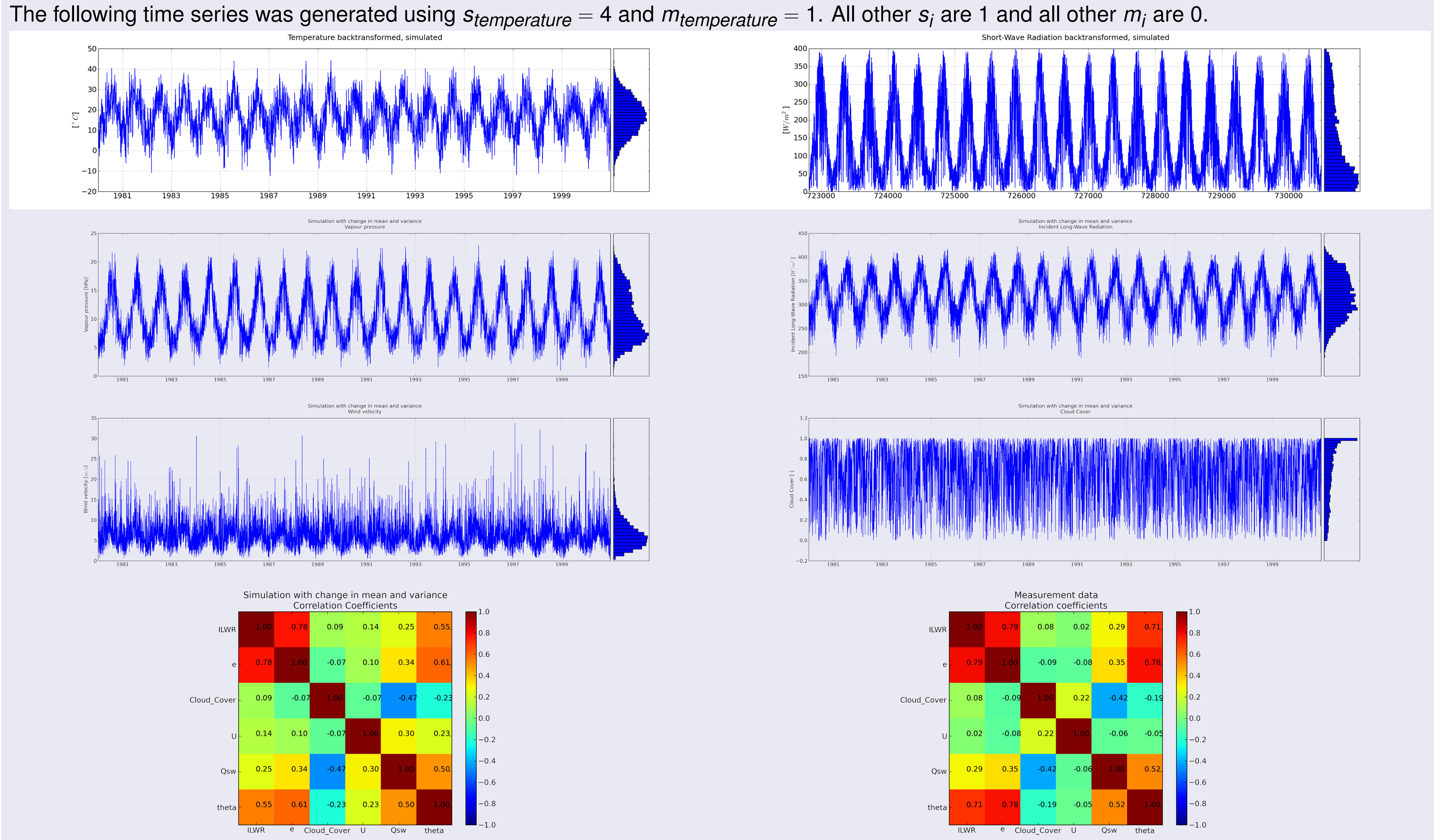
*S* is a matrix of the form

$$S = \begin{bmatrix} s_0 & 0 & \dots & 0 \\ 0 & s_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_K \end{bmatrix} \tag{5}$$

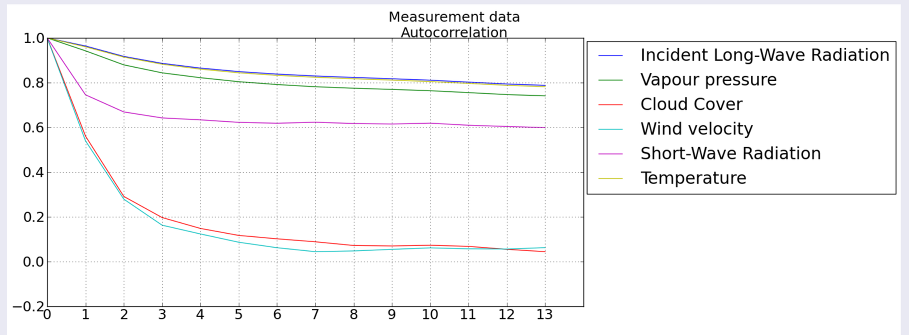
$s_i$ parameters to change variances, *m* is a vector to change means, respectively.
The covariance matrix of the time series $\tilde{y}_i$ ($\tilde{\Sigma}_y$) is slightly changed, since $\tilde{\Sigma}_u \neq Cov(S\tilde{u}_t)$. Yet the dependency structure does not change too drastically, since $\widetilde{\Sigma}_y$ is mainly influenced by $A_i$ and $M_i$.

## VARMA Simulation results: Changed mean and variance of temperature

The following time series was generated using $s_{temperature} = 4$ and $m_{temperature} = 1$. All other $s_i$ are 1 and all other $m_i$ are 0.




## Autocorrelations of measured and simulated data



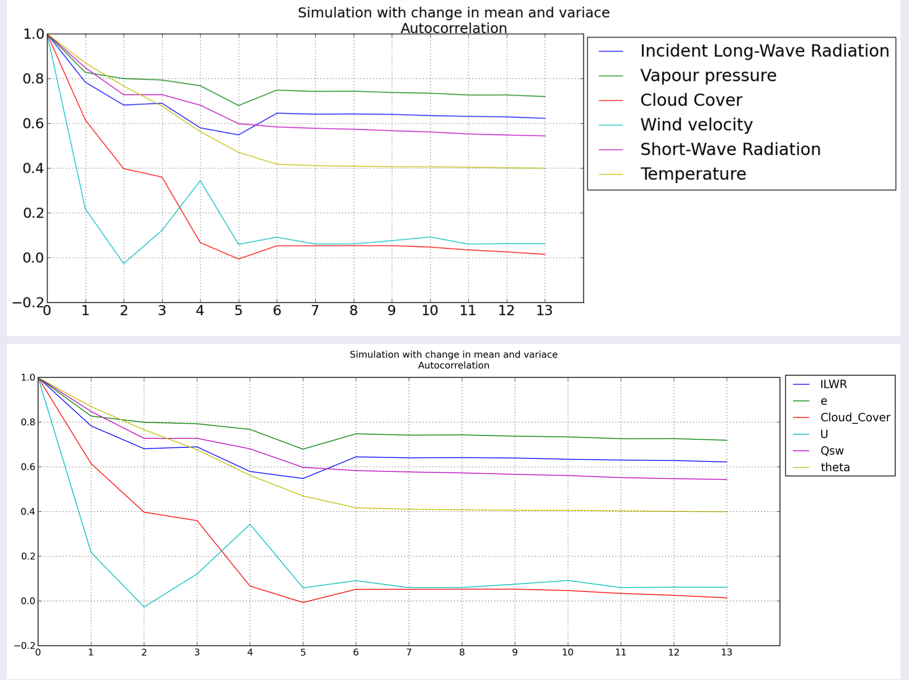




Figure: The different autocorrelations for the wind velocity probably come from the bad fit of the seasonal distribution.

## Summary

- Data that exhibits seasonalities can be transformed by describing its distribution parameters with triangular functions.
- Ordinary VARMA processes can be easily modified to simulate scenarios.

## Literature

H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer Verlag Berlin Heidelberg 2005