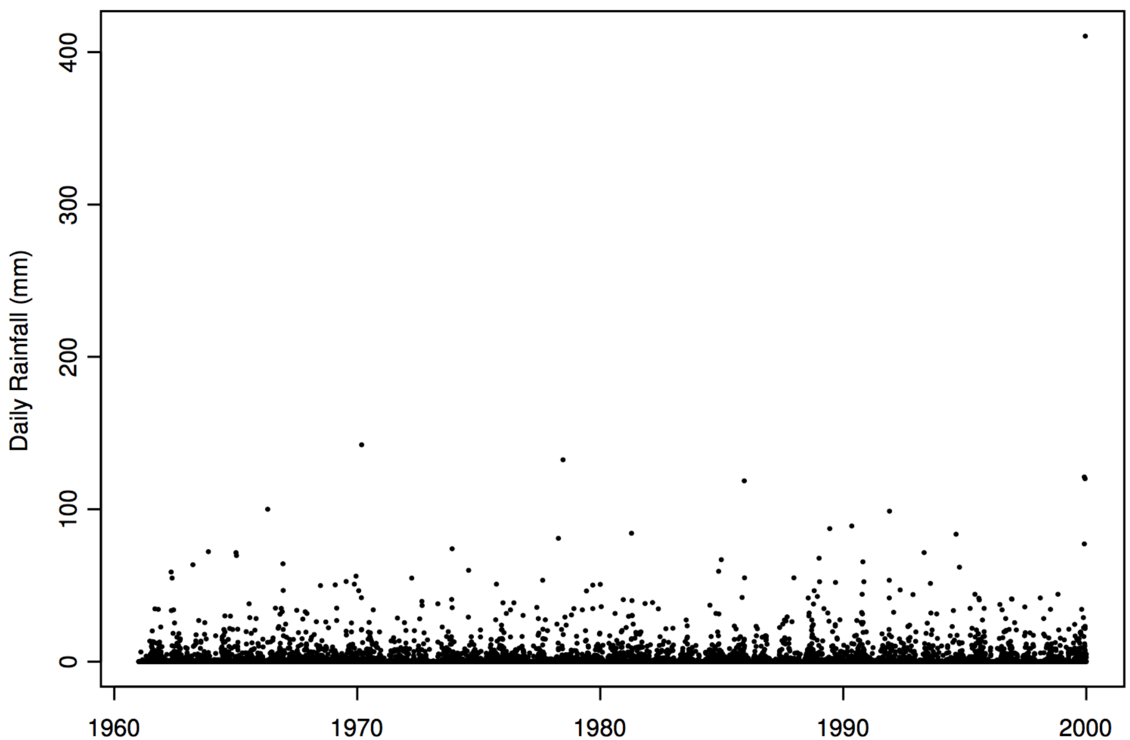
# Variability and Extremes

# How to model extremes?

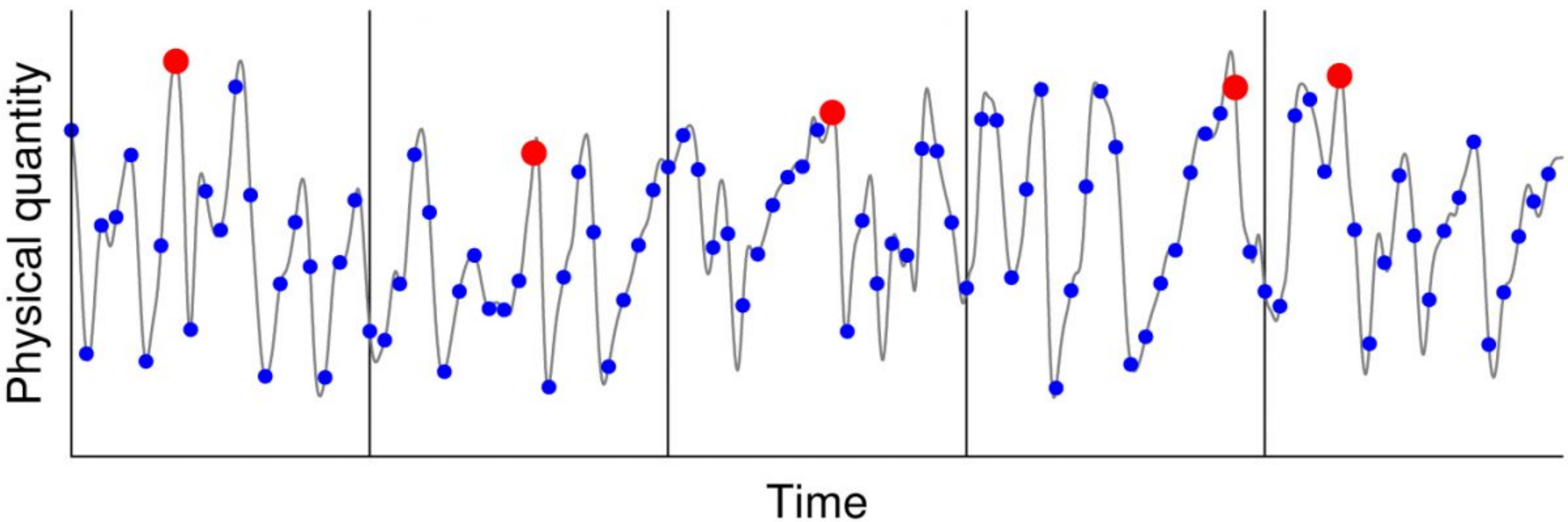
- Normal distribution not applicable to describe extremes
- Distributions often „heavy-tailed“: strong, rare outliers
- Example: Daily Rainfall in Venezuela




Source: Coles, Pericchi, Sisson: A Fully Probabilistic Approach to Extreme Rainfall Modeling (2003)

# The block-maxima approach

- group data into blocks of the same size, take for each block maximum value




- the resulting block maxima can be described with a Generalized Extreme Value (GEV) distribution

# The GEV distribution

- Task: The GEV distribution has three parameters. Find out their meanings!
- Simulate and plot GEV distribution data using the following code:

```
library(EnvStats)
plot(rgevd(500, x, y, z))
```

# The GEV distribution

- The three GEV distribution parameters are:
  - location ($\mu$)
  - scale ($\sigma > 0$)
  - shape ($\xi$)
- $\mu$ is a location (or shift) parameter:

$$\text{if } X \sim \mathrm{GEV}(\mu, \sigma, \xi), \text{ then } X+a \sim \mathrm{GEV}(\mu+a, \sigma, \xi)$$

- $\sigma$ describes the variability of the distribution:

$$\text{if } X \sim \mathrm{GEV}(\mu, \sigma, \xi), \text{ then } b \cdot X \sim \mathrm{GEV}(b \cdot \mu, b \cdot \sigma, \xi)$$

- $\xi$ describes the heavy-tailedness of the distribution:

$\xi > 0$: negative extremes ; $\xi < 0$: positive extremes

- the larger the absolute value of $\xi$, the stronger the extremes are

(Remark: In the literature, a different parametrization is also sometimes used. It replaces $\xi$ with $-\xi$ so that under that parametrization, positive values of $\xi$ correspond to positive extremes)

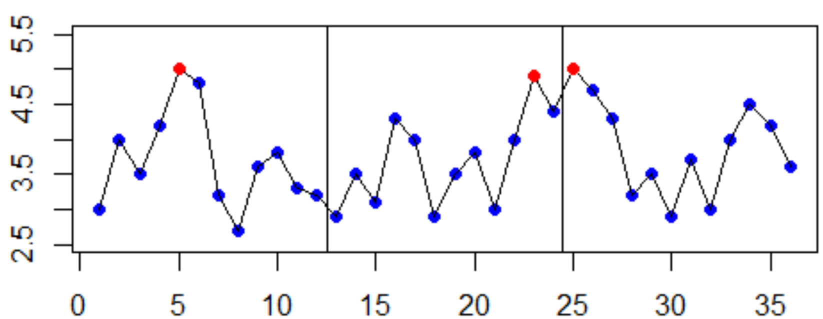
# When to apply the GEV distribution

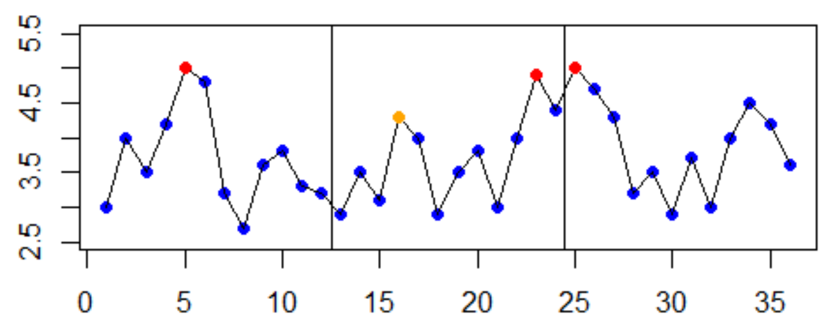
- The GEV distribution can be used to describe the distribution of block maxima if:
  - the block size is sufficiently large
  - the underlying (not block-maximized) data are independent
  - the distribution of the underlying data stays the same over time
- It can be used (almost) regardless of how the underlying data are distributed
- The GEV distribution is also often applicable in case of cyclic seasonality and if one season has a dominant effect on the maxima

- Example: annual maxima of daily data

# Decluserization

- How to deal with climate data that are not independent?
- Declustering algorithm: If the maxima from two adjacent blocks are very near to each other, it is assumed that they are due to the same climatological event
- In that case, the smaller of the two block maxima is discarded, and instead of it, the highest value of the block that is sufficiently far away from the other block is used



# Example Data

- Open the file „retreat.R“ and change the path in the first line to load „retreat_data.Rdata“
- Data set: monthly temperature from AWI-ESM-1-1-LR, historical run (1850-2000): `temperature_monthly`
- Dimensions: `lon, lat, t_m`

- Run lines 1 through 38 and test the plotting function:

```
# January 1850:
plot_map(temperature_monthly[,,1],lo=-50,hi=50)
# July 1850:
plot_map(temperature_monthly[,,7],lo=-50,hi=50)
```

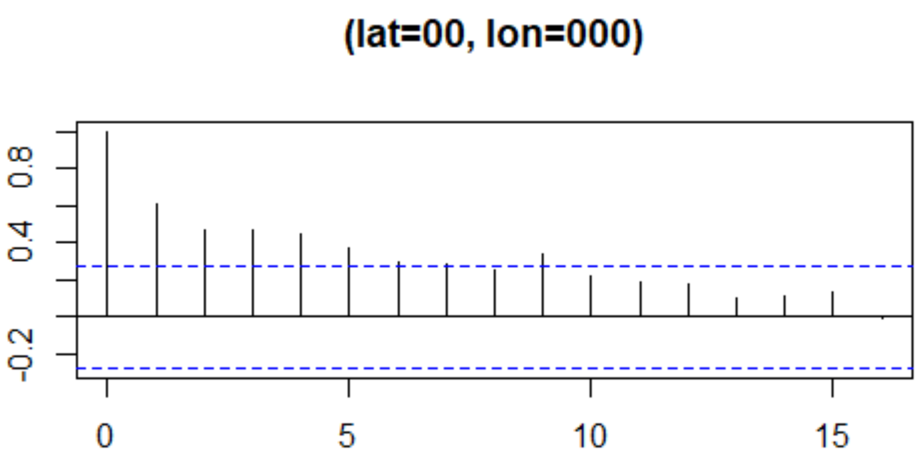
- Plot a time series using `get_timeseries(data, lat, lon)`:

```
plot(t_m, get_timeseries(temperature_monthly,
                                   00, 000), type=„l“)
```

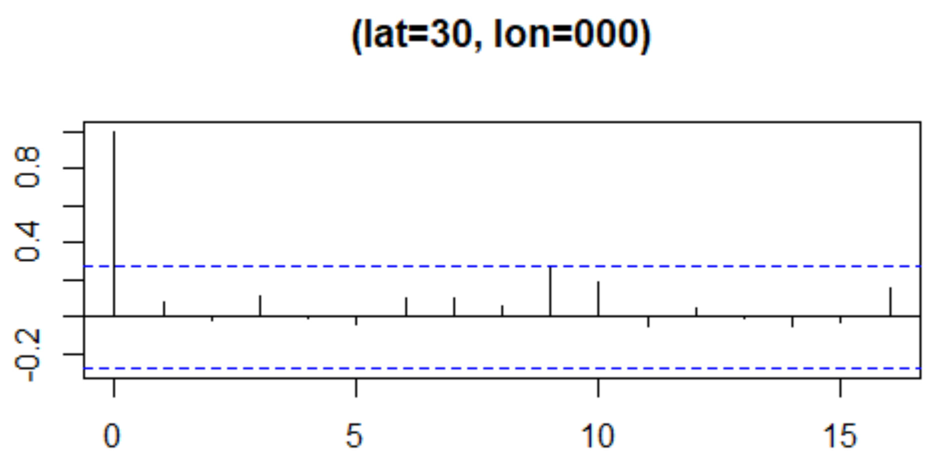
# Example Data

- Follow lines 42 through 97: The functions `block_maximize` and `block_maximize_declustered` are defined and used to group the data into blocks
- Check for autocorrelation:

```
acf(get_timeseries(temperature_threeyearly, 0, 0))
```




Autocorrelation (decadal trends?)

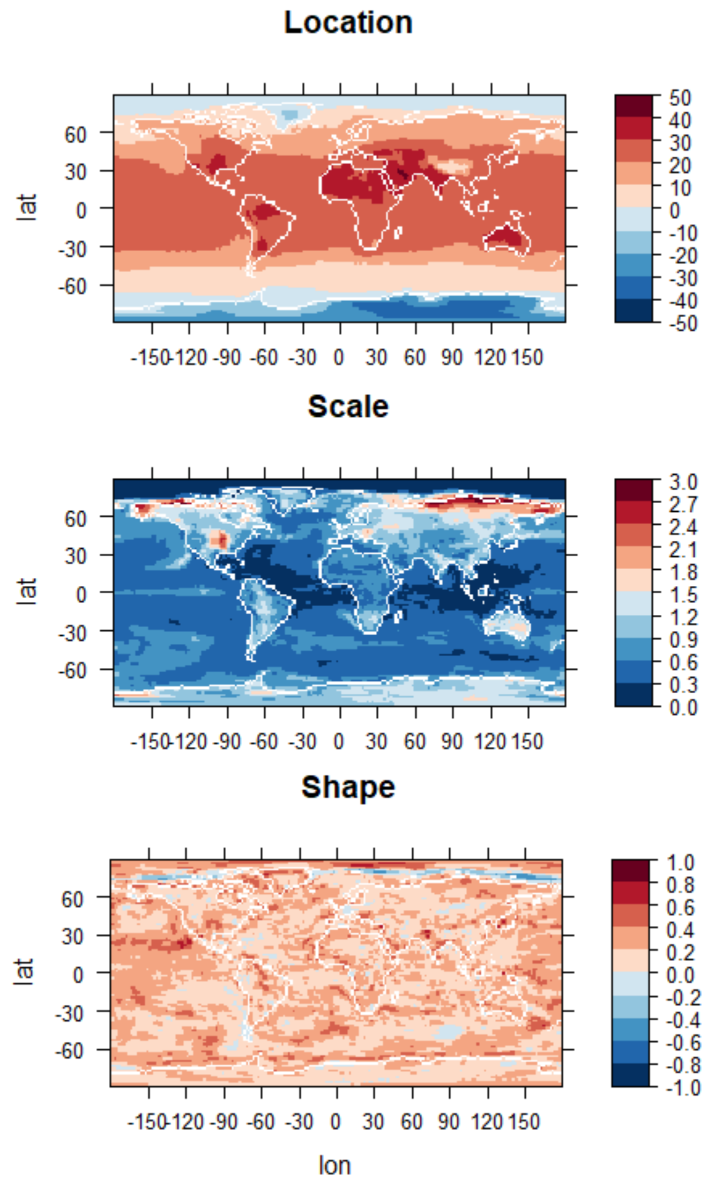



No significant autocorrelation

# Fitting a GEV distribution

- Follow lines 100 through 106: The function `egevd` from the R package „EnvStats“ is used to fit GEV distributions to the block-maximized data
- Different estimation methods exist. For small sample sizes, the PWME method (Probability Weighted Moments Estimator) is better suited than the default Maximum Likelihood method
- Inspect the geographical distribution of the GEV parameters using `plot_map`

# Fitted GEV parameters



# Goodness-of-Fit Test

- We can use a Kolmogorov-Smirnov Test (KS-Test) to test the goodness of fit
- Null hypothesis: The data are GEV distributed with given parameters
- Use a KS-Test on a time series and with the estimated GEV parameters (Lines 108 through 114)

# Non-stationary GEV distributions

- To investigate a changing climate, a non-stationary GEV distribution must be used
- The parameters are allowed to change over time:

$$X(t) \sim GEV(\mu(t),\sigma(t),\xi(t))$$

- Time-dependent GEV distributions can be fitted using Maximum Likelihood, numerical optimization is necessary

# A simple non-stationary model

- We use the following model:

$$\mu(t) = \mu_{start} + \mu_{change} \cdot (t - t_{start})/(t_{end} - t_{start})$$

$\sigma$ constant over time

$\xi$ constant over time

- four parameters: $\mu_{start}$, $\mu_{change}$, $\sigma$, $\xi$
- Fit model to data using the function

```
MLE_est(time_series, print=F, plot=F)
```

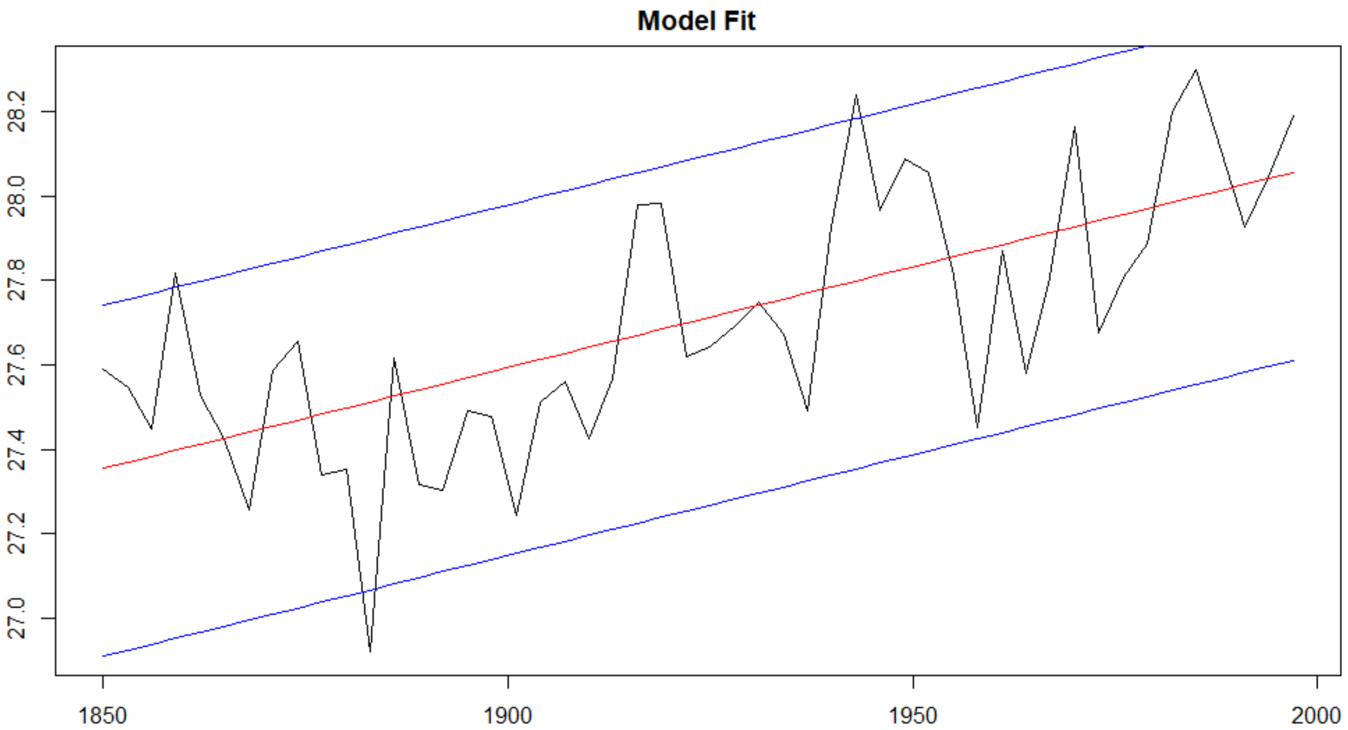
# The function MLE_est

- The non-stationary model for each GEV parameter is contained as subfunction in `MLE_est`
- For each parameter to be estimated, upper and lower bounds must be prescribed
- For the numerical optimization, the function `optim` is used
- `optim` is run several times with different starting values, to ensure that a global maximum is found
- use `print=T` to check the results of the iterations. If they are too different, use more iterations or narrower parameter bounds

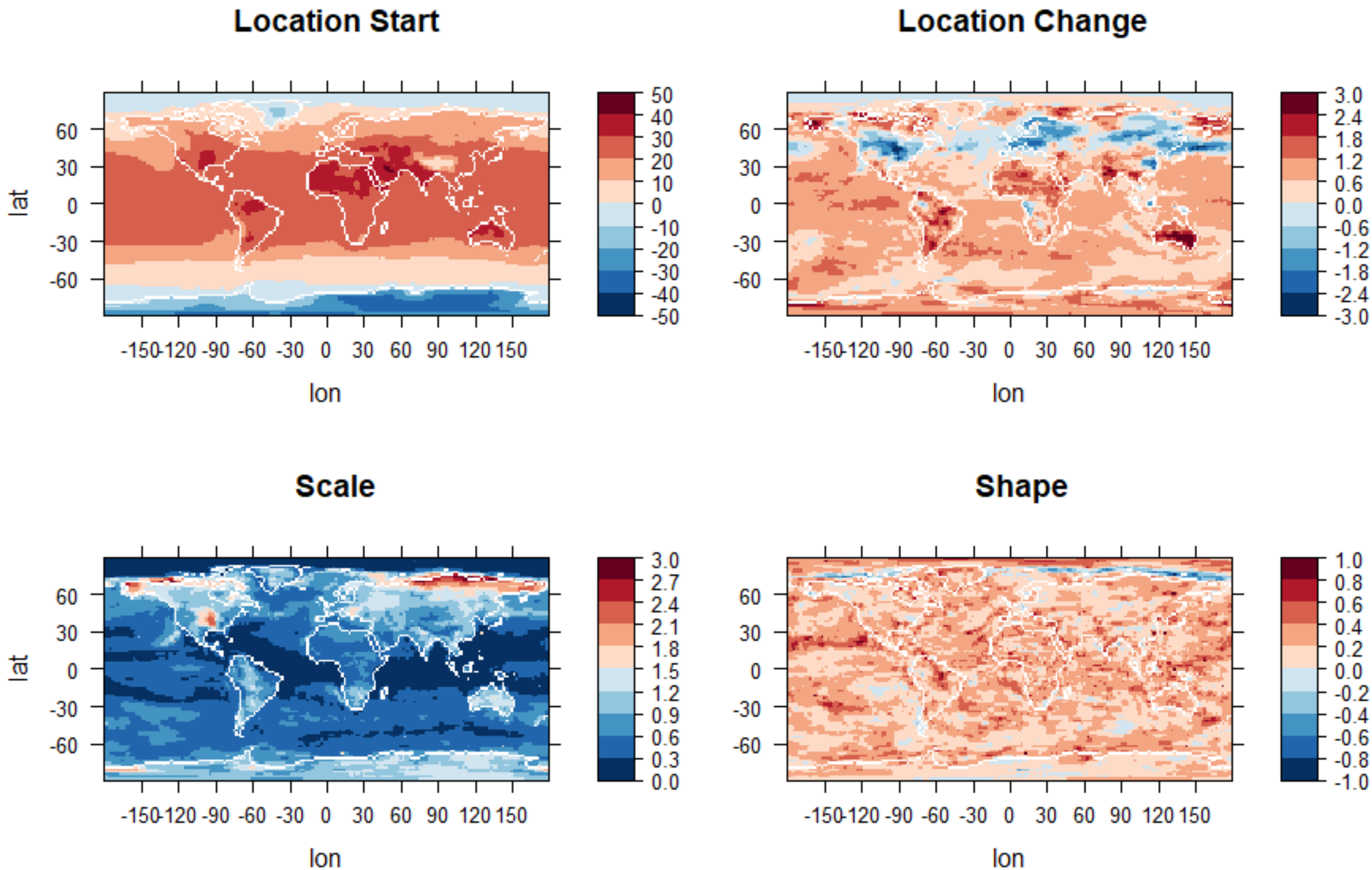
# The function MLE_est

- use the option `plot=T` to see a plot of the time series together with the median and 95% quantiles of the fitted model

```
MLE_est(get_timeseries(temperature_threeyearly,
                        00, 000), plot=T)
```



# Results


Location Start
Location Change
Scale
Shape
lat
lon
60 30 0 -30 -60
-150 -120 -90 -60 -30 0 30 60 90 120 150
50 40 30 20 10 0 -10 -20 -30 -40 -50
3.0 2.4 1.8 1.2 0.6 0.0 -0.6 -1.2 -1.8 -2.4 -3.0
3.0 2.7 2.4 2.1 1.8 1.5 1.2 0.9 0.6 0.3 0.0
1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0

# Task: Adapt MLE_est

- Task: Adapt the function `MLE_est` to the following model:

$$\mu(t) = \mu_{start} + \mu_{change} \cdot (t-t_{start})/(t_{end}-t_{start})$$

$$\sigma(t) = \sigma_{start} + (\sigma_{end}-\sigma_{start}) \cdot (t-t_{start})/(t_{end}-t_{start})$$

$\xi$ constant over time

# Solution

```
parnames <- c(„loc_start“, „loc_change“,
                  „scale_start“, „scale_end“, „shape)

t_start = tm[1] ; t_end = tm[length(tm)]
loc_model <- function(par) {
  return(par[1] + par[2]*(tm-t_start)/(t_end-t_start)
}
sc_model <- function(par) {
  return(par[3] + (par[4]-par[3])*(tm-t_start)/(t_end-
t_start)
}
sha_model <- function(par) {
  return(rep(par[5], length(tm))
}
parameters_lower_bound <- c(min(ts)-1, -3, 0.01, 0.01, -1)
parameters_upper_bound <- c(max(ts)+1,  3,    5,    5,  1)
```