INTRODUCTION OF A GLM TO THE RAINFALL SIMULATION FOR THE MANAGEMENT OF THE NESTOS RIVER BASIN

Iordanidis Ilias[1]
Anagnostopoulos Petros[2]

1: PhD Student, AUTh, Dept.of Civil Engineering, Hydraulics Laboratory, Thessaloniki 54124, Greece
Project Engineer, YDROFILI S.A.

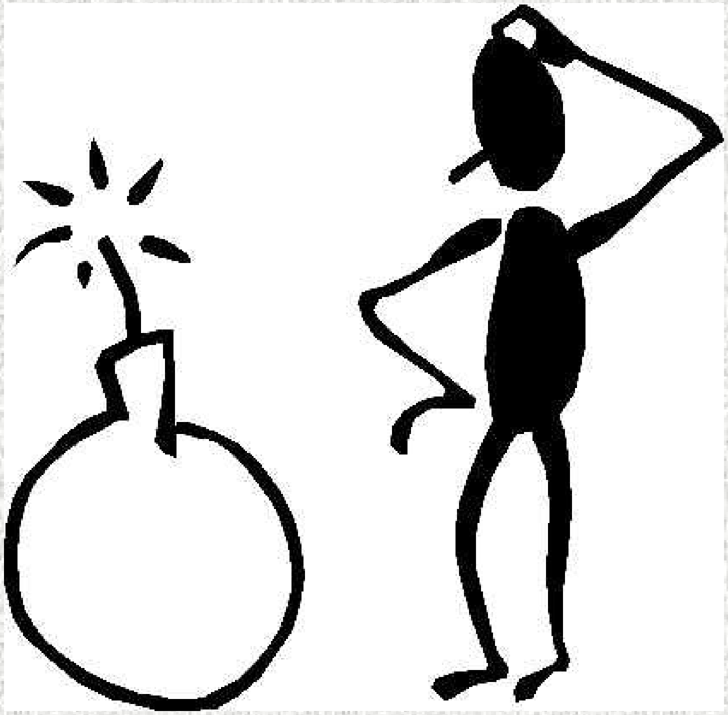
2: Professor, AUTh, Dept.of Civil Engineering, Hydraulics Laboratory, Thessaloniki 54124, Greece

e-mail: iordanid@civil.auth.gr

IN A FEW WORDS…

Data not always available to researchers for various reasons.

This results in difficulties in Research, Design Projects or Management Decisions



THE REASONS

- Data do not exist
- Data cover a narrow time span
- Data availability is limited due to restrictions by the respective authorities
- Study areas are shared by neighbouring countries which are reluctant to share their data

A PROPOSED SOLUTION

The use of a Generalised Linear Model (GLM) in order to produce stochastic series to compensate for the lack of data.

The selection of a reliable GLM will give a good first insight for the behaviour of a basin for a long time period.

THE MODEL

The model selected was GlimClim

GlimClim is a freeware program for research purposes, created by Dr. Richard Chandler of Imperial College London.

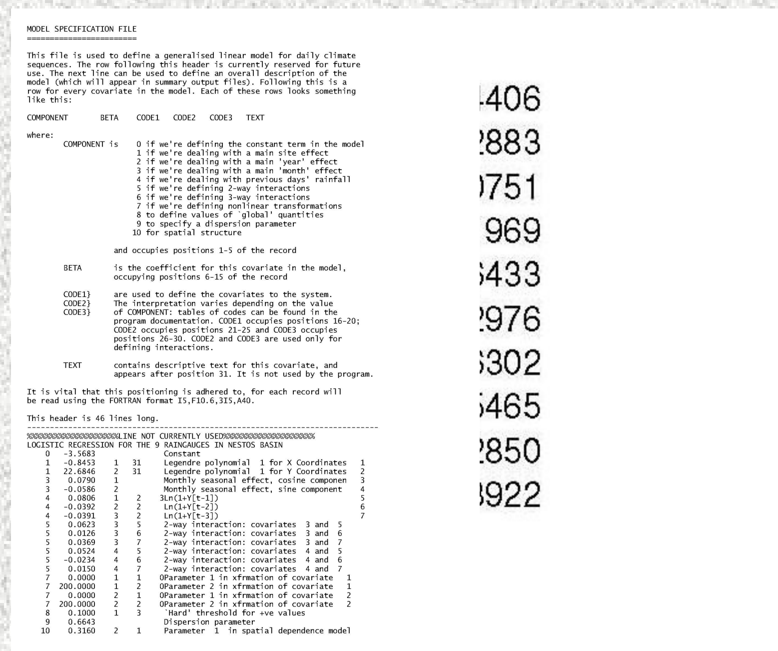
Incorporates the theory of GLMs in a light and user friendly package

HOW DOES IT WORK?

?

Input Data:

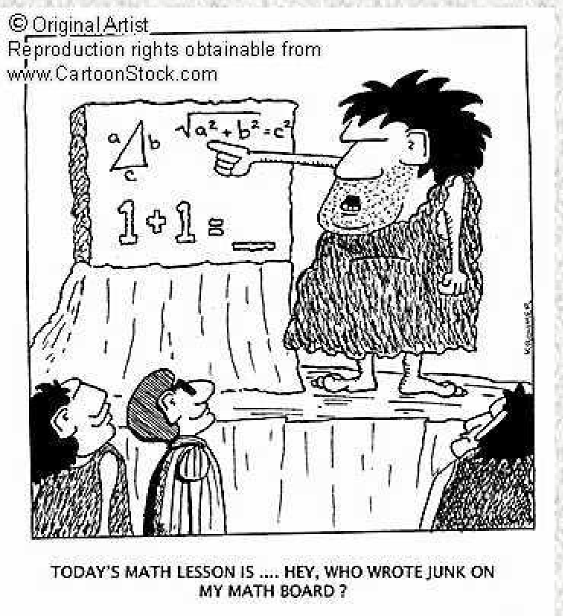
- Daily Rainfall Data
- Station Coordinates
- Model Parameters



```
MODEL SPECIFICATION FILE
========================

This file is used to define a generalised linear model for daily climate
sequences. The row following this header is currently reserved for future
use. The next line can be used to define an overall description of the
model (which will appear in summary output files). Following this is a
row for every covariate in the model. Each of these rows looks something
like this:

COMPONENT       BETA    CODE1   CODE2   CODE3   TEXT

where:
        COMPONENT is    0 if we're defining the constant term in the model
                        1 if we're dealing with a main site effect
                        2 if we're dealing with a main 'year' effect
                        3 if we're dealing with a main 'month' effect
                        4 if we're dealing with previous days' rainfall
                        5 if we're defining 2-way interactions
                        6 if we're defining 3-way interactions
                        7 if we're defining nonlinear transformations
                        8 to define values of `global' quantities
                        9 to specify a dispersion parameter
                       10 for spatial structure

                   and occupies positions 1-5 of the record

        BETA       is the coefficient for this covariate in the model,
                   occupying positions 6-15 of the record

        CODE1}     are used to define the covariates to the system.
        CODE2}     The interpretation varies depending on the value
        CODE3}     of COMPONENT: tables of codes can be found in the
                   program documentation. CODE1 occupies positions 16-20;
                   CODE2 occupies positions 21-25 and CODE3 occupies
                   positions 26-30. CODE2 and CODE3 are used only for
                   defining interactions.

        TEXT       contains descriptive text for this covariate, and
                   appears after position 31. It is not used by the program.

It is vital that this positioning is adhered to, for each record will
be read using the FORTRAN format I5,F10.6,3I5,A40.

This header is 46 lines long.
--------------------------------------------------------------------------------
%%%%%%%%%%%%%%%%%%%%LINE NOT CURRENTLY USED%%%%%%%%%%%%%%%%%%%%
LOGISTIC REGRESSION FOR THE 9 RAINGAUGES IN NESTOS BASIN
    0   -3.5683               Constant
    1   -0.8453    1   31     Legendre polynomial  1 for X Coordinates   1
    1   22.6846    2   31     Legendre polynomial  1 for Y Coordinates   2
    3    0.0790    1          Monthly seasonal effect, cosine componen   3
    3   -0.0586    2          Monthly seasonal effect, sine component    4
    4    0.0806    1    2    3Ln(1+Y[t-1])                               5
    4   -0.0392    2    2     Ln(1+Y[t-2])                               6
    4   -0.0391    3    2     Ln(1+Y[t-3])                               7
    5    0.0623    3    5     2-way interaction: covariates  3 and  5
    5    0.0126    3    6     2-way interaction: covariates  3 and  6
    5    0.0369    3    7     2-way interaction: covariates  3 and  7
    5    0.0524    4    5     2-way interaction: covariates  4 and  5
    5   -0.0234    4    6     2-way interaction: covariates  4 and  6
    5    0.0150    4    7     2-way interaction: covariates  4 and  7
    7    0.0000    1    1    0Parameter 1 in xfrmation of covariate   1
    7  200.0000    1    2    0Parameter 2 in xfrmation of covariate   1
    7    0.0000    2    1    0Parameter 1 in xfrmation of covariate   2
    7  200.0000    2    2    0Parameter 2 in xfrmation of covariate   2
    8    0.1000    1    3     `Hard' threshold for +ve values
    9    0.6643               Dispersion parameter
   10    0.3160    2    1     Parameter  1  in spatial dependence model
```

MATHEMATICAL BACKGROUND

(a very brief overview)



GlimClim consists of two mathematical models and the simulation software:

- Logistic model (Events model)
- Gamma Model (Magnitude Model)
- Simrain (simulation software)

THE LOGISTIC MODEL

Creates the rainfall events for the study period (I / 0 model)

$$\ln\left(\frac{p_i}{1-p_i}\right) = x_i^T \beta$$

p_i : probability of rain on a given day

β : vector of factors

x_i : vector of variables

THE GAMMA MODEL

Calculates the amount of a rainfall of a non-zero event (amounts model)

$$\ln(\mu_i) = \xi_i^T \gamma$$

μ_i: mean of the Gamma distribution for the i^{th} day

ξ_i: vector of variables

γ : vector of factors

CALIBRATION PROCESS

The model parameters are fitted using the available data.

Users can pre-define the interactions between variables.

Complexity of interactions is gradually increased until a sufficient level of reliability is reached (measured by the value of log-likelihood)

The Logistic model is fitted first and then the Gamma model.

When the calibration is finished, the user can continue with the…

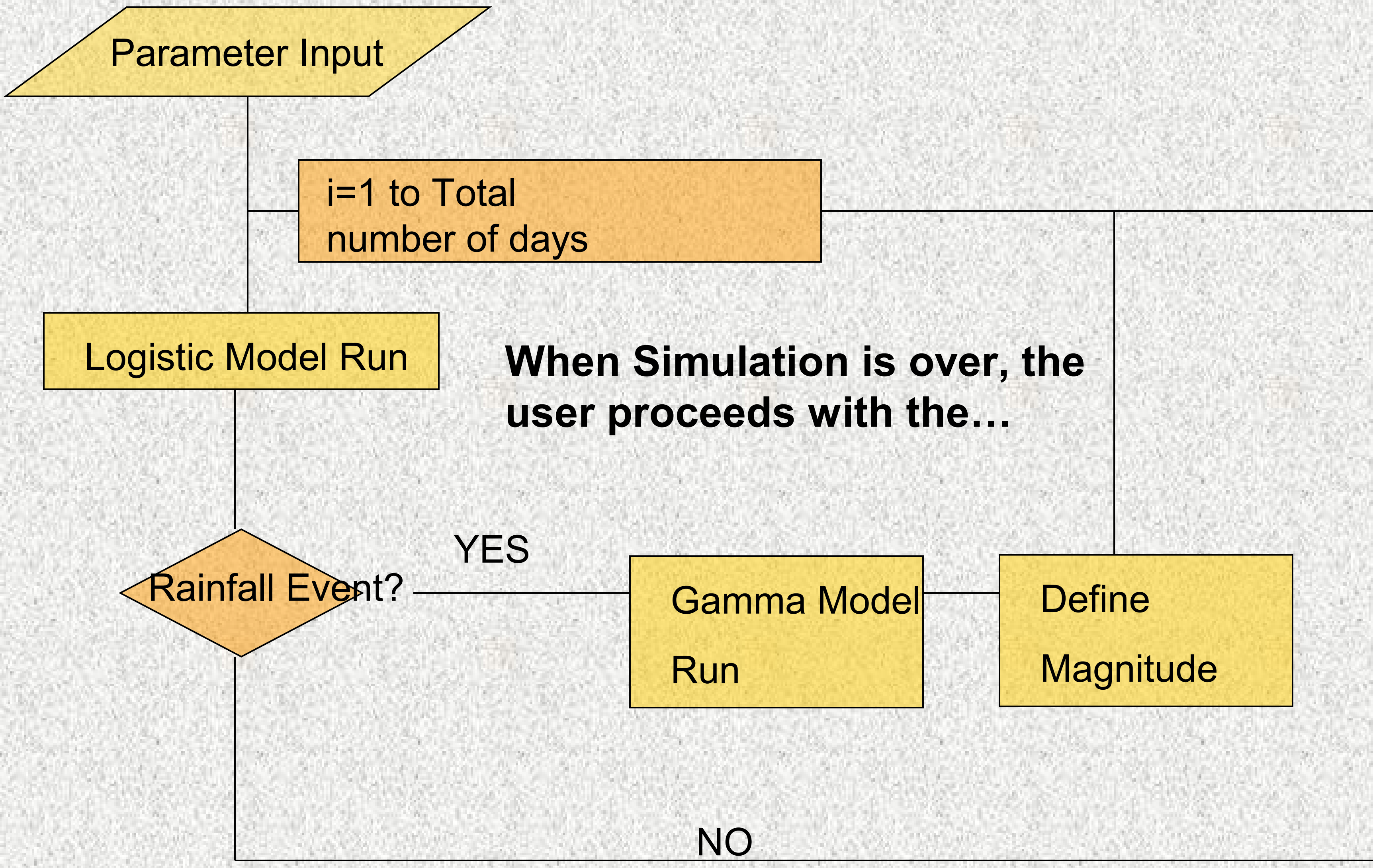
SIMULATION

Simulation process is very automated.

Users define the desired time span and format of the outputs and proceeds with the simulation process

SIMULATION DIAGRAM

(when running the SimRain component)



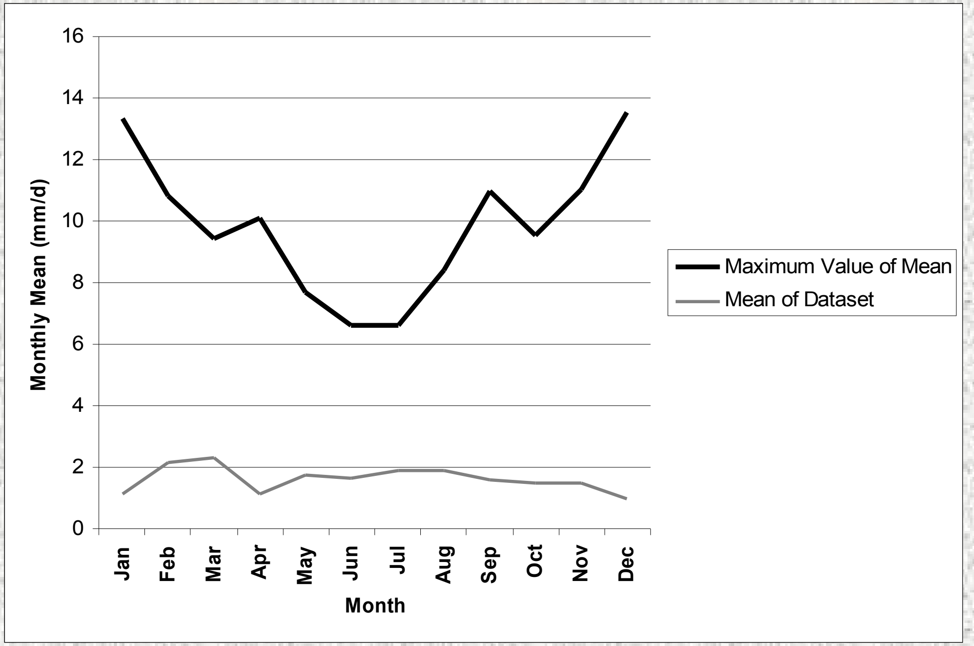
When Simulation is over, the user proceeds with the…

STATISTICAL CHECKS

Statistical checks are important in order to verify the reliability of the results.

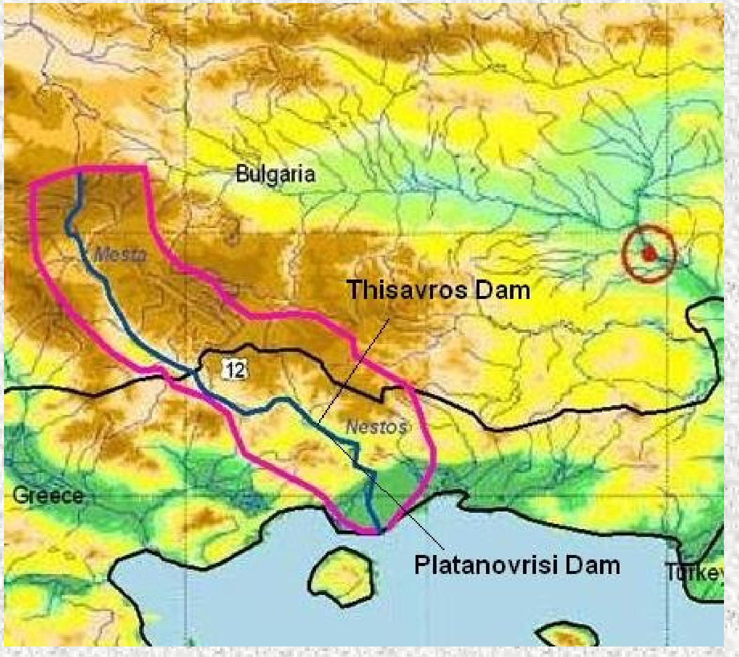
A common check :

To examine if the values certain statistical figures (mean, standard deviation etc.) of the data are within the maximum and minimum of those values for the simulated data.



AREA OF STUDY

The Basin of the Nestos River



The river has a total length of 250km and is shared by Greece and Bulgaria.

150 km run in Greece, where the Thisavros and Platanovrisi Dam are also located.

PURPOSE OF THE STUDY

To create a sufficient and reliable stochastic rainfall dataset for the basin, which could be used in rainfall-runoff analyses.

RAINFALL GAUGES

Data from 9 gauges were used

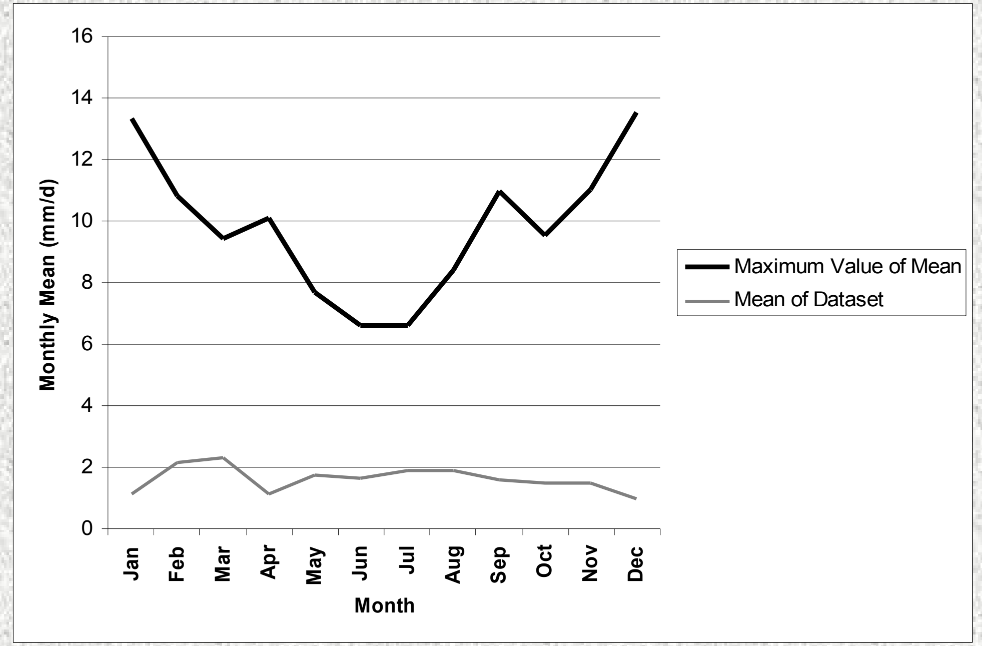
AVAILABLE DATA

Available data cover a time span of 20 years (daily step)

The dataset is no complete for this period of time

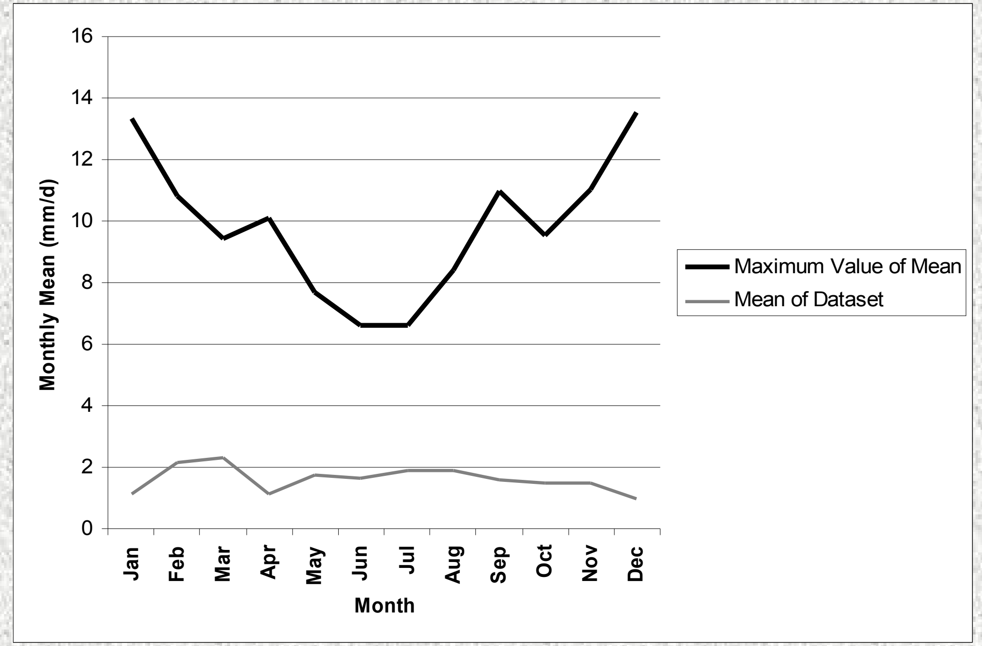
There is a season of drought for 4 years in the dataset.

SIMULATION



Simulation was conducted according to the software's requirements and specifications. The results were statistically checked, and extreme values were also determined in the results.

RESULTS



The results appear to be consistent with rainfall conditions in the area.

This fulfills the primary goal, to obtain an insight to basin rainfall status

Extreme values also appear between long periods of time, indicating that the software can handle extreme values consistently as well.

COMMENTS ON THE USE OF GlimClim

- Most reliable for basins up to 100x100 km.
- Can be used to produce stochastic datasets for future periods
- Useful to fill-in missing data
- Works better in Monte Carlo-type analyses.
- Any data deficiency introduces uncertainties to the results.

FUTURE APPLICATION

To use GlimClim in order to obtain stochastic data for a neighbouring area with no data available.

CONCLUSIONS

GlimClim produces reliable results even with datasets which are not complete.

Extreme values are also handled consistently by the programme.

Researchers can use GlimClim in order to continue their research despite difficulties.

APPLICATION IN SHARED STUDY AREAS

The GLM can be used to create datasets for non-accessible areas

and give some first insight to the behaviour of a basin.

It is an easy to use tool which can help overcome some difficulties when a researcher is trying to obtain data.

However……

No software should ever replace co-operation between scientists and countries.

"They don't trust each other to share research."

THANK YOU!