Estimation of uncertainty in satellite-based area average rainfall for crop yield forecasting using sequential simulation

David Grimes
Teo Chee Kiat
Bernard Gomez

TAMSAT*
Dept of Meteorology
University of Reading
U.K.

Thanks to the Gambian Meteorology and Agriculture Services for provision of rainfall and crop yield data

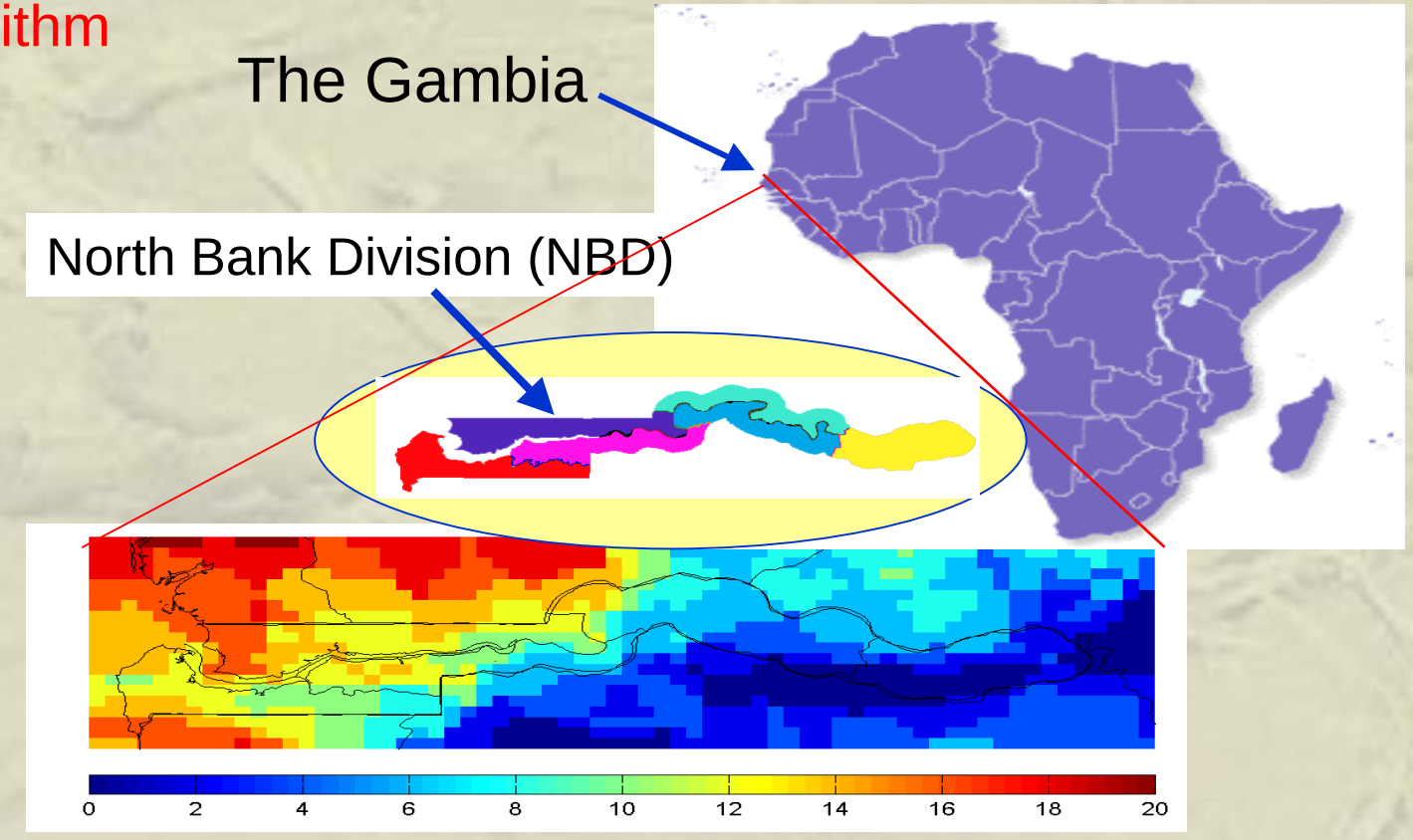
TAMSAT = *T*ropical *A*pplications of *M*eteorology using *SAT*ellite data

Overview of talk

- Case study – groundnut yield in the Gambia
- Principles of method - Sequential Simulation – step by step guide
- Results and validation
- Application to crop yield forecasting

Case study

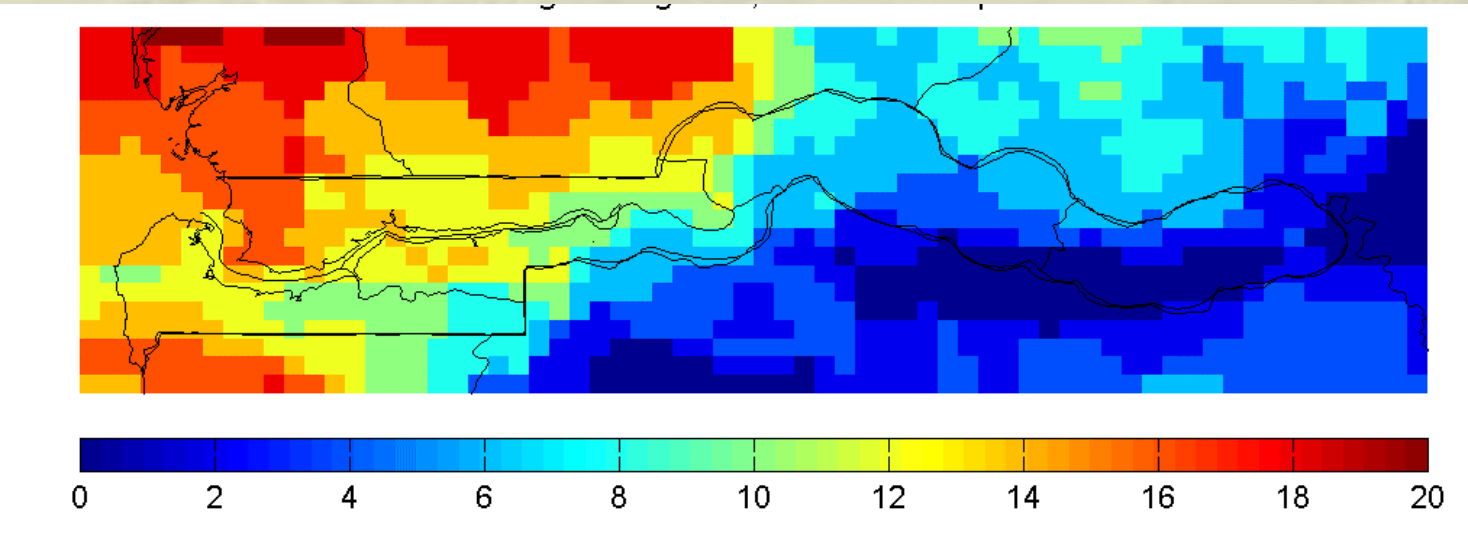
- Location- the Gambia
- Area average daily rainfall for crop yield forecasting for each district is provided by Cold Cloud Duration (CCD)
- NB uncertainty estimation is applicable to any rainfall algorithm



Rainfall field 7/8/01

We can estimate the mean areal rainfall by averaging over all pixels within the target area

- **but how do we assess the *uncertainty* on the mean rainfall amount?**



We must take account of:

1. uncertainty in number of rainy pixels
2. uncertainty in rainfall amount in each pixel
3. spatial correlation between pixels in respect of ‘raininess’ and rainfall amount

Overview of uncertainty estimation method

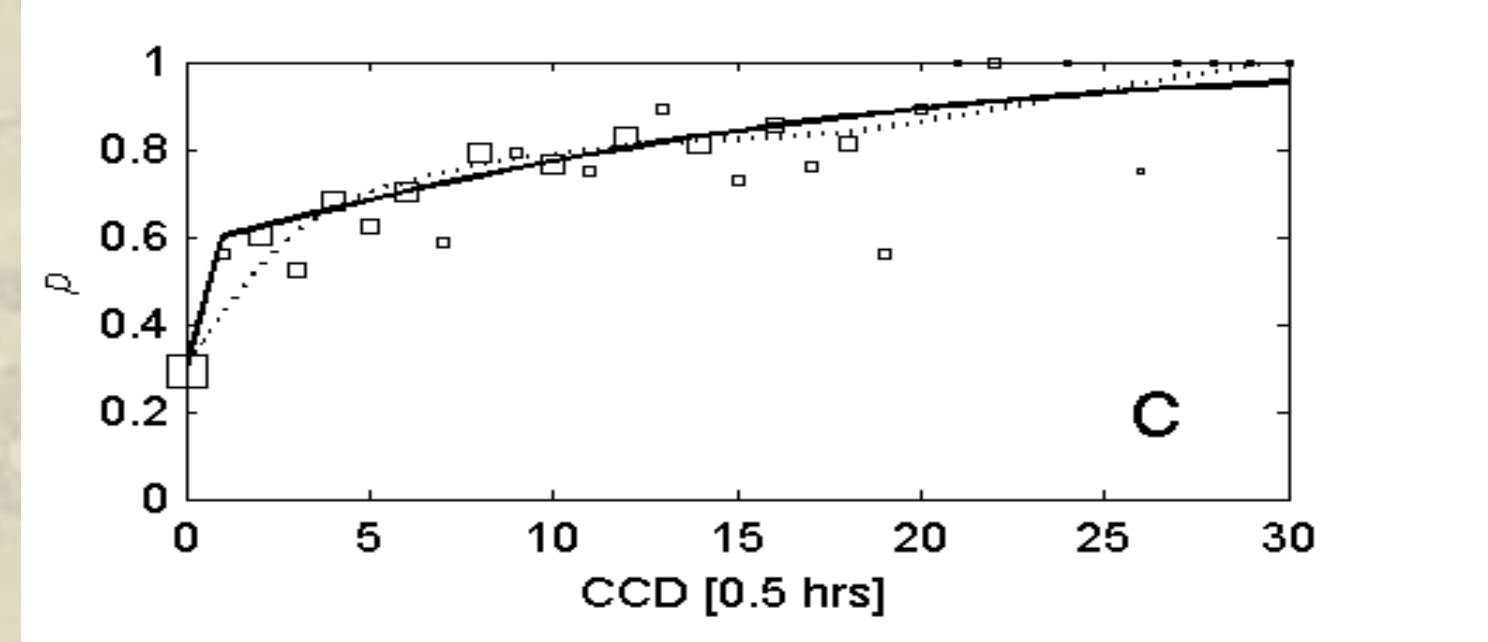
1. **Compare statistics of CCD and gauge data.**
   - *Convert gauge data to same spatial scale as satellite data*
   - *Estimate probability of rain/no-rain for given CCD*
   - *Determine distribution of rainfall amount for given CCD.*

2. **Use the statistical relationships to generate an ensemble of rainfall fields**
   - *For each pixel, the ensemble of values must agree with the observed statistics*
   - *For each ensemble member, the inter-pixel correlation must agree with the observations*

3. **Calculate area average rainfall from each ensemble member**

4. **Estimate the uncertainty from the distribution of average rainfall values over the ensemble**

Step 1: Comparison of satellite and gauge data – probability of rain



$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \times CCD, CCD > 0$$

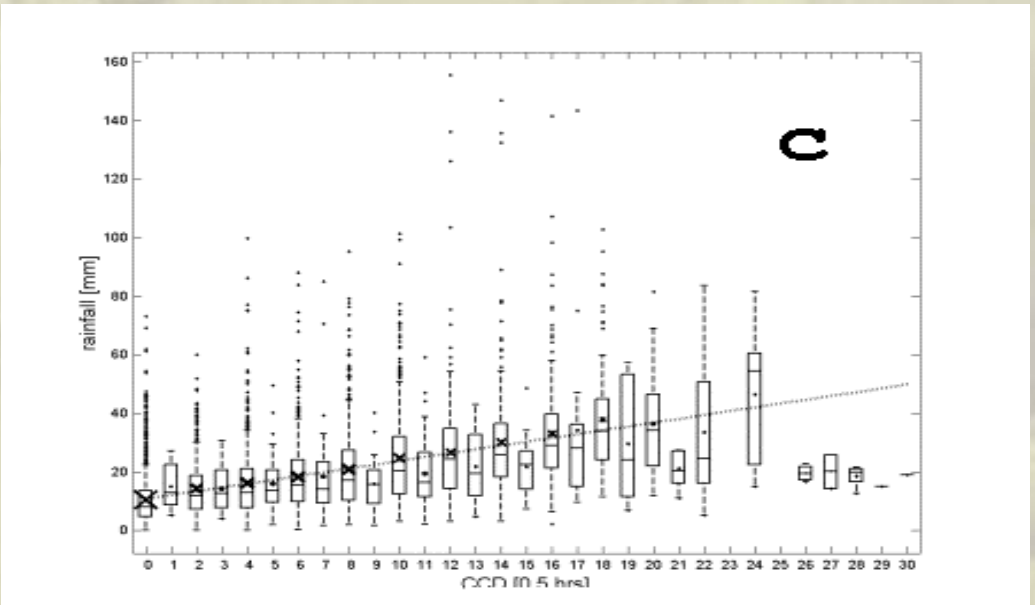
$$p = p_0 \; for \; CCD = 0 \qquad (1)$$

Use existing data to calibrate for p_0, b_0, b_1

Step 1: Comparison of satellite and gauge data – rainfall amount

The rainfall distribution for a pixel *when rainy* is described in terms of the mean μ_+ and variance σ_+^2 as shown below.

- The variance is estimated as an function of μ_*.
- Existing data are used to calibrate a_0, a_1, A, θ



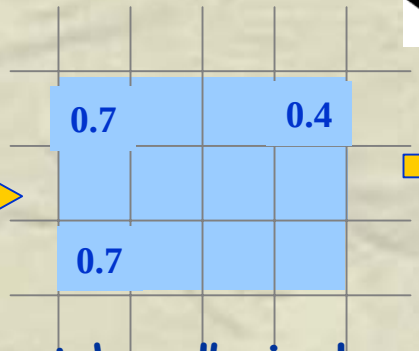
$$\mu_+ \equiv E\left[Z_V \mid z_V > 0, CCD\right] = a_0 + a_1 \times CCD \qquad (2)$$

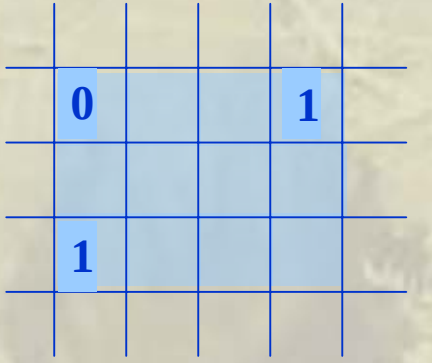
$$\sigma_+^2 \equiv Var(Z_V \mid z_V > 0, CCD) = A\mu^\theta \qquad (3)$$

Step 2: Generate ensemble of rainfall fields - Sequential indicator Simulation

Simulate rain/no rain map from CCD

Initial seeding

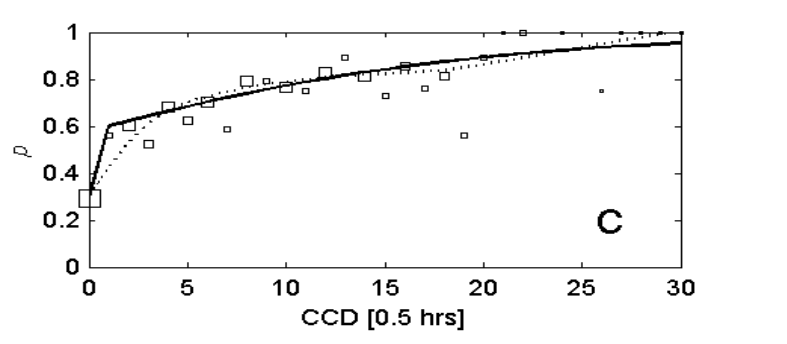




1. Randomly select 'seed' pixels

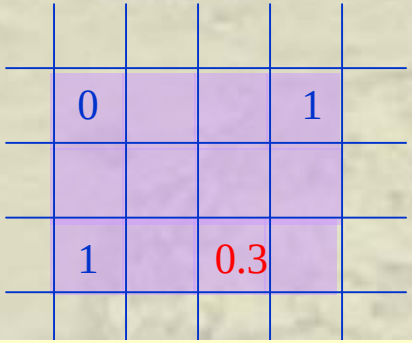
3. Assign 1 or 0 representing rain/no rain based on the prior probability .

2. Use the CCD values to give a prior probability of rain

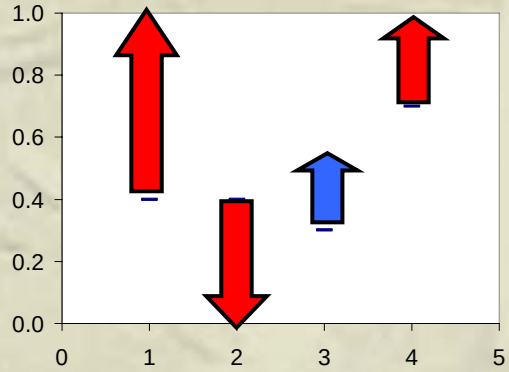


Step 2: Generate ensemble of rainfall fields - Sequential indicator Simulation

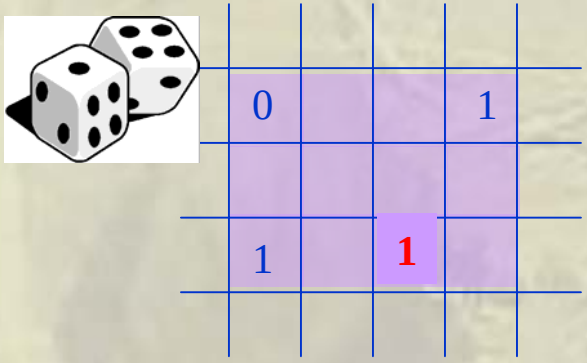
Simulate rain/no rain map from CCD –



1. Select a target pixel, and compute the *prior* probability from CCD..



2. Adjust the prior probability of the target by kriging the residuals of the seed pixels.

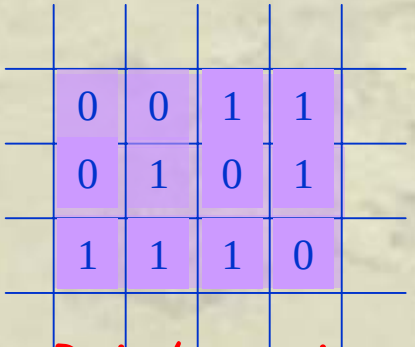


3. Assign 1 or 0 taking into account the final probability

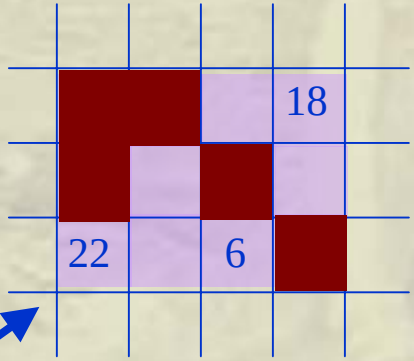
4. The new value is included in the 'sampled' information set and the procedure repeats until all the pixels in the image are covered.

Step 2: Generate ensemble of rainfall fields - Sequential Gaussian Simulation

Simulate rainfall amount - initial seeding

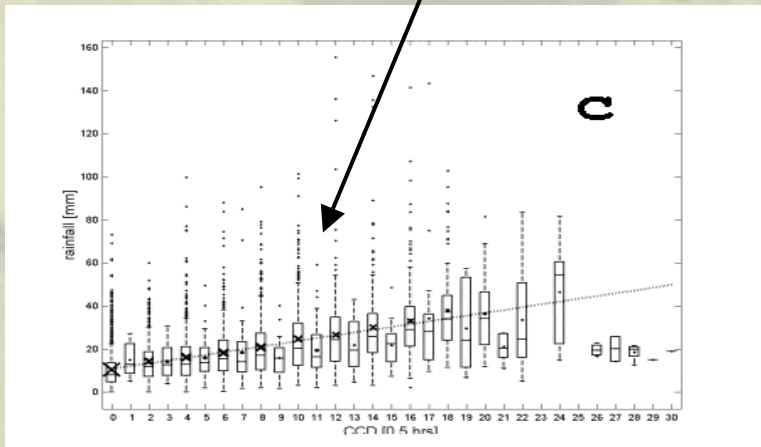


Rain/no-rain map



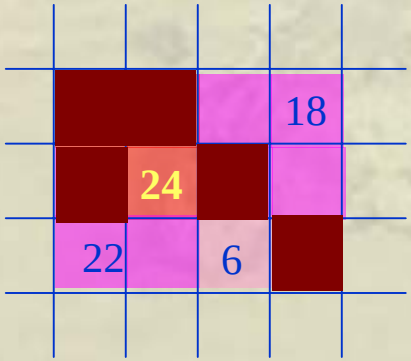
Randomly seed the rainy pixels and select rainfall amounts from calibration distribution

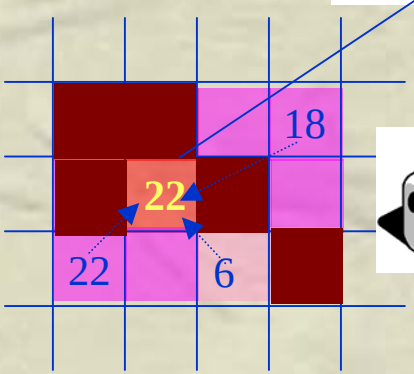
Original CCD values

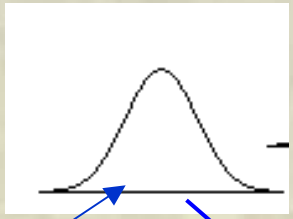


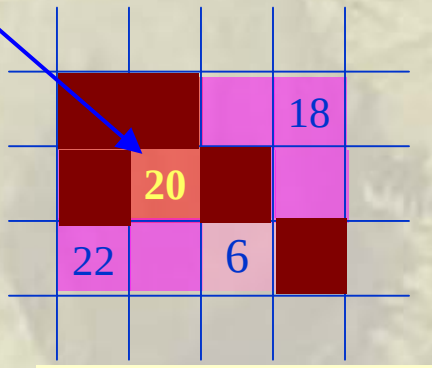
Step 2: Generate ensemble of rainfall fields – Sequential Gaussian Simulation

Simulate rainfall amount – target pixel









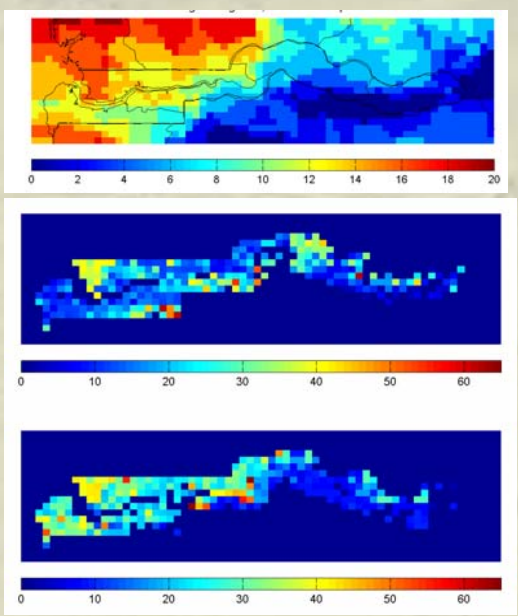
1. Select target pixel and assign value based on calibration equation

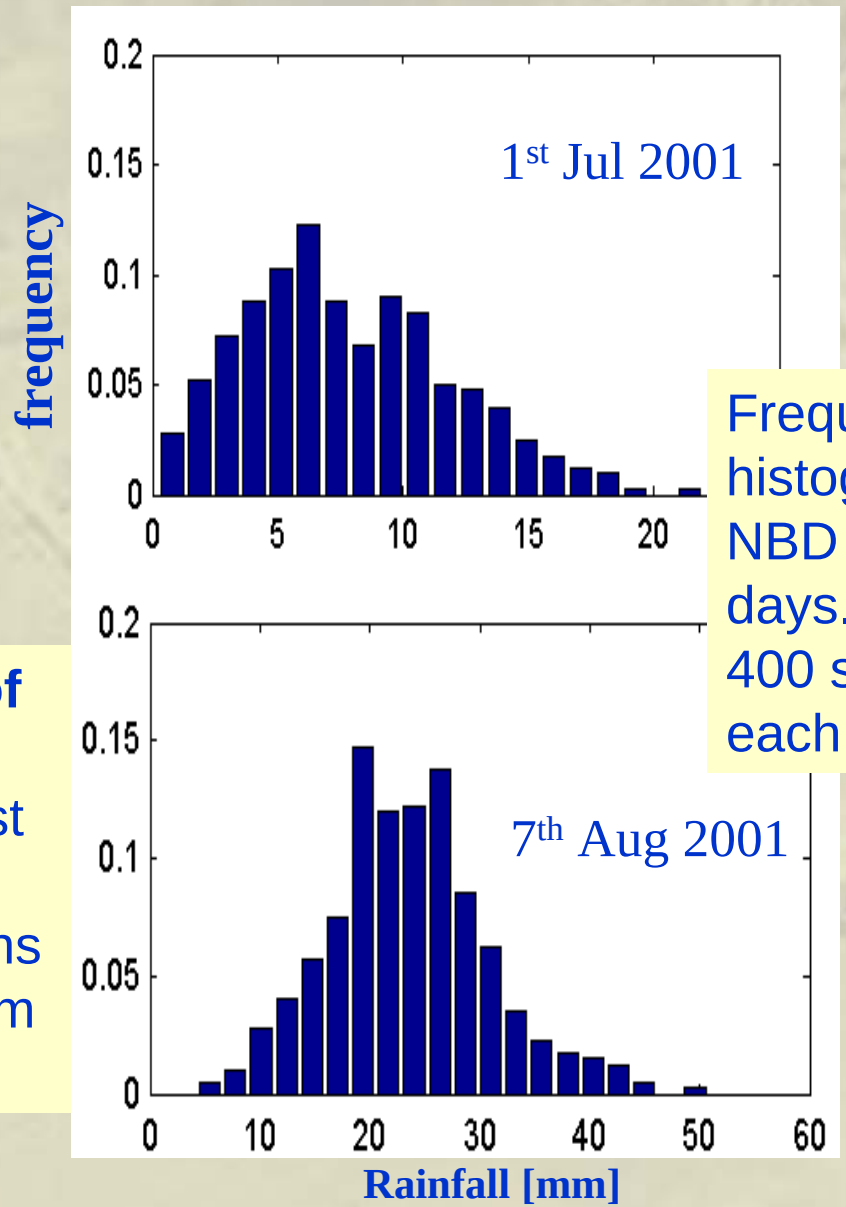
2. Apply residual kriging to previously sampled pixels to update target value and distribution about new value.

3. Select from distribution to give new value

4. Select new target pixel and repeat process

Step 3, 4: Estimation of uncertainty from histogram of area amounts





CCD image and examples of simulated rainfall.
Upper panel: CCD 7th August 2001.
Bottom two panels: realisations of rainfall (mm) generated from the CCD information.

Frequency histogram for NBD for two days.
400 simulations each day

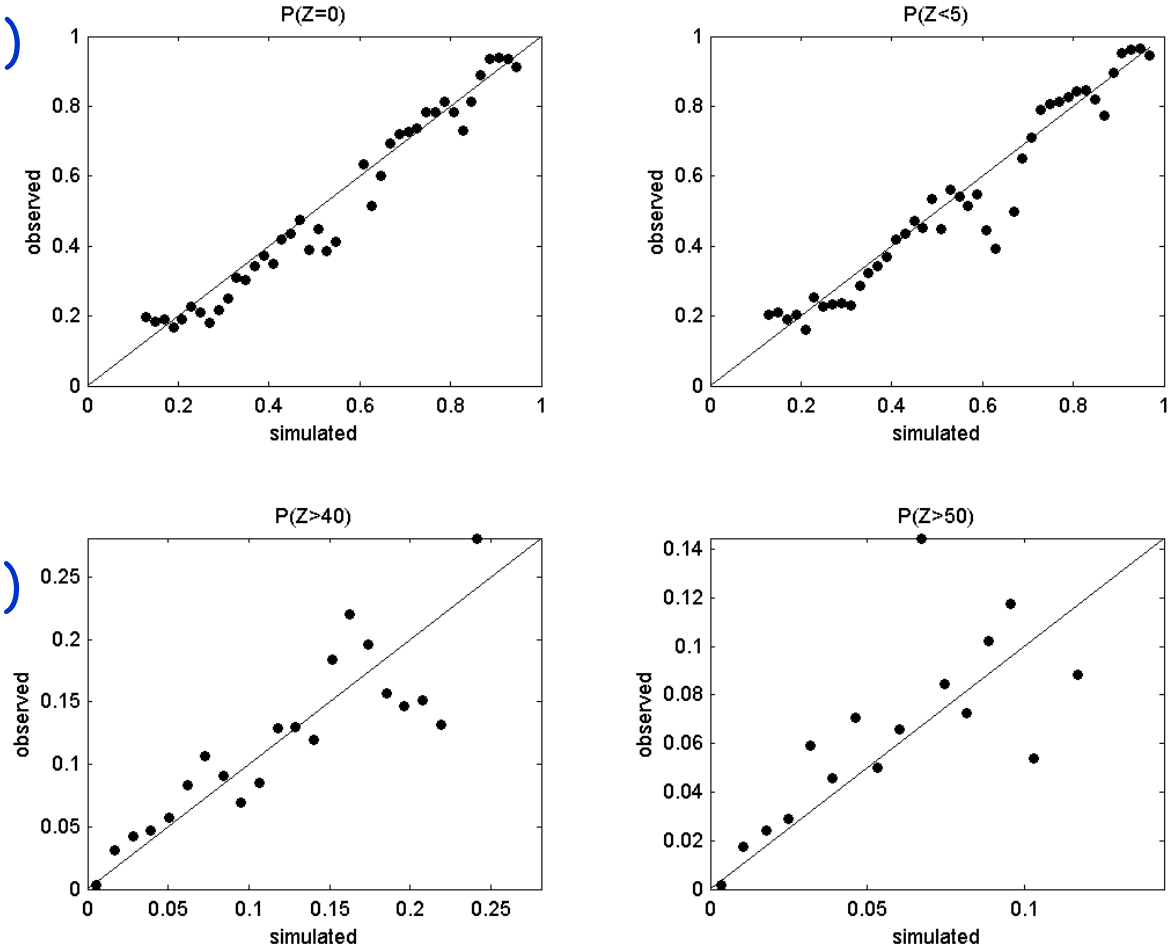
Validation of ensemble statistics

P(rain=0)

P(rain<5)

P(rain>40)

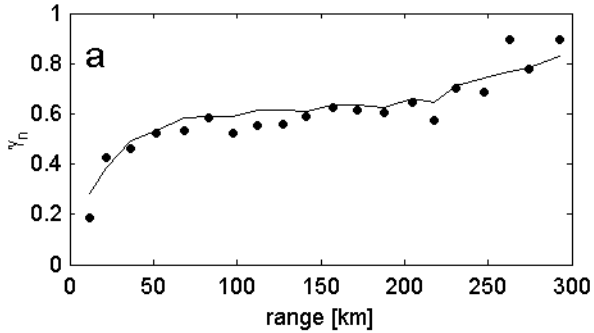
P(rain>50)

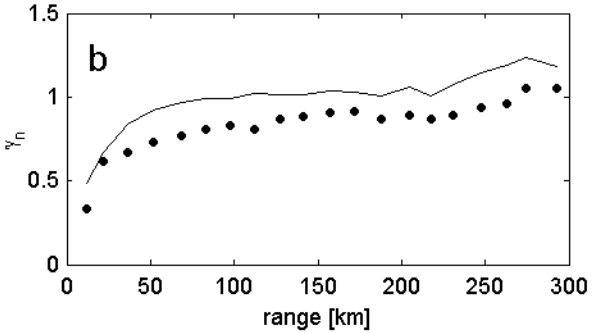


Comparison of ensemble and kriged gauge distributions for various thresholds

Validation of ensemble statistics

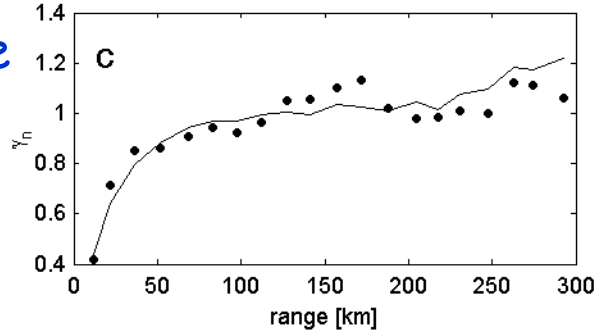
May

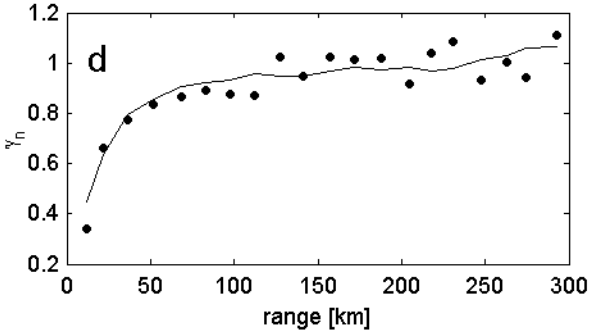




August

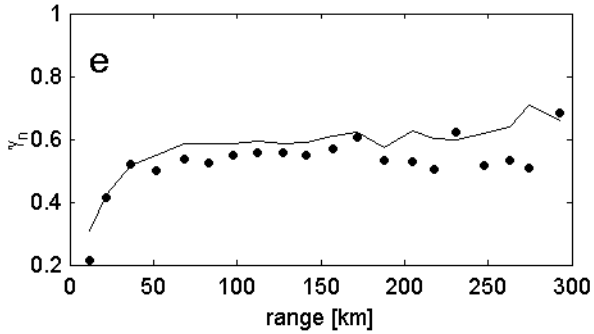
June





September

July

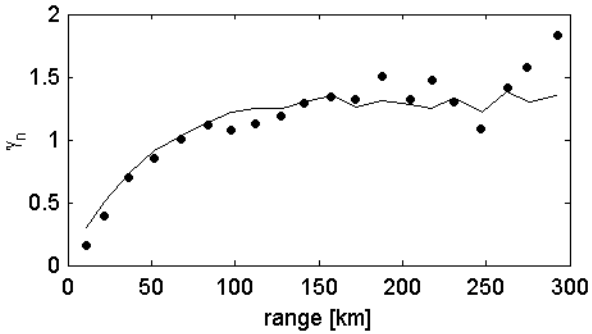


Comparison of rain/no-rain variograms for simulation and kriged gauge data

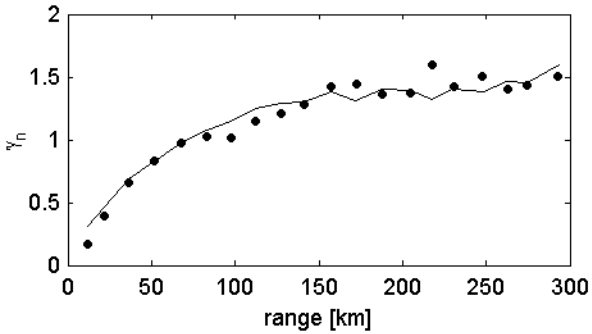
- • = simulation
- ____ = kriged gge data

Validation of ensemble statistics

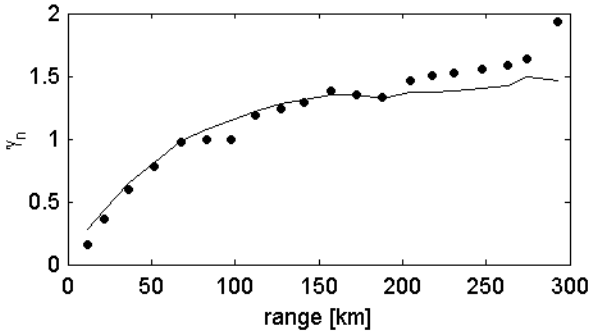
May



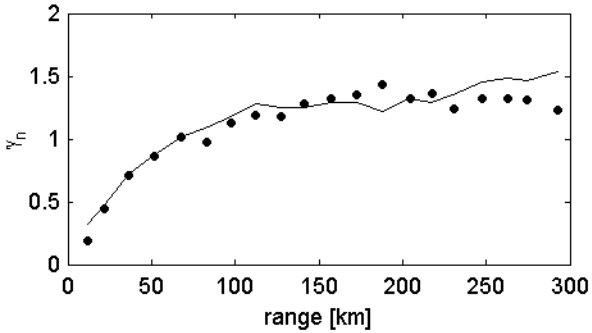
August



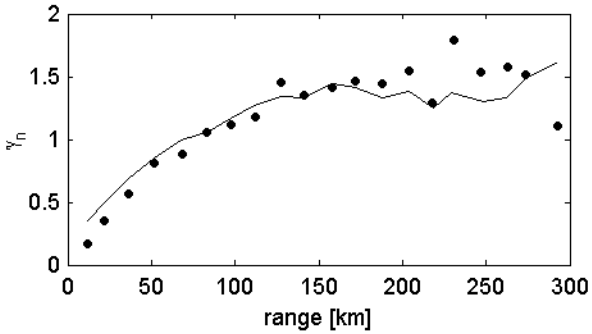
June



September



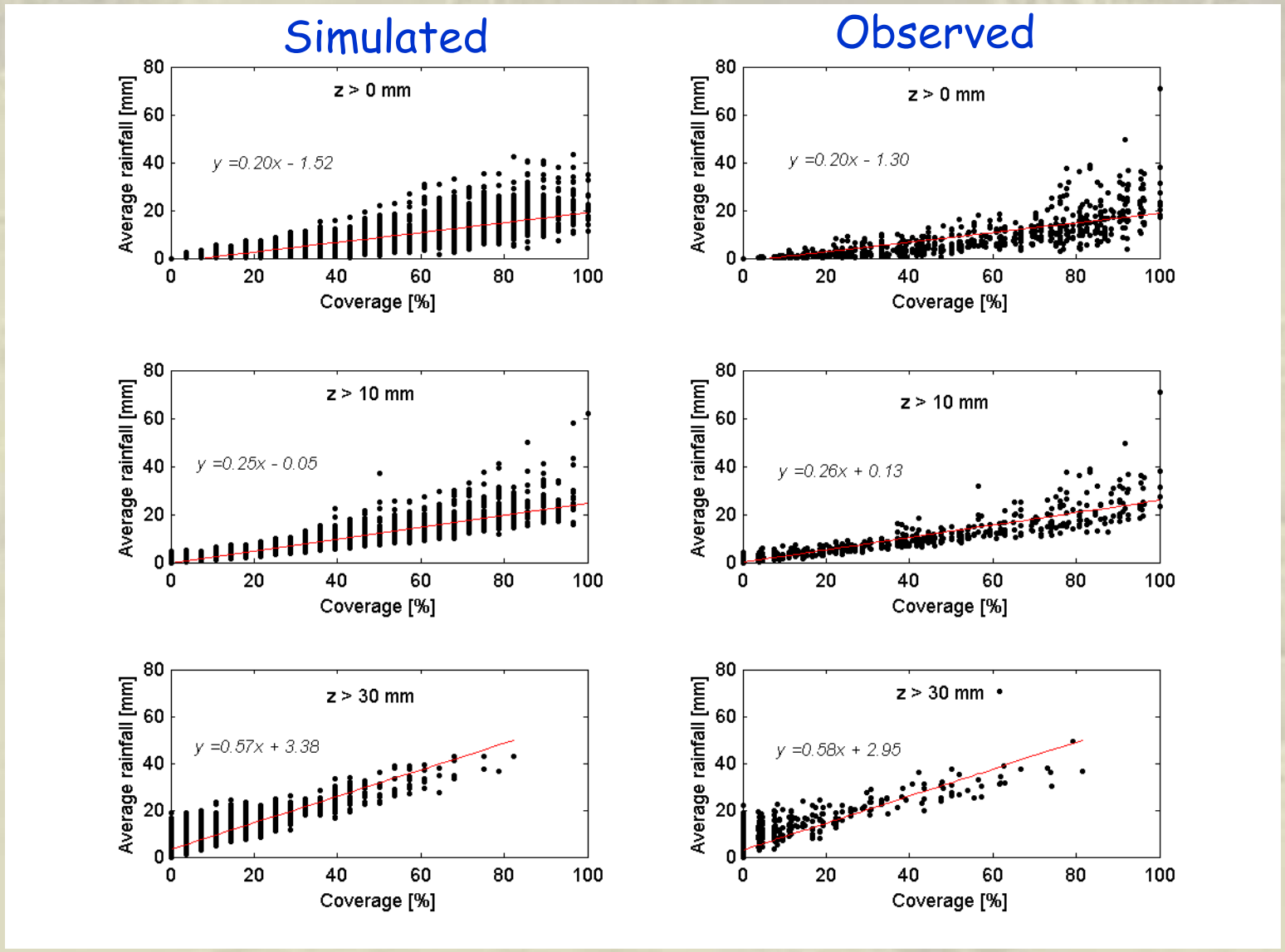
July



Comparison of rainfall amount variograms for simulation and kriged gauge data

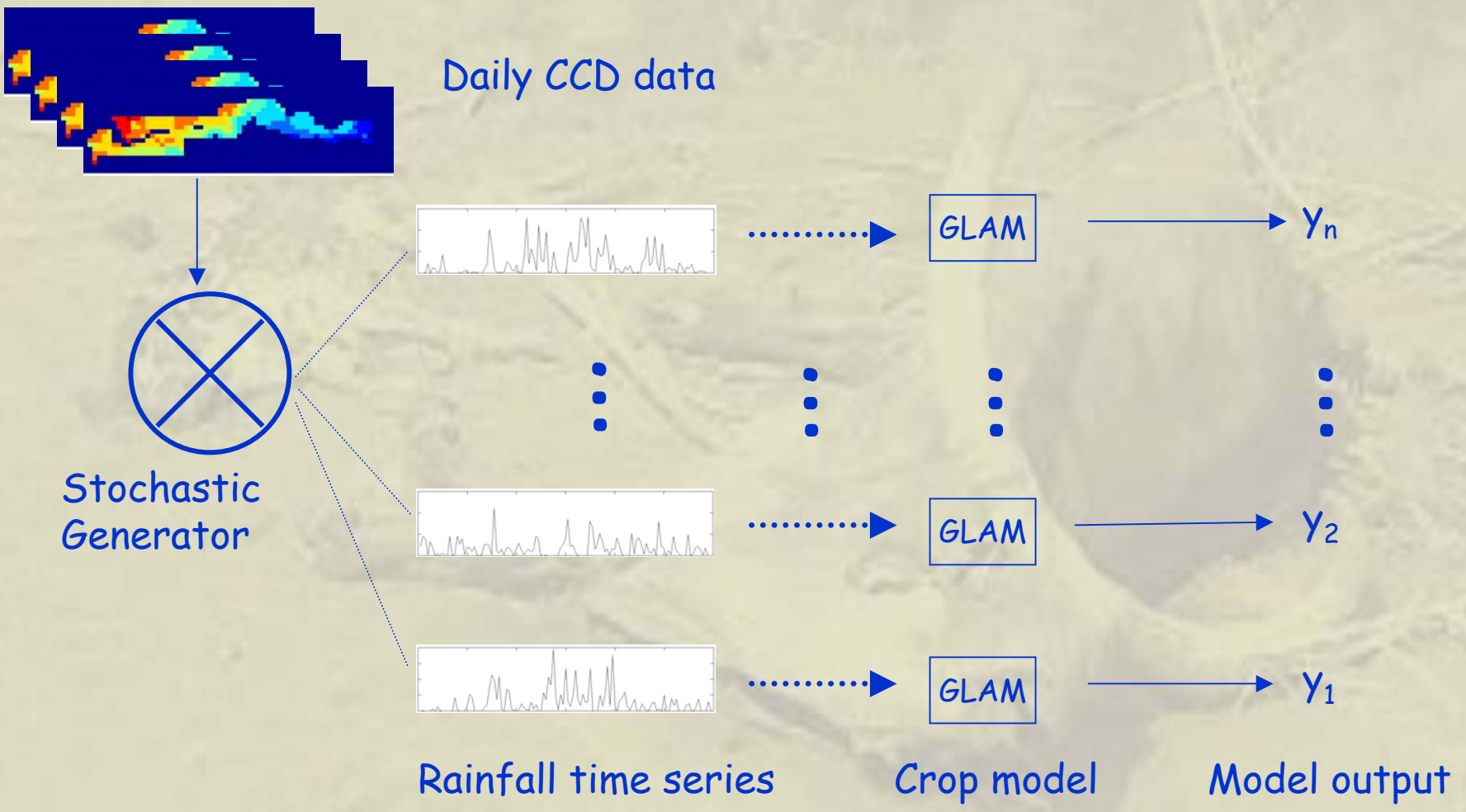
- • = simulation
- ____ = kriged gge data

Validation of ensemble statistics



Areal rainfall – Fractional coverage relationship

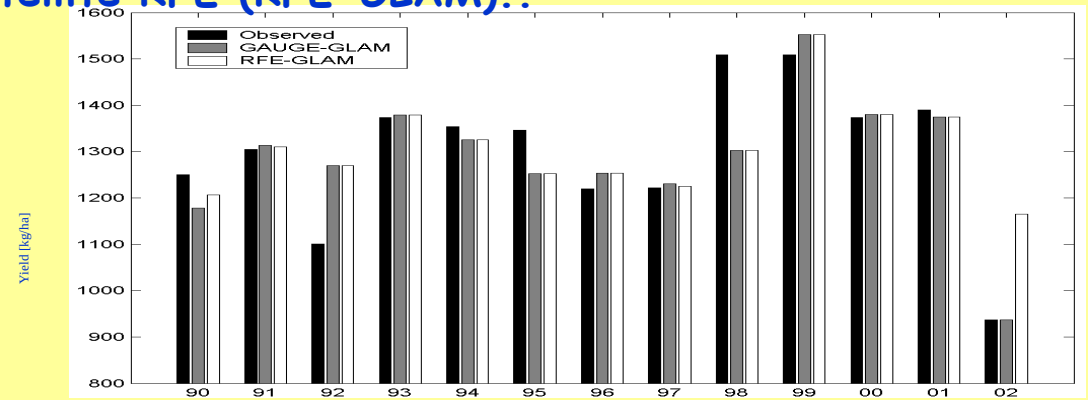
Ensemble method of investigating crop model prediction using GLAM crop model



Groundnuts in Gambia - yield predictions

Case Study: Results from GLAM

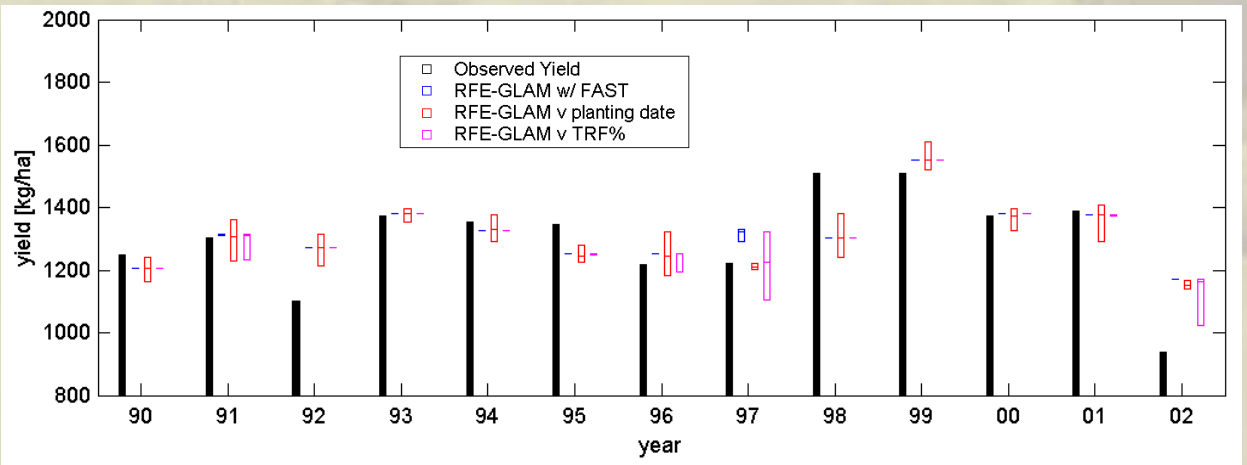
NBD groundnut yields from 1990 to 2002 have been forecast using raingauge data (GAUGE-GLAM) and satellite RFE (RFE-GLAM)..



observed

GLAM gauge

GLAM satellite



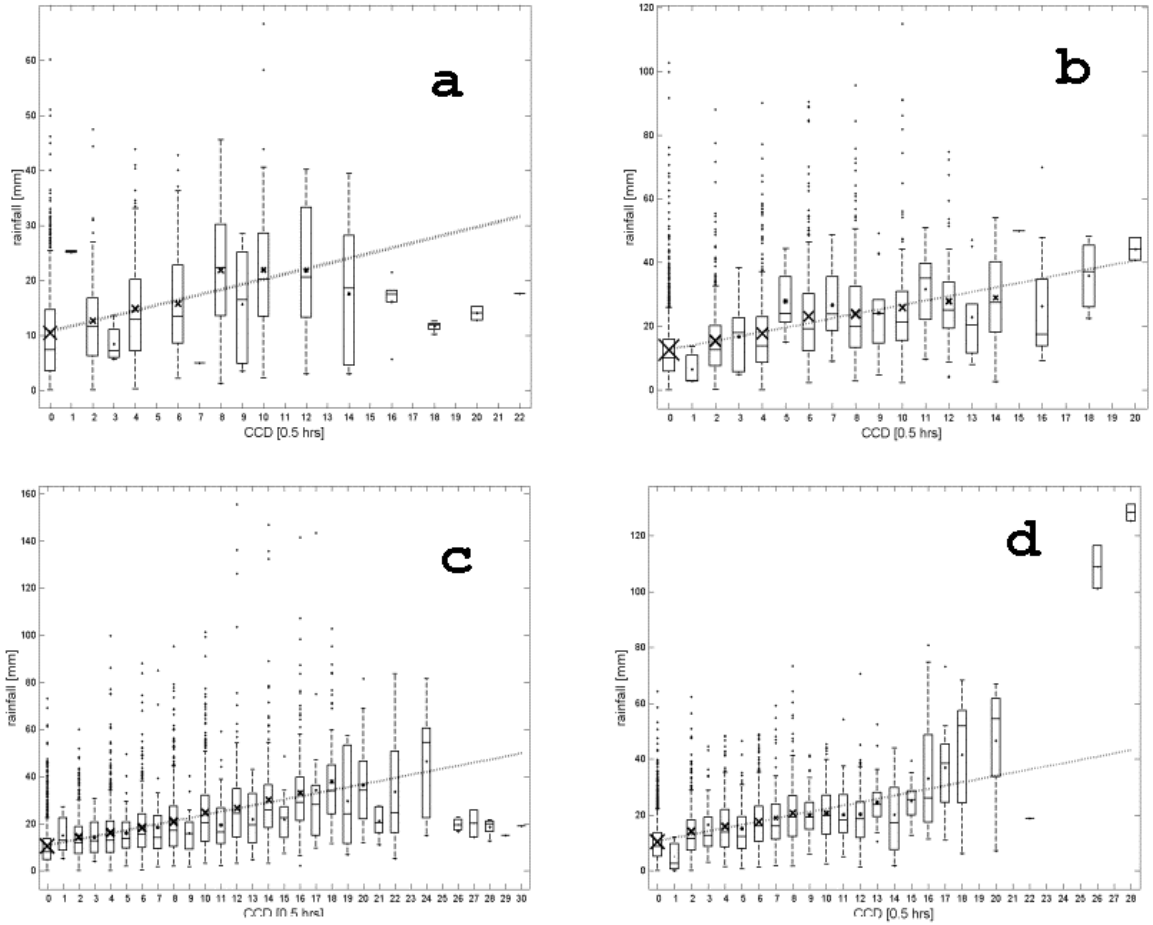
GLAM satellite showing errors

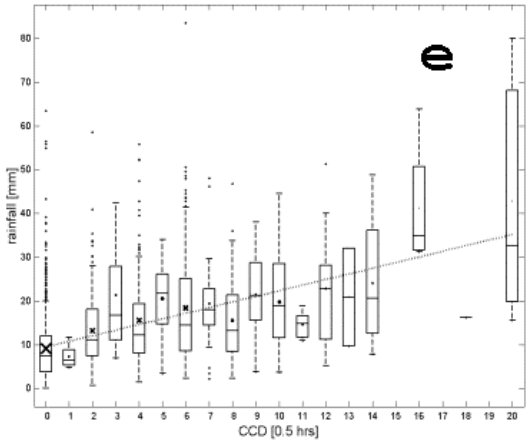
area rainfall

sowing date

Conclusions

- To estimate uncertainty on satellite of area average rainfall, we need to take into account
  - uncertainty on rainfall area
  - uncertainty on rainfall amount
  - spatial correlation pattern for rainfall
- Sequential Simulation allows an ensemble of rainfall estimates to be generated taking into account these factors
- Results using the GLAM crop model show that the model is relatively insensitive to random variations in daily rainfall
- More work needs to be done to take into account temporal correlation of rainfall amount
- This methodology is applicable to any study (e.g hydrological modelling) in which uncertainty on area rainfall amounts is required.

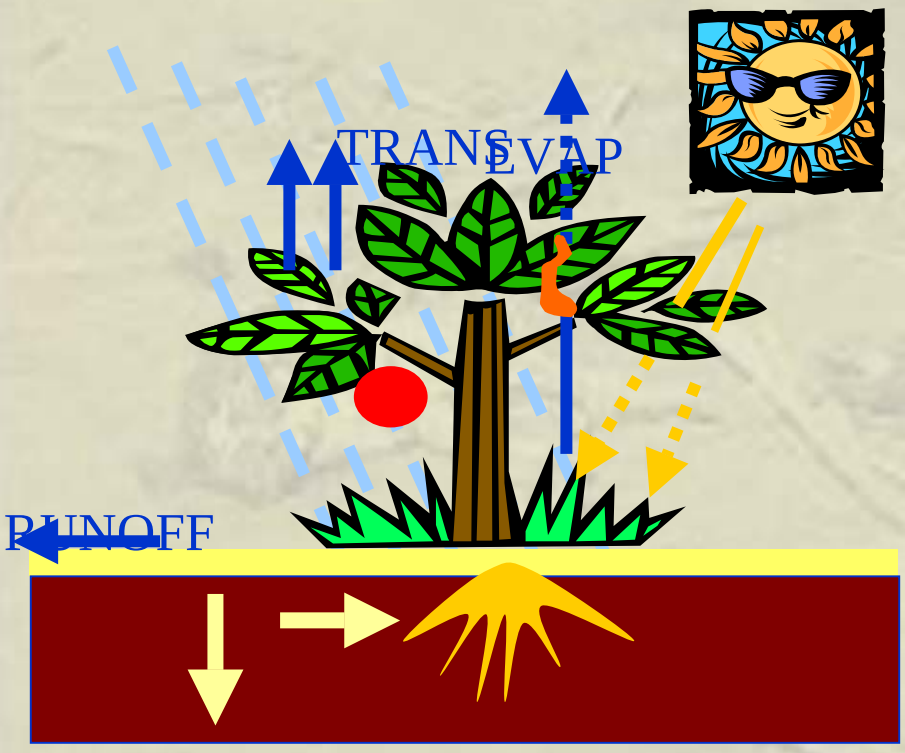




Boxplots of daily pixel rainfall versus CCD and fitted model for various months: '×' – sample mean for each CCD bin. Size of marker is proportional to the number of data available for each CCD bin; dotted line: regression fit. (a) June; (b) July; (c) August; (d) September; (e) October

Calibration for a_0 and a_1 using all data 1988-2002

Factors contributing to uncertainty in crop yield forecast



We need to consider

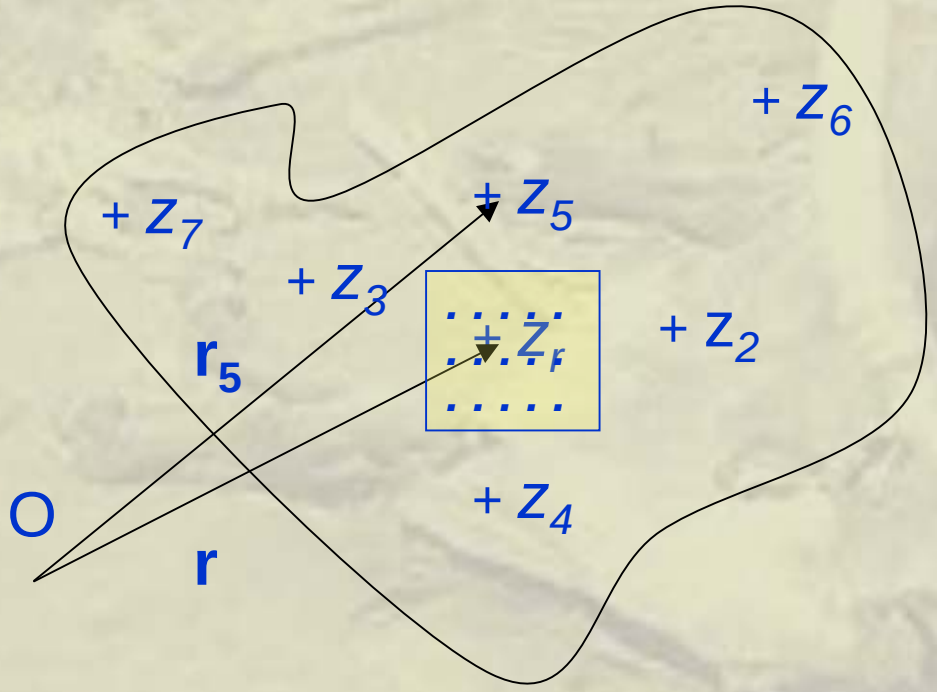
- uncertainty in model parameters
- uncertainty in evaporation estimate
- uncertainty in areal rainfall estimate

This work focuses on the estimation of uncertainty in satellite-based areal rainfall

The methodology is also applicable to hydrological modelling

Step 1: Comparison of satellite and gauge data

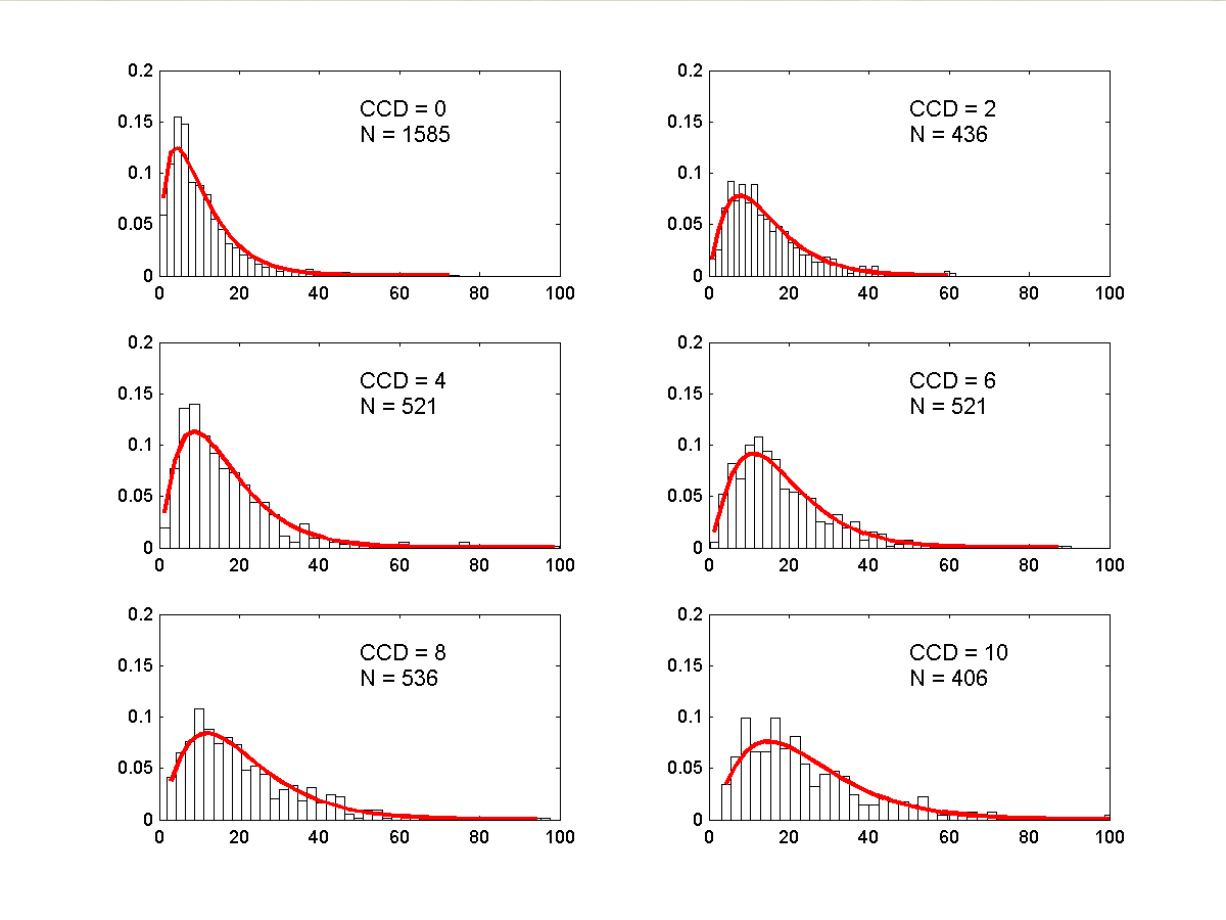
Use **block kriging** to generate pixel-gauge values for each satellite pixel



Step 1: Comparison of satellite and gauge data – rainfall distribution

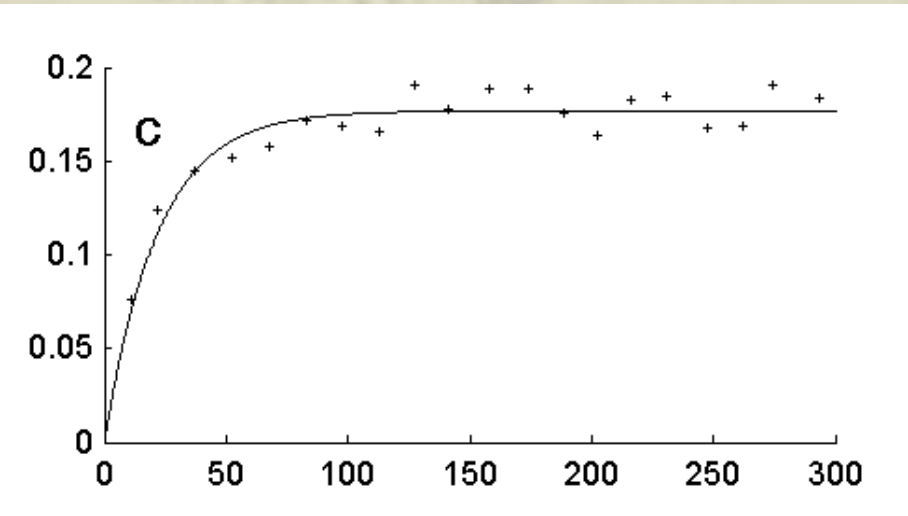
We now have a distribution of rainfall amounts for each CCD value.

We assume a gamma distribution



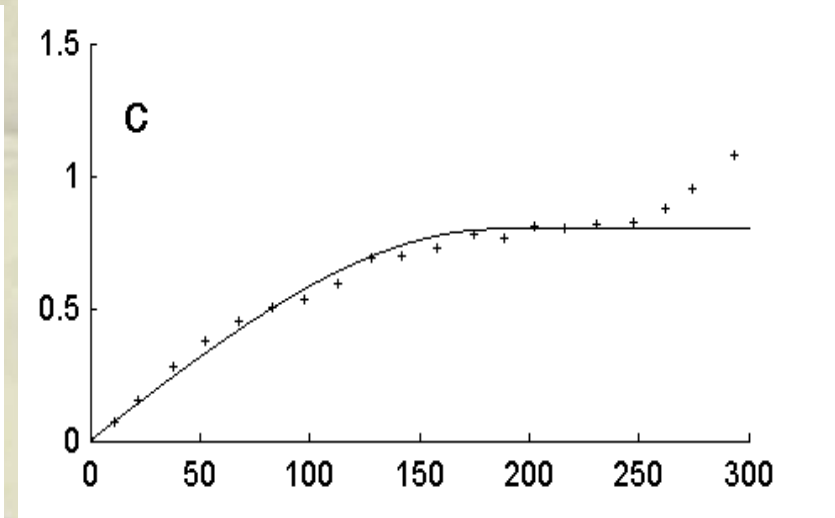
Step 2: Determine spatial correlation of rainfall field

Generate variograms for rainy/non rainy pixels and rainfall amount within rainy area



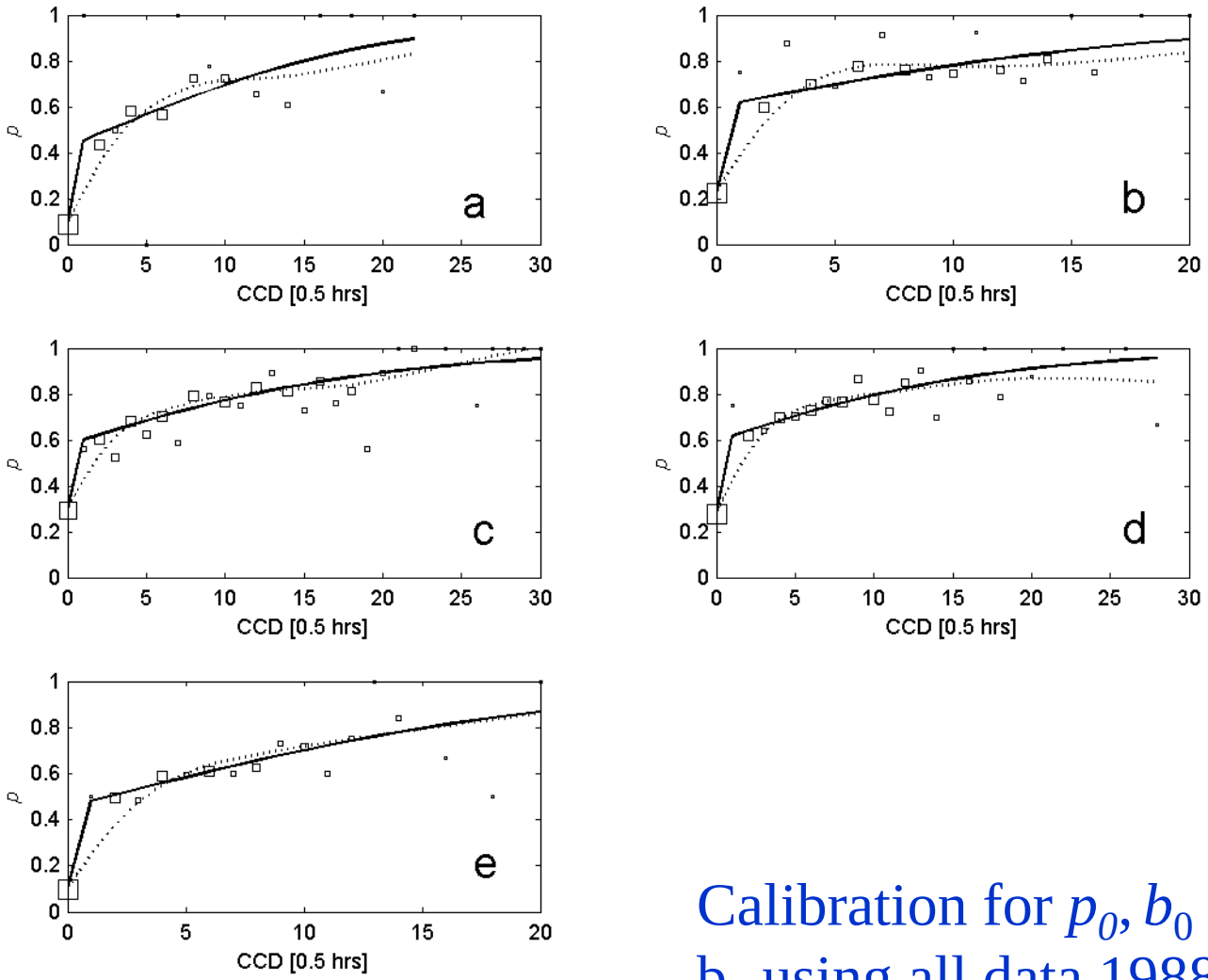
Pixel separation/ km

Indicator variogram (raininess)



Pixel separation/ km

Rainfall amount variogram



Calibration for p_0, b_0 and b_1 using all data 1988-2002

Scatter plots of the observed rainfall occurrence frequencies as well as the model of rainfall probability as a function of CCD. Solid line: Logistic Model; Dotted line: Weighted spline fit with smoothing factor of 0.8

Probabilistic rainfall estimation summary

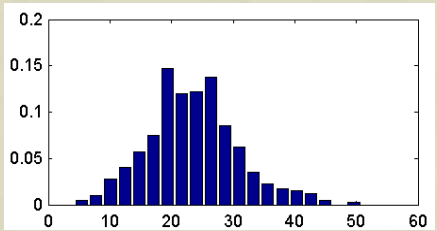
Daily CCD data

Initial Seeding using prior probability of rain

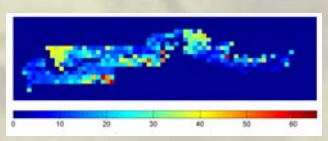
Determine rain/no rain map by Sequential Indicator Simulation

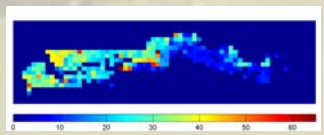
Initial seeding of positive rainfall amount on rainy area using prior positive rain distribution

Determine rainfall amount for all rainy pixels by Sequential Gaussian Simulation..



Σ





Aggregating pixel rainfall for each realisation to give a sample of division rainfall

Multiple realisations of rainfall map

Note:

Variogram shown are not normalised in terms of the sill value.

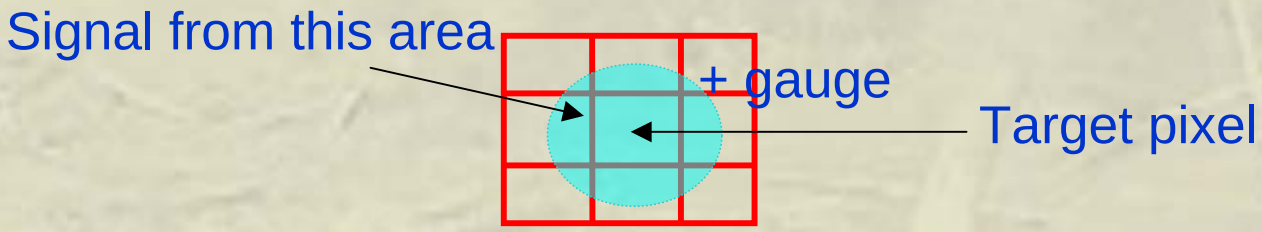
Variogram for positive rainfall is constructed from LU decomposition.

Indicator variogram worked out from the block indicator estimates.

All variogram model uses the exponential model.

Strictly speaking residual variogram is needed for simulation. For processes with constant mean or at least constant in mean for a local neighbourhood, the residual variogram is identical to that of the variable variogram. Since both the estimates for the block indicator, I_v as well as the LUD assumes at least locally stationarity in mean, variogram of the residuals $R_I = I_v - E(I_v)$ and $R_Y = Y_v - E(Y_v)$ will be the same as I_v and Y_v at least locally.

The problem of gauge-satellite comparison



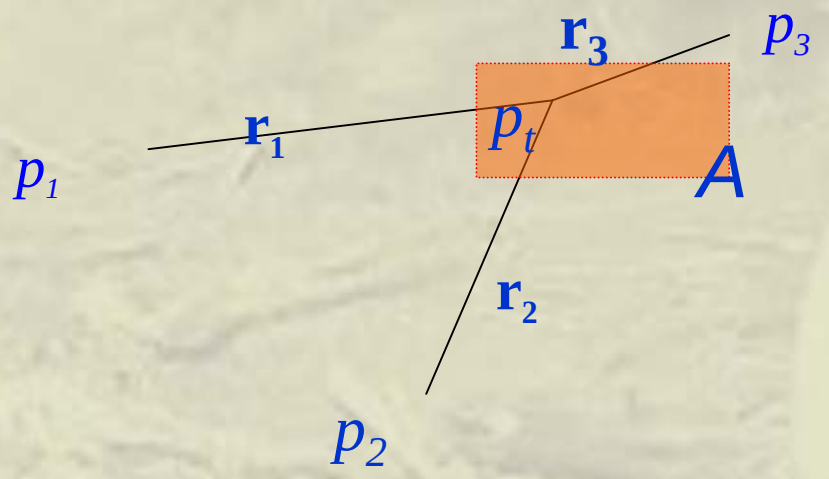
Problem

Gauge gives point value but satellite gives area average

Question

How do we compare at same spatial scale?

Block kriging - pixel averages from gauge data

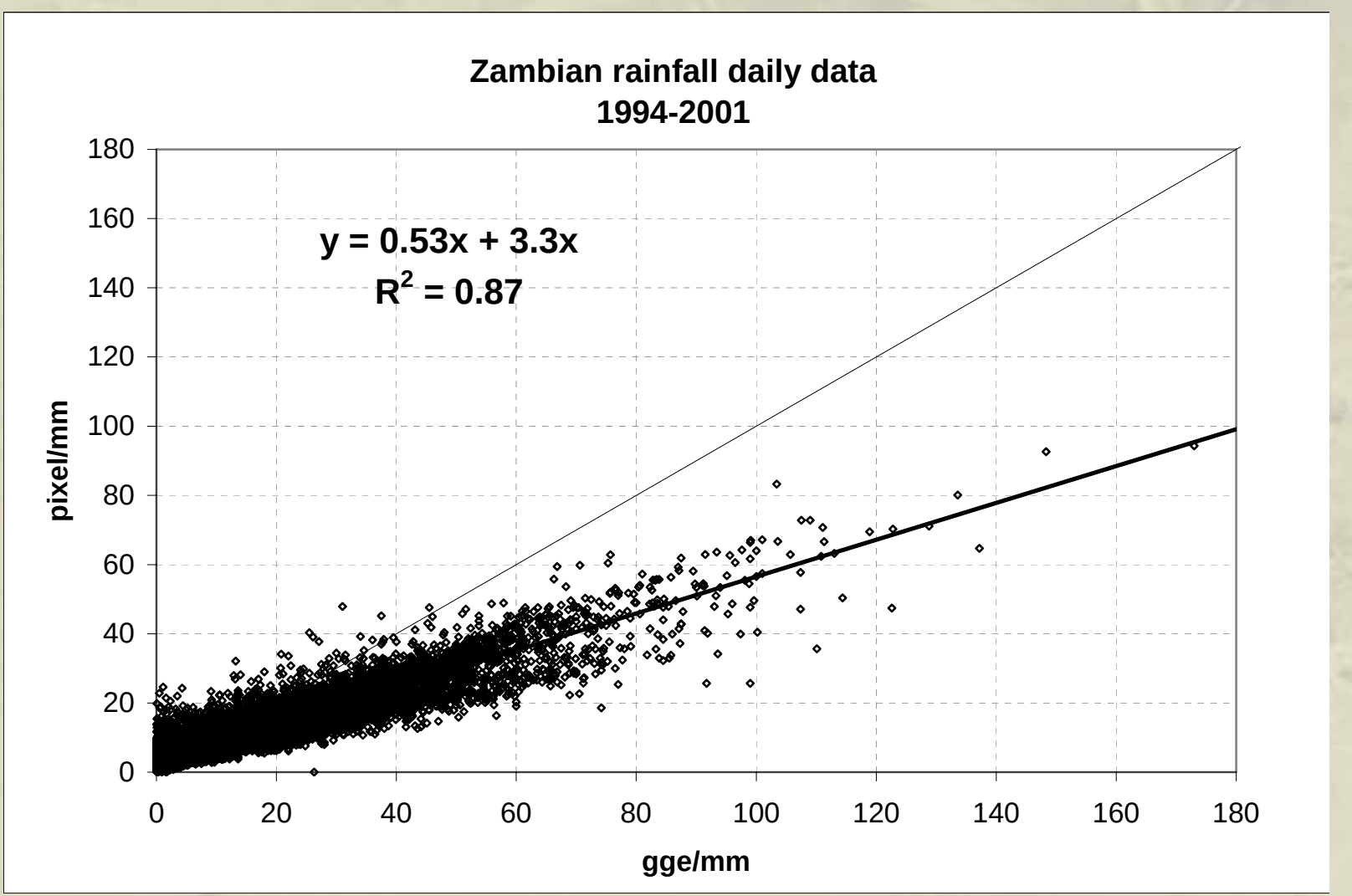


$$p_t = \sum_{i=1}^{n} \lambda_i p_i ;$$

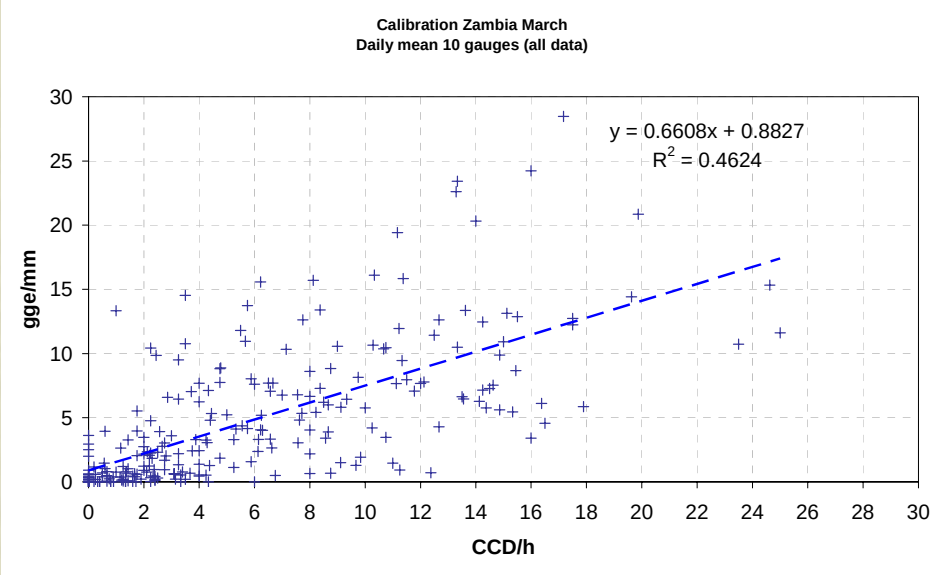
$$\sum_{i=1}^{n} \lambda_i = 1$$

- Rainfall at point *t* is weighted sum of surrounding gauges
- Weights λ_i depend on correlation between rainfall measurements at distance r_i
- Kriging automatically provides an error estimate for each value
- For kriged areal estimate (e.g. for pixel area *A*), simply average over all p_t in area *A.*

Comparison of pixel average v. raw gauge data for Zambia

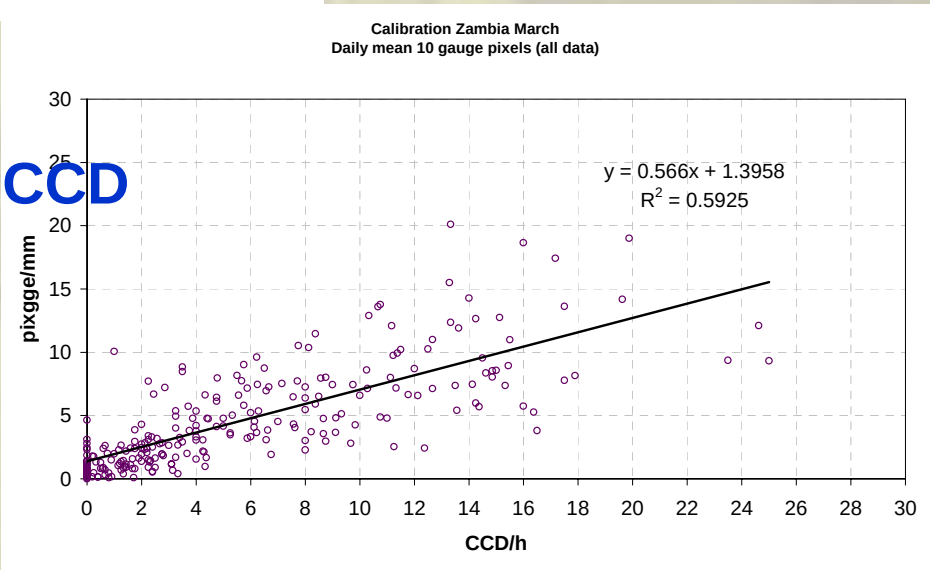


Daily calibrations using pixel and gauge data for Zambia, March (10 gauges)



mean gauge v CCD

mean kriged pixel v CCD

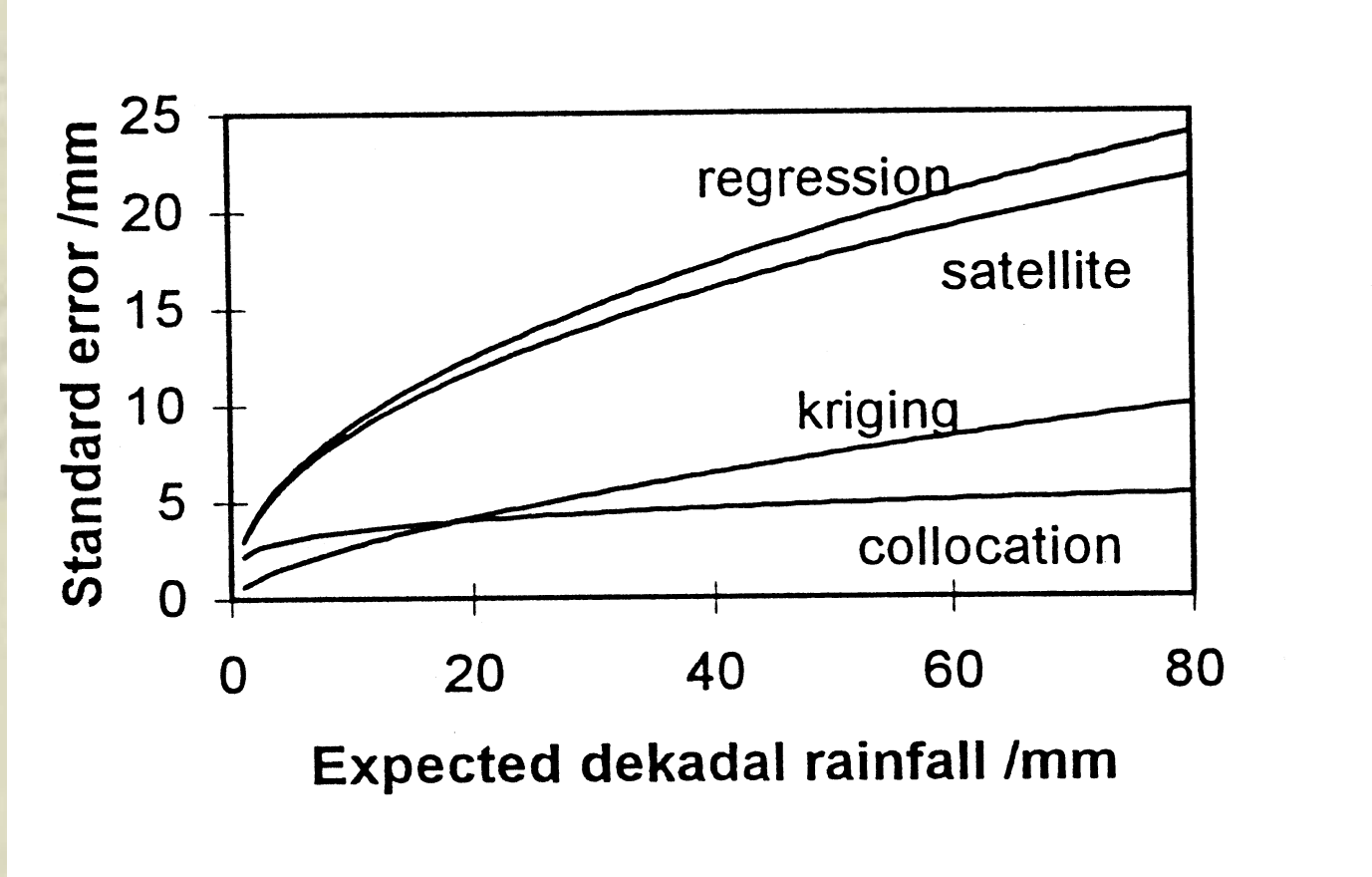


Merging gauge and satellite data

1. Calculate kriged estimates P_k and errors e_k for each pixel using gauge data
2. Estimate errors e_s for satellite value P_s for each pixel
3. Calculate best (merged) estimate P_m as

$$P_m = \frac{e_k^2}{\left(e_k^2 + e_s^2\right)} P_s + \frac{e_s^2}{\left(e_k^2 + e_s^2\right)} P_k$$

4. The error e_m on this estimate is given by

$$e_m^2 = \frac{e_s^2 e_k^2}{e_s^2 + e_k^2}$$

Estimation of satellite error