Department of Geography and Earth Sciences

WRM 625

Hydrological Modeling: Calibration, Verification, Validation, and Sensitivity Analysis

Lecture 3

PHASES OF MODEL EVALUATION

- Levels of model evaluation

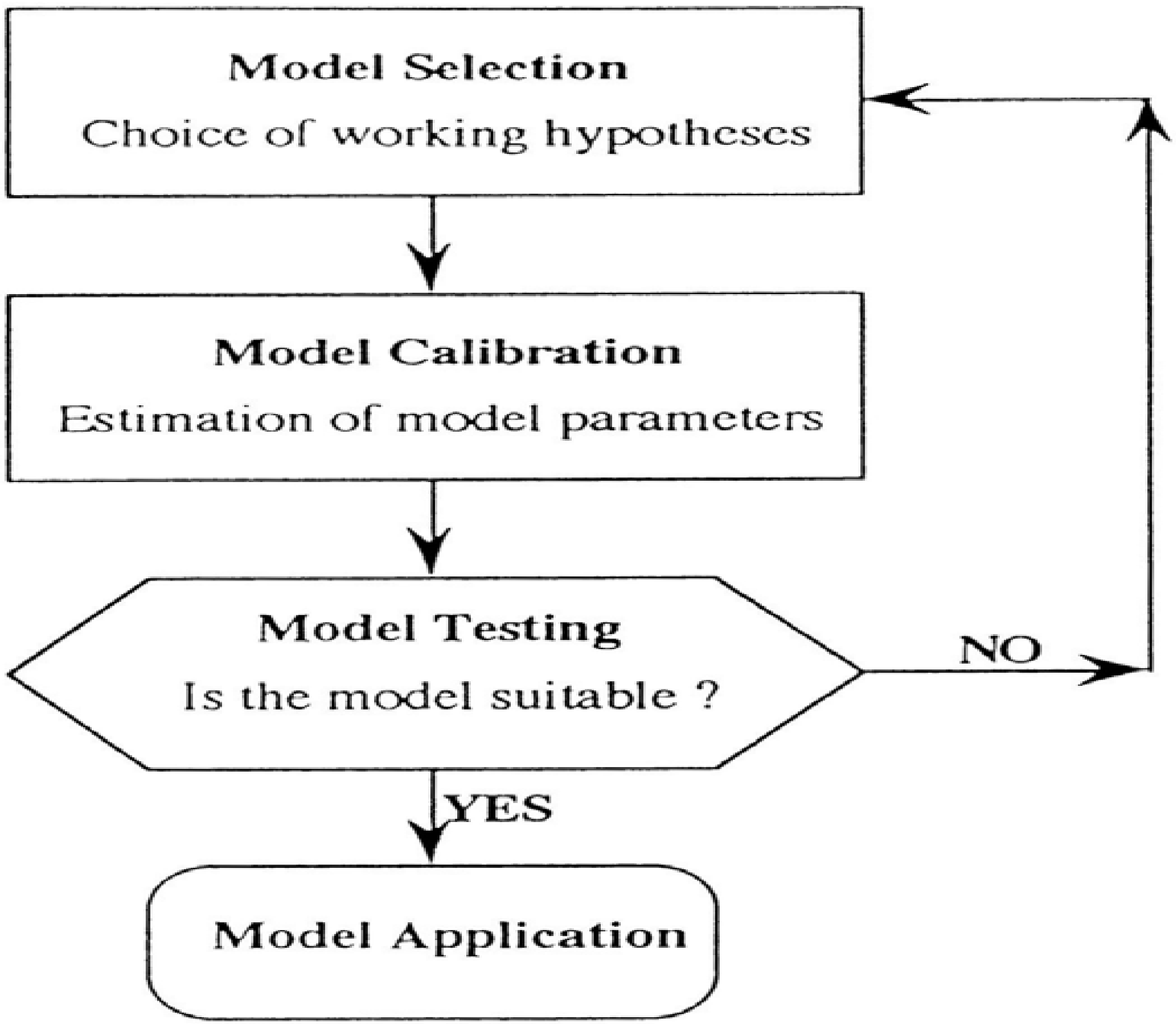
  (i) model selection – choice of working hypotheses

  (ii) model calibration - estimation of the parameter values

  (iii) model validation - testing the fitted model to verify its accuracy; and

  (iv) estimation of its range of applicability

Phases of Model Evaluation



Calibration

- All models have unknown constants used to represent the physical process.
- These constants are called model parameters and must be assigned fixed numerical values before the model may be used to predict the runoff .
- There must be agreement between Observed and simulated flow
- The process by which the parameters are selected is called model "calibration".

Model Parameters

a) **Physical parameters:** represent physically measurable properties of the watershed.

– Examples the area of the watershed, the fraction of the watershed area that is impervious, the surface area of the streams and open water bodies, surface slopes.

(b) Process parameters: process parameters represent watershed properties that are not directly measurable.

Examples : the average or "effective" depth of surface soil moisture storage, the effective lateral interflow rate, the coefficient of nonlinearity controlling rate of percolation to the groundwater storage

Parameter determination

- **Parameter specification:** prior knowledge about the watershed properties and behaviour to specify initial estimates for the parameters of the model
  - Physical parameters: from field measurements and fixed
  - Process parameters: estimates of the range determined based on knowledge of the watershed and adjusted.

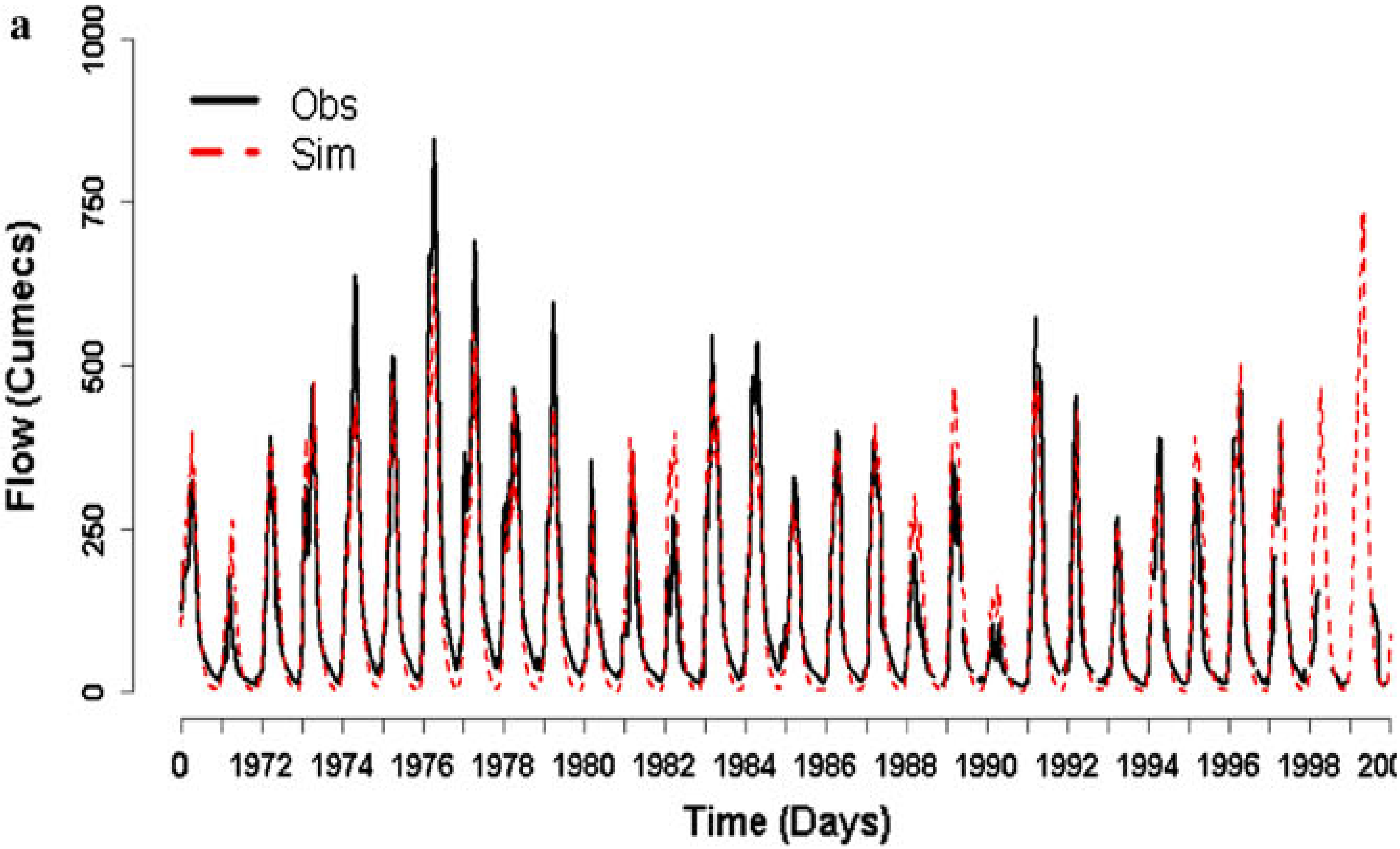
Parameter determination

- **Parameter estimation:** Set initial range and adjust parameters manually or automatically

Calibration: Checking Goodness of Fit

- Visual comparison between simulated and observed data – look for trends in errors
  - A learned art, subjective
  - Use appropriate graph
- Statistical and regression based performance measures
  - Consider mean daily discharge as calibration target
  - Q = observed
  - S = simulated

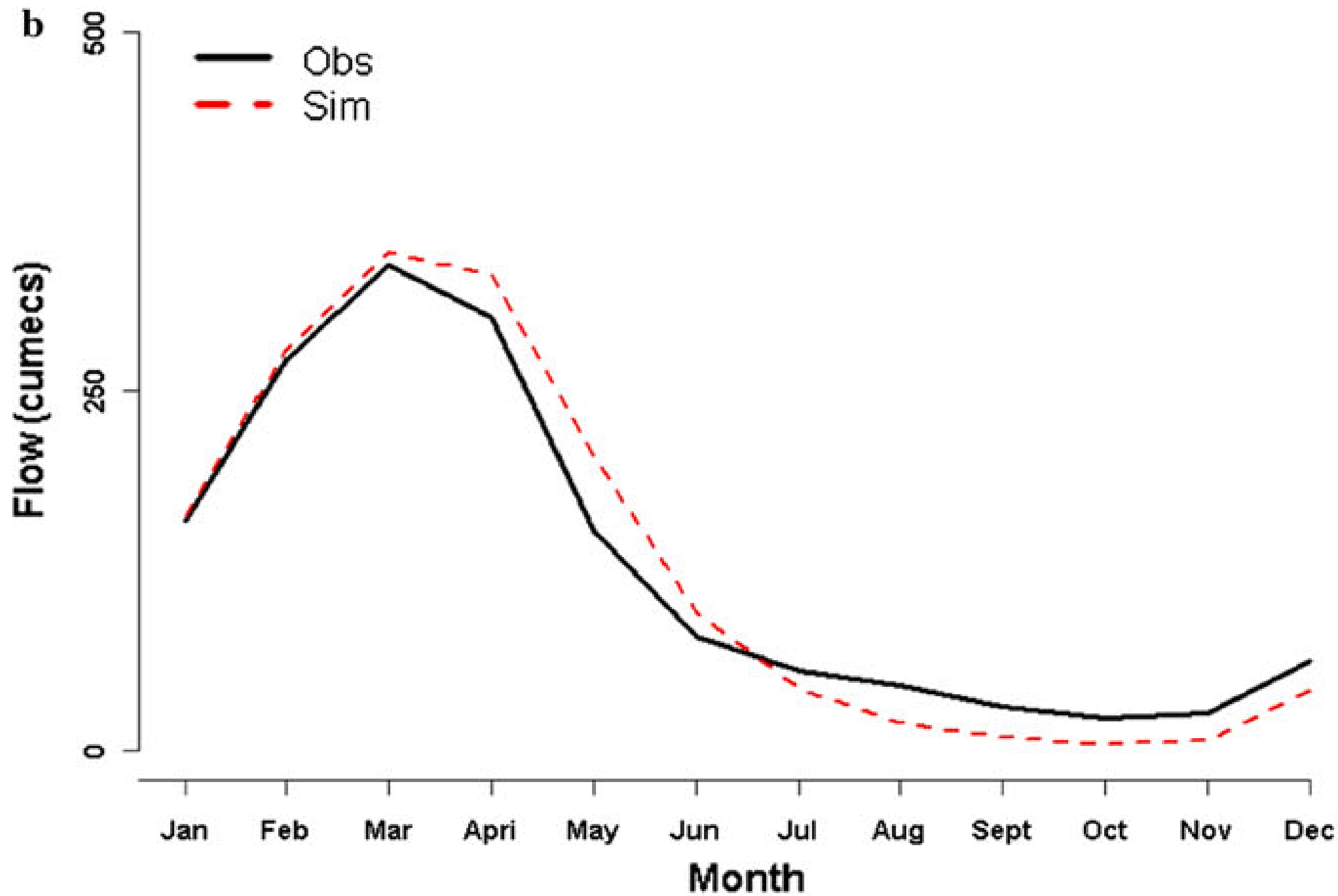
Calibration and Validation



Split Sample Test for the WASMOD-D in the Upper Kafue River Basin: Calibration 1971–1986) and validation (1987–2001)

Source: Ngongondo et al. 2013, Stoch Environ Res Risk Assess

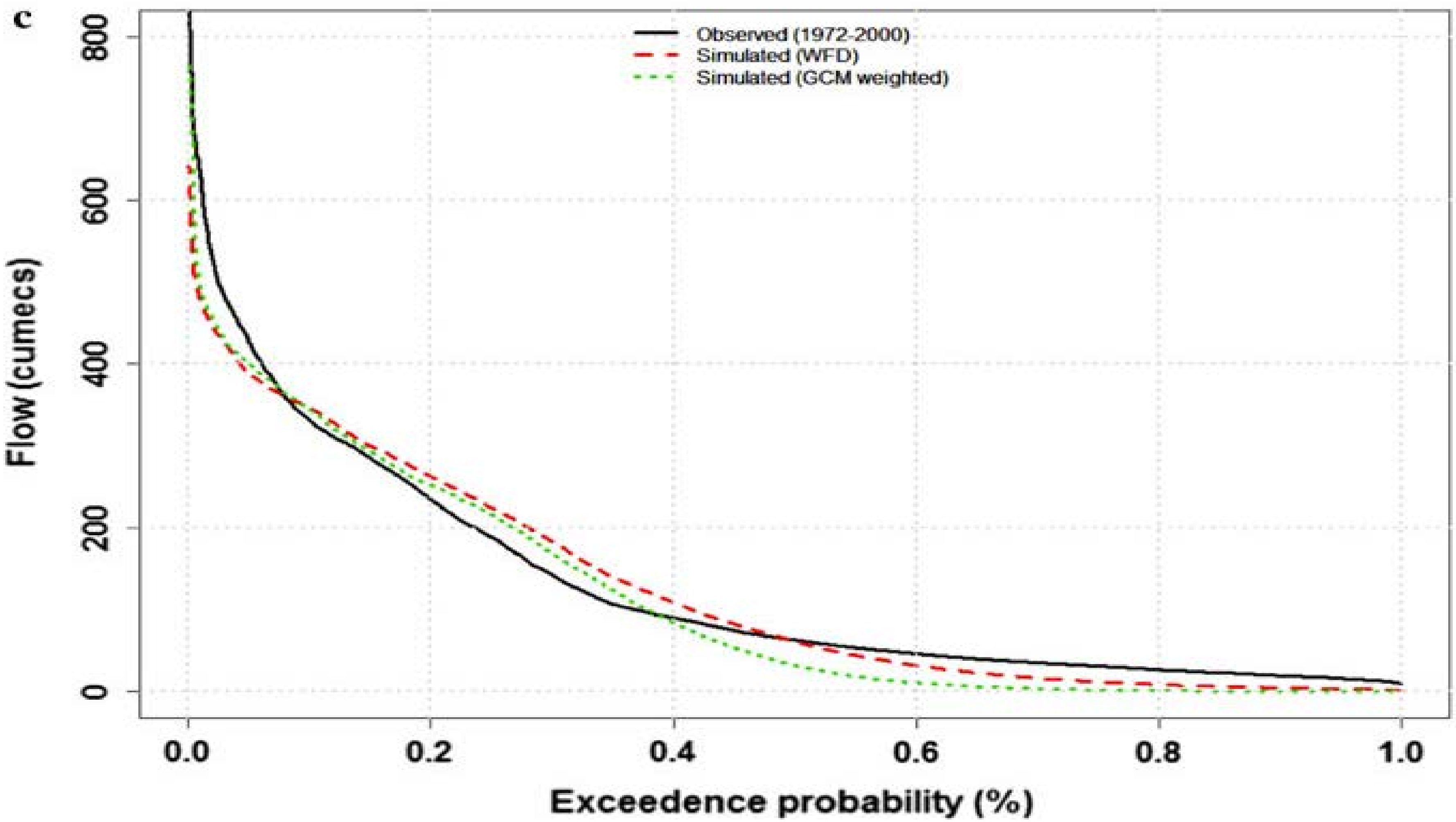
Calibration and Validation



Split Sample Test for the WASMOD-D in the Upper Kafue River Basin: Calibration 1971–1986) and validation (1987–2001)

Source: Ngongondo et al. 2013, Stoch Environ Res Risk Assess

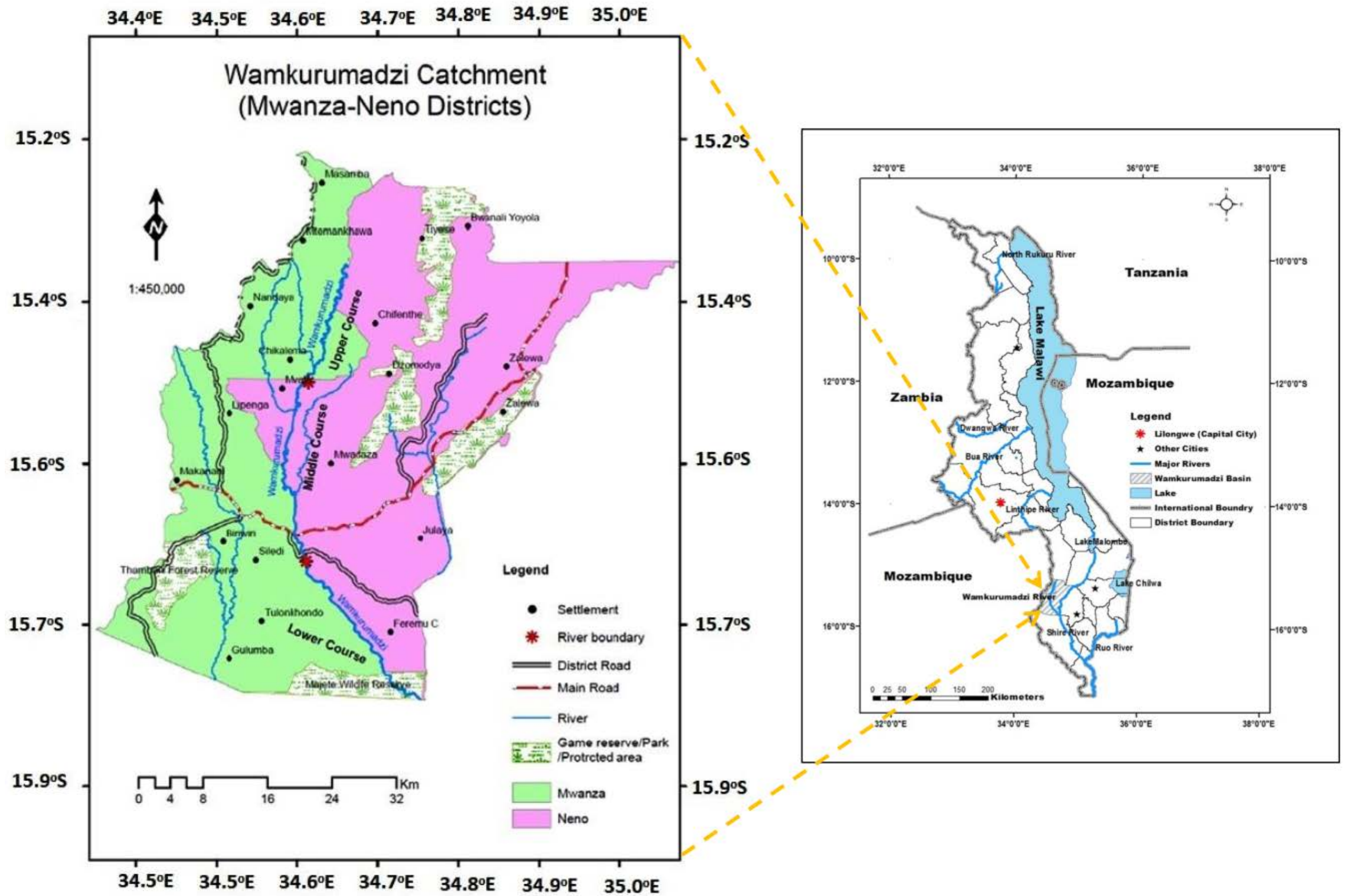
Calibration and Validation



DC for observed, WFD simulated and GCM weighted discharge for the upper Kafue River at Machiya Ferry (1972–2000)

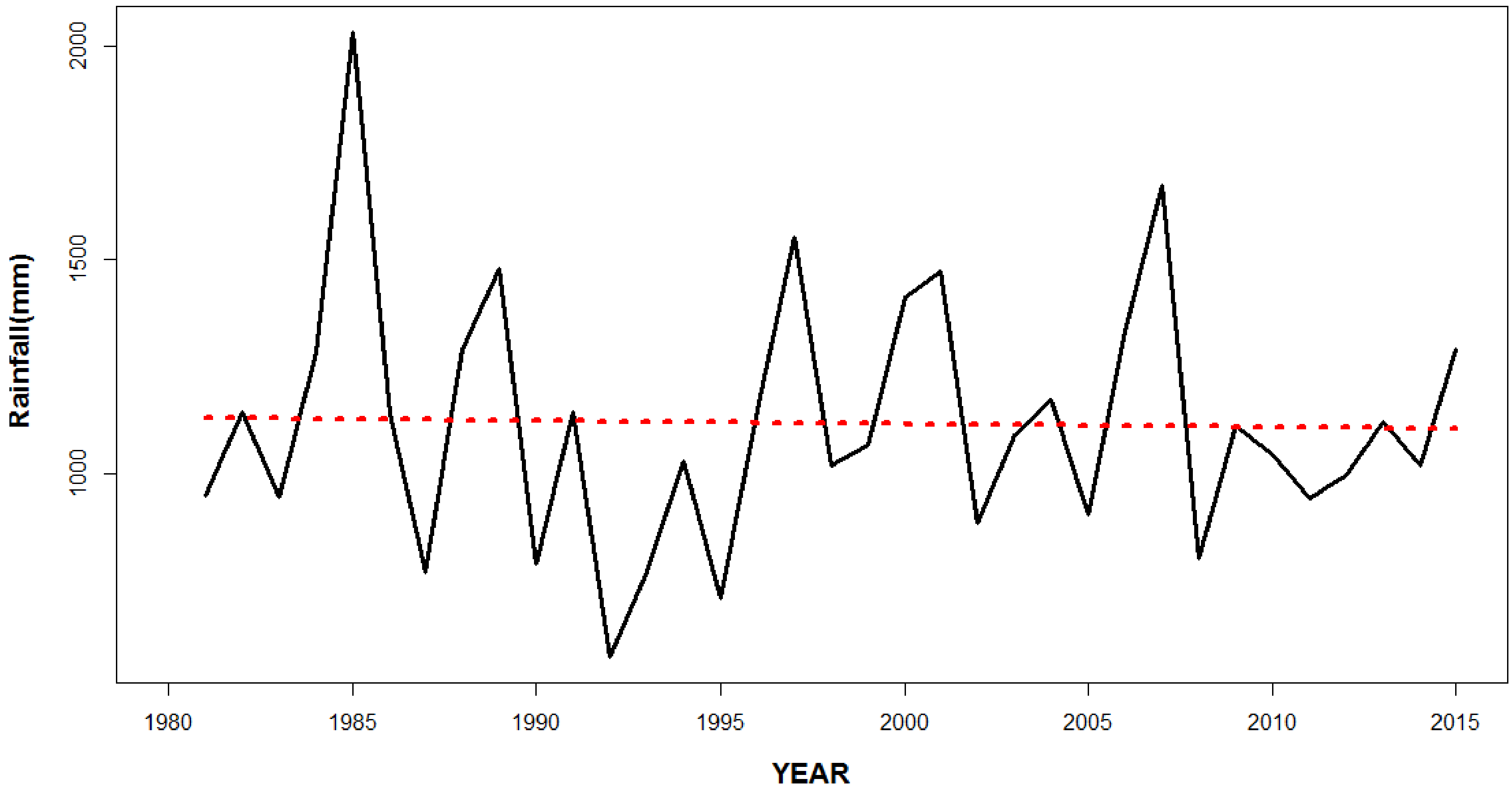
Source: Ngongondo et al. 2013, Stoch Environ Res Risk Assess

Calibration and Validation



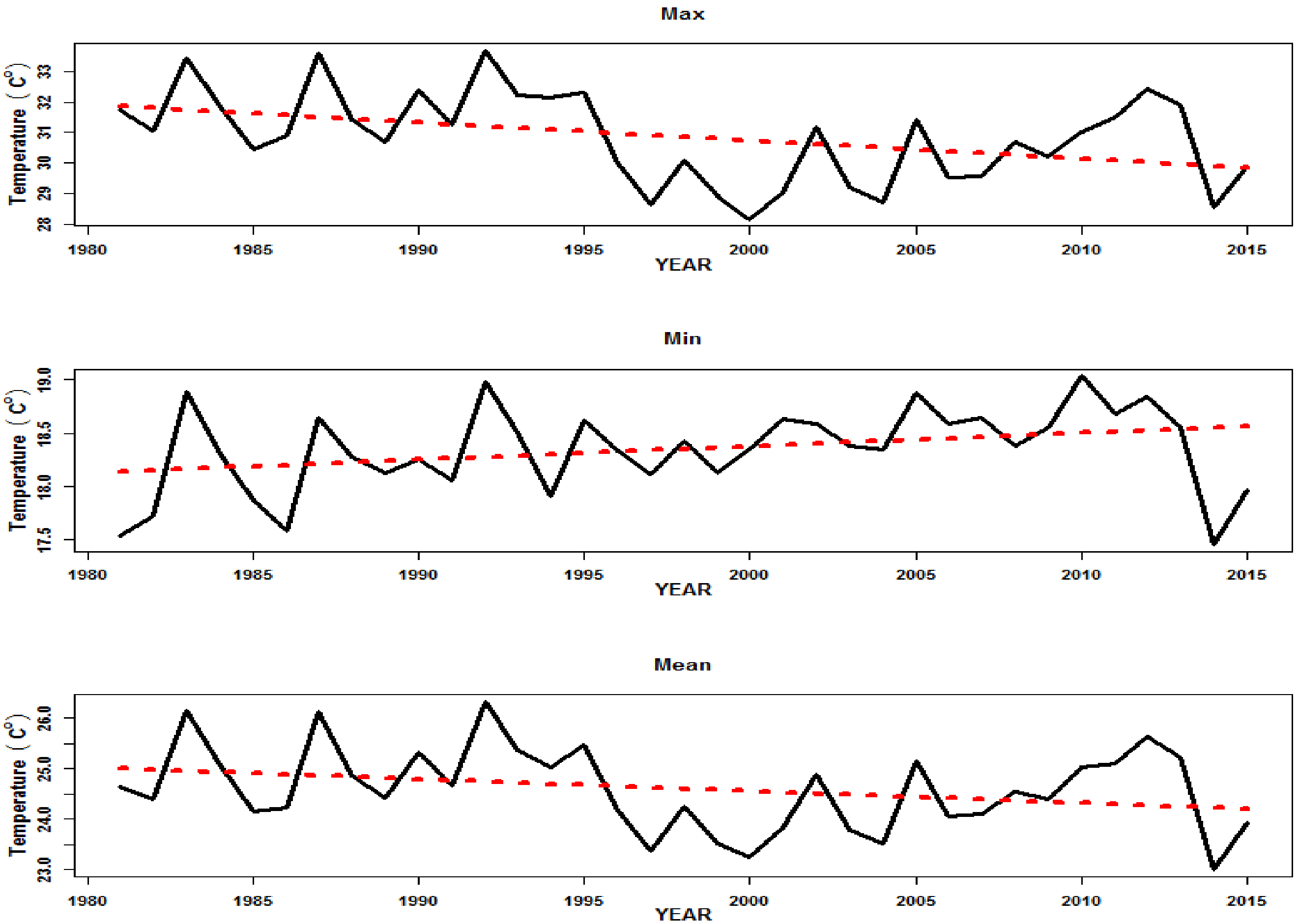
Nkhoma et al. (2020), Water and Climate Change

Wamkurumadzi



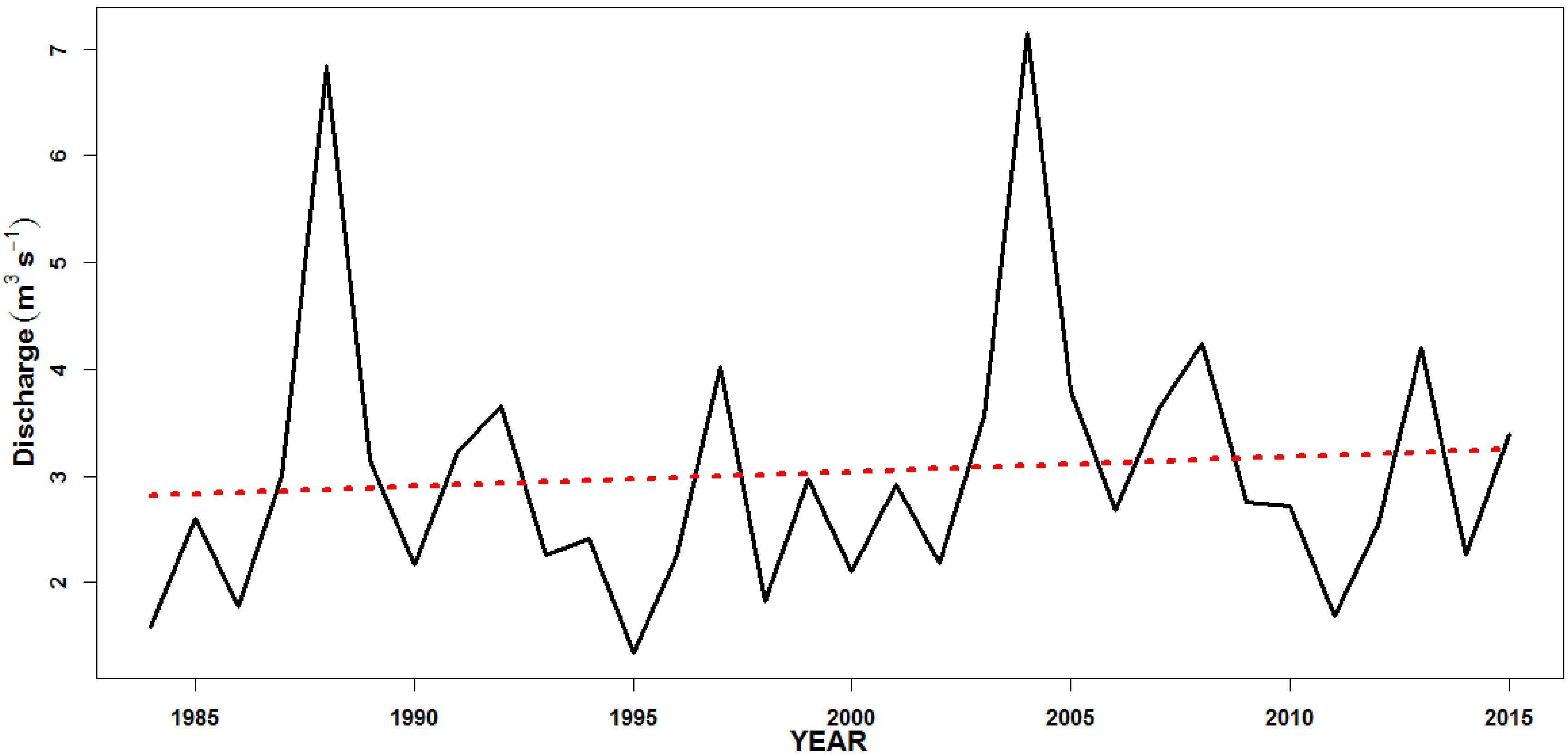
Nkhoma et al. (2020), Water and Climate Change

Wamkurumadzi



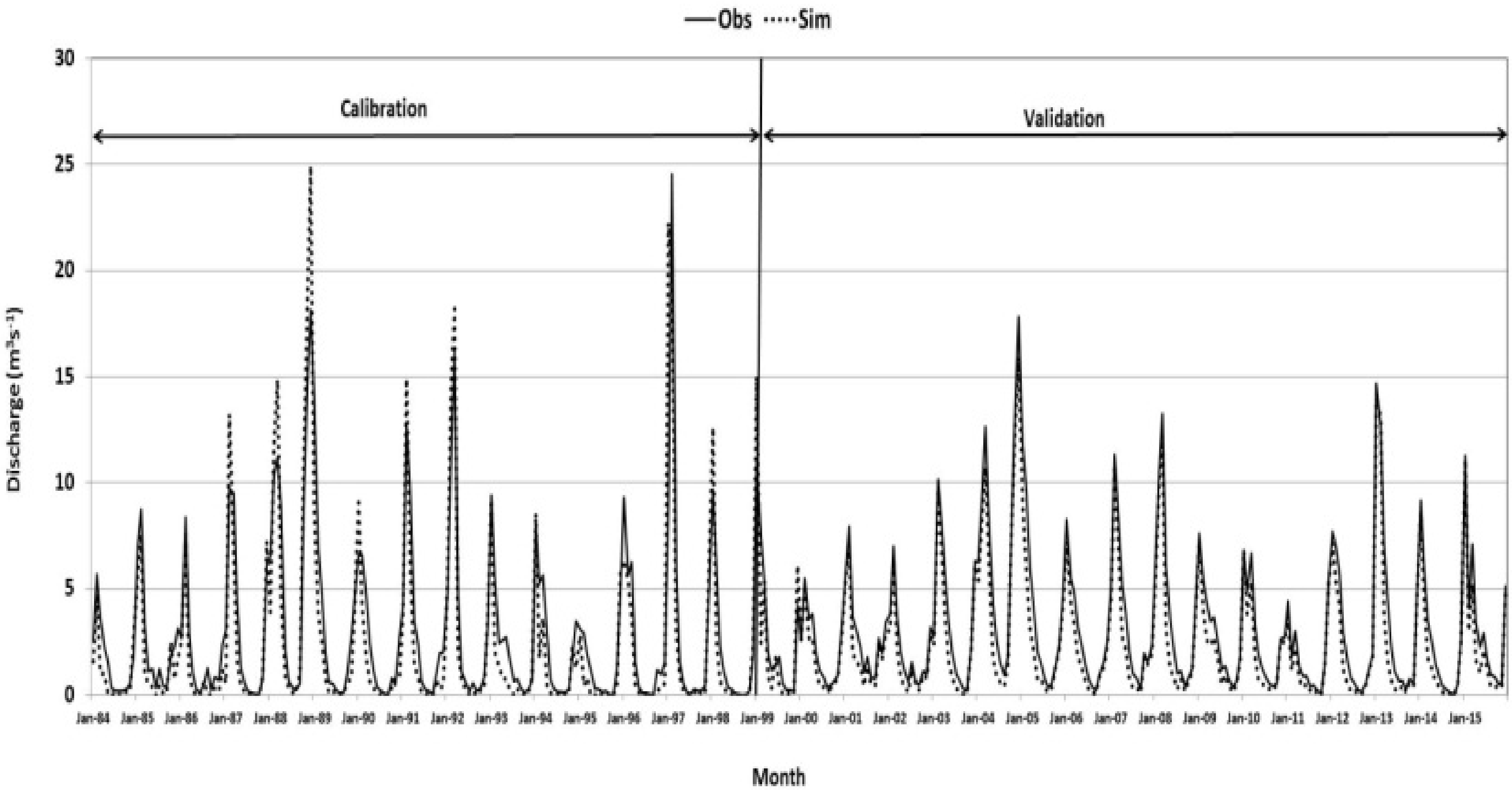
Nkhoma et al. (2020), Water and Climate Change

Wamkurumadzi



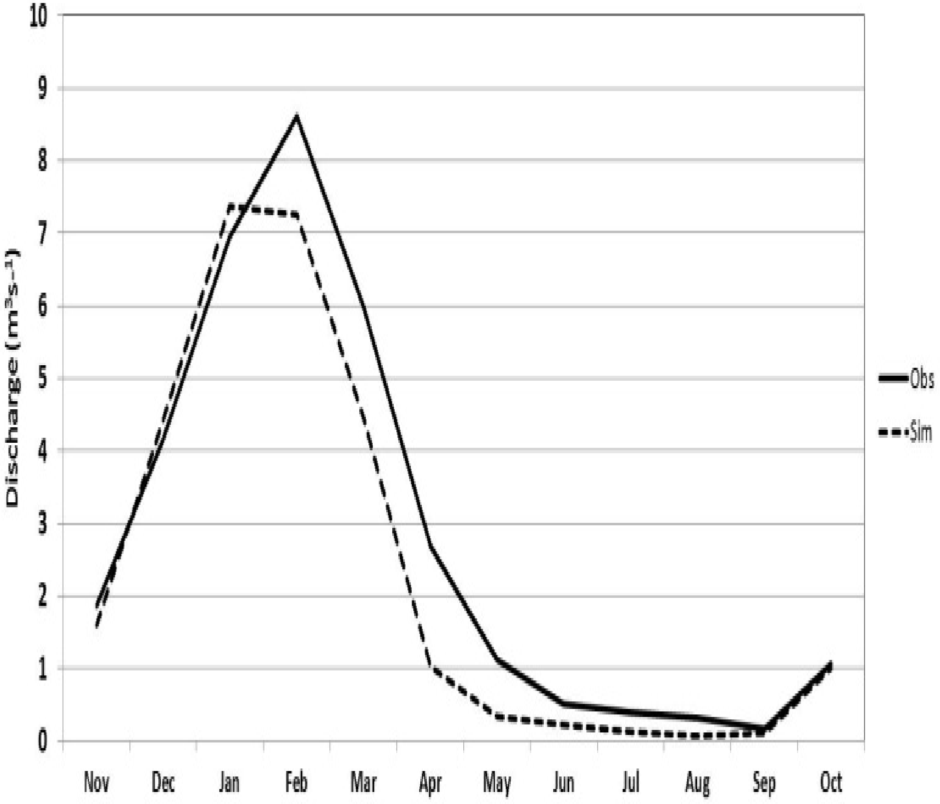
Nkhoma et al. (2020), Water and Climate Change

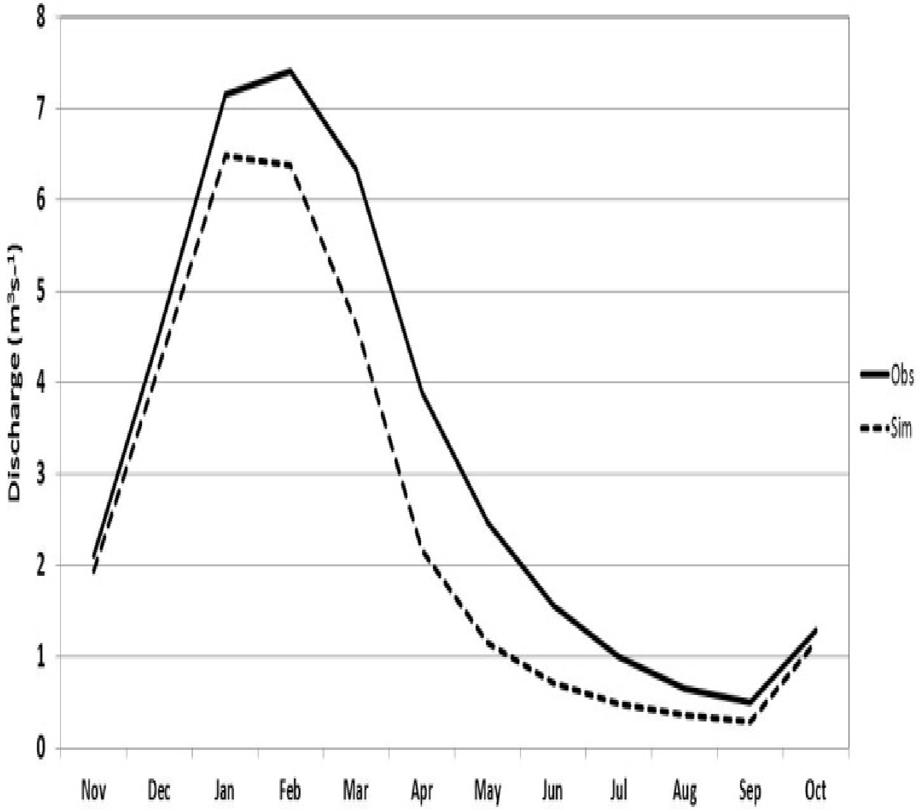
Wamkurumadzi



Nkhoma et al. (2020), Water and Climate Change

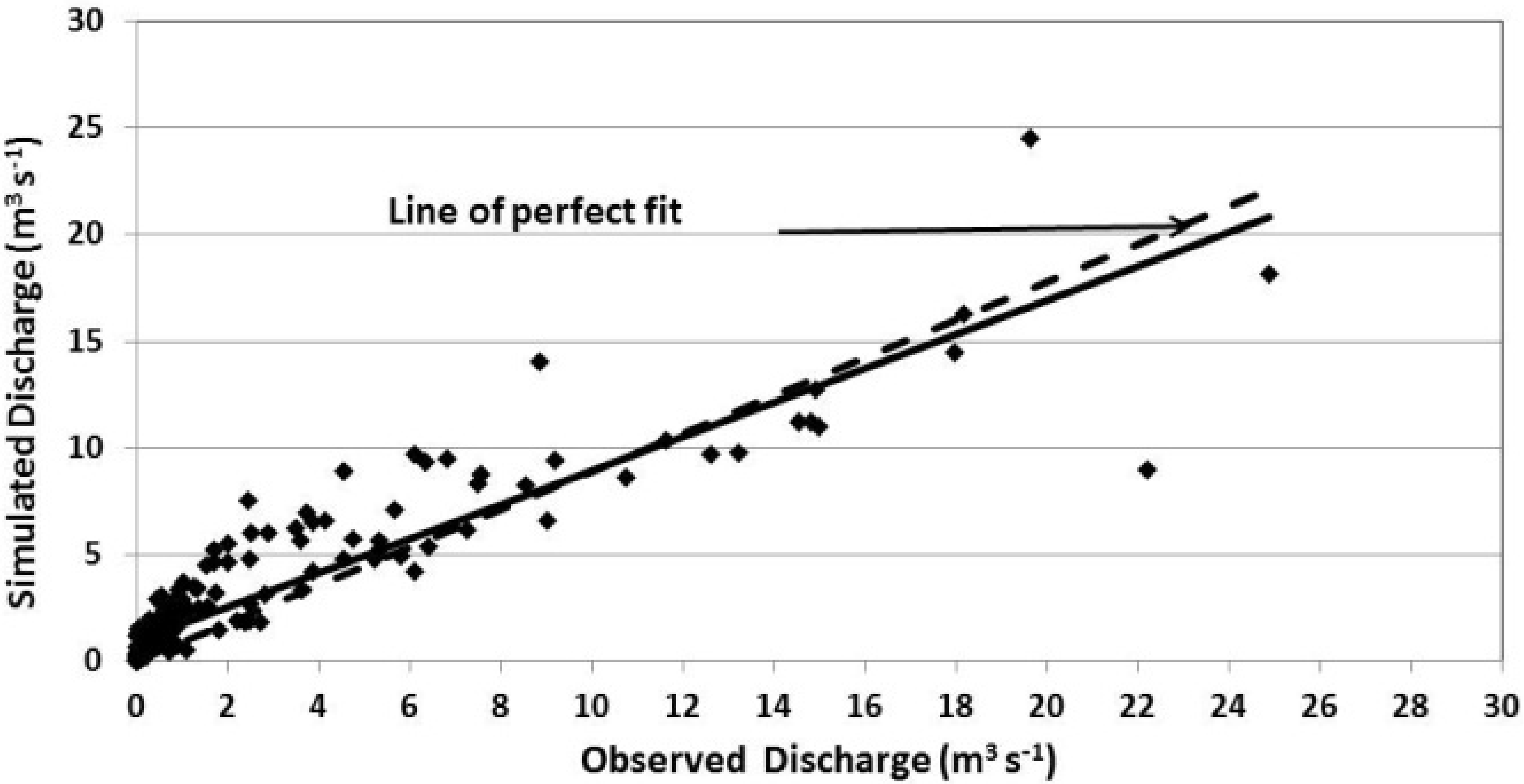
Wamkurumadzi





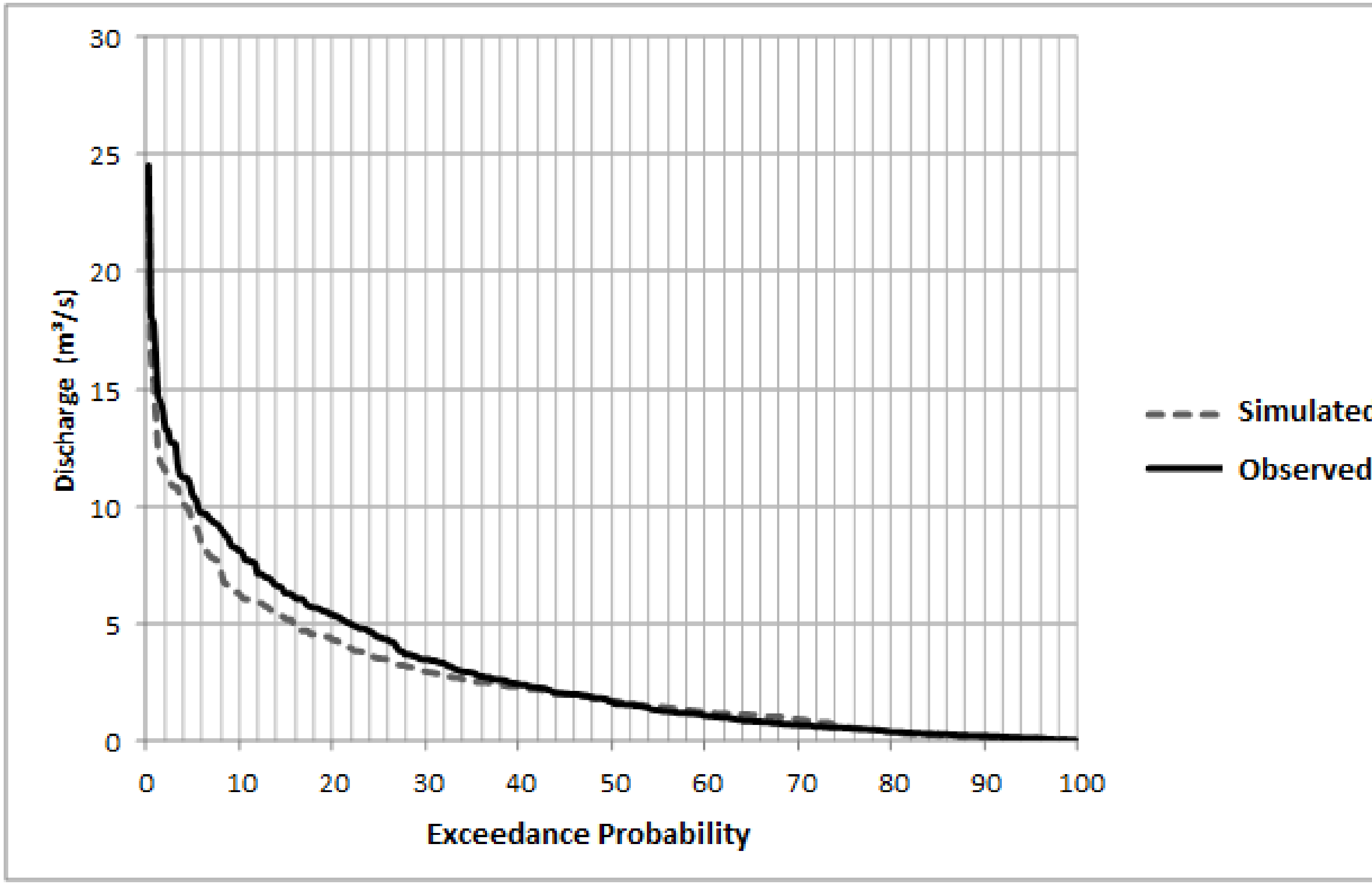
Nkhoma et al. (2020), Water and Climate Change

Wakurumadzi



Nkhoma et al. (2020), Water and Climate Change

Wakurumadzi



Nkhoma et al. (2020), Water and Climate Change

Means and Bias

$$\overline{S} = \frac{\sum_{i=1}^{N} S_i}{N} \qquad \overline{Q} = \frac{\sum_{i=1}^{N} Q_i}{N}$$

$$Percent\ Bias = \frac{\sum_{i=1}^{N} S_i - Q_i}{\sum_{i=1}^{N} Q_i}(100)$$

- Common calibration strategy: fix bias first, revisit periodically, goal of no bias
- PBIAS ± 25 %

- Maximum Error:

$$ME = \max(S_i - Q_i) \ \ for\ i = 1\, to\ N$$

- Percent Average Absolute Error

$$PAAE = \left\{ \frac{\frac{1}{N}\sum_{i=1}^{N}\left|S_i - Q_i\right|}{\overline{Q}} \right\}(100)$$

Sum of Squares of Errors

- Most common basis for statistical goodness of fit
  - e.g., least squares regression, seek to minimize

$$\sum_{i=1}^{N}(S_i - Q_i)^2$$

Root Mean Squared Error

$$RMSE = \left[\frac{\sum_{i=1}^{N} (S_i - Q_i)^2}{N} \right]^{1/2}$$

- Size of error usually related to size of events or values, thus *RMSE* typically smaller for dry periods, small watersheds (for example)

Percent RMSE

- Normalize *RMSE* by mean observed

$$PRMSE = \frac{RMSE}{\overline{Q}}(100)$$

- Because the magnitude of *RMSE* varies with magnitude of values, by minimizing *RMSE* only, which part of hydrographs are primarily best fit in calibration?

RSR

- Ratio of the Root Mean Square Error to the standard deviation of measured data

$$RSR = \frac{RMSE}{STDEV_{Obs}}$$

$$RSR \leq 0.70$$

Nash-Sutcliffe

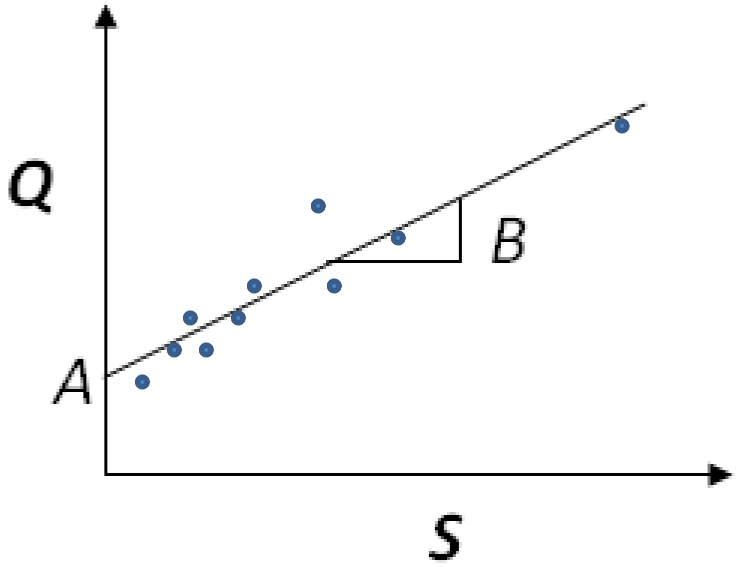
- Very popular method of evaluating calibration

$$R_{NS}^2 = 1 - \frac{\sum_{i=1}^{N} (S_i - Q_i)^2}{\sum_{i=1}^{N} (Q_i - \overline{Q})^2}$$

$$R^2{}_{NS} \geq 0.5$$

- **Reading**: McCuen, R. H., Evaluation of the Nash—Sutcliffe efficiency index, *Journal of Hydrologic Engineering*, 11(6), 597-602, 2006 (note: author uses different variables)

Line of Best Fit



$$Q = BS + A$$

Analyze as in regression: hypothesis testing on A and B, residual analysis, correlation coefficient…

Recommended Measures

- Moriasi et al. (2007):

  Nash–Sutcliffe (NS) coefficient (Nash and Sutcliffe 1970), percent bias error (PBIAS) and ratio of the root mean square error to the standard deviation of measured data (RSR)

Statistical Indicators

Calibration									
Period	Mean discharge $(m^3 s^{-1})$		NS	PBIAS	RSR	Days	R^2	r
	Obs	Sim						
1971–1986	156.4	158	0.83	1.0	0.38	0.96	0.86	0.93
1987–2001	102.04	110.6	0.82	−2.8	0.43	0.95	0.82	0.90

Validation								
Period	Mean discharge $(m^3 s^{-1})$		NS	PBIAS	RSR	Days	R^2	r
	Obs	Sim						
1987–2001	102.04	106.40	0.82	−6.9	0.43	0.95	0.82	0.91
1971–1986	156.4	154.8	0.83	−2.2	0.41	0.95	0.83	0.91

Model performance by simple split-sample test in the upper Kafue River basin during 1971–1986 and 1987–2001

Source: Ngongondo et al. 2013, Stoch Environ Res Risk Assess

Statistical Indicators

Calibration									
Period	Mean discharge ($m^3 s^{-1}$)		NS	PBIAS	RSR	Days	R^2	r
	Obs	Sim						
1971–1980	170.8	165.5	0.84	−3.1	0.40	0.95	0.84	0.92
1981–1990	127.5	114.8	0.81	−8.2	0.43	0.95	0.83	0.91

Validation								
Period	Mean discharge ($m^3 s^{-1}$)		NS	PBIAS	RSR	Days	R^2	r
	Obs	Sim						
1981–1990	127.5	120.3	0.82	−3.9	0.42	0.95	0.82	0.91
1971–1980	170.8	168.2	0.84	−1.6	0.40	0.96	0.84	0.92

Model performance by differential split-sample during test wet (1971–1980) and dry years (1981–1990) for the upper Kafue River basin

Source: Ngongondo et al. 2013, Stoch Environ Res Risk Assess

Equifinality

- Multiple combinations of parameters can lead to similar results
- Issue with both multi-parameter lumped models (e.g., SAC-SMA) and spatially distributed models (e.g., CASC-2D)

- **Reading**: Ebel, B. A. and K. Loague, Physics-based hydrologic-response simulation: Seeing through the fog of equifinality, *Hydrological Processes,* 20(13), 2887–2900, 2006

End Of Lecture 3