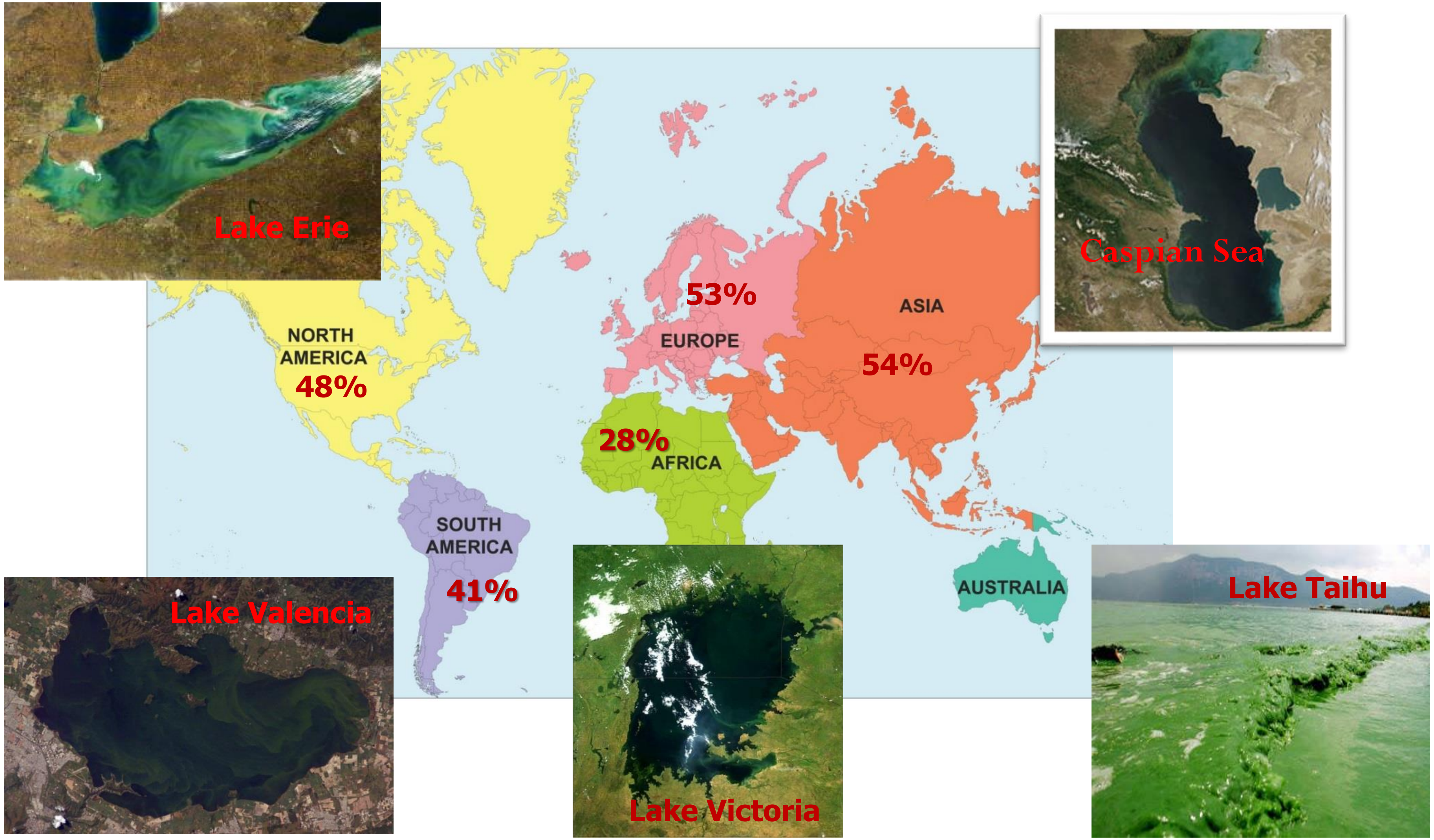
Understanding Harmful Algal Bloom Dynamics in a Mediterranean Hypereutrophic Reservoir insights from a Bayesian Network and a Structural Equation Model

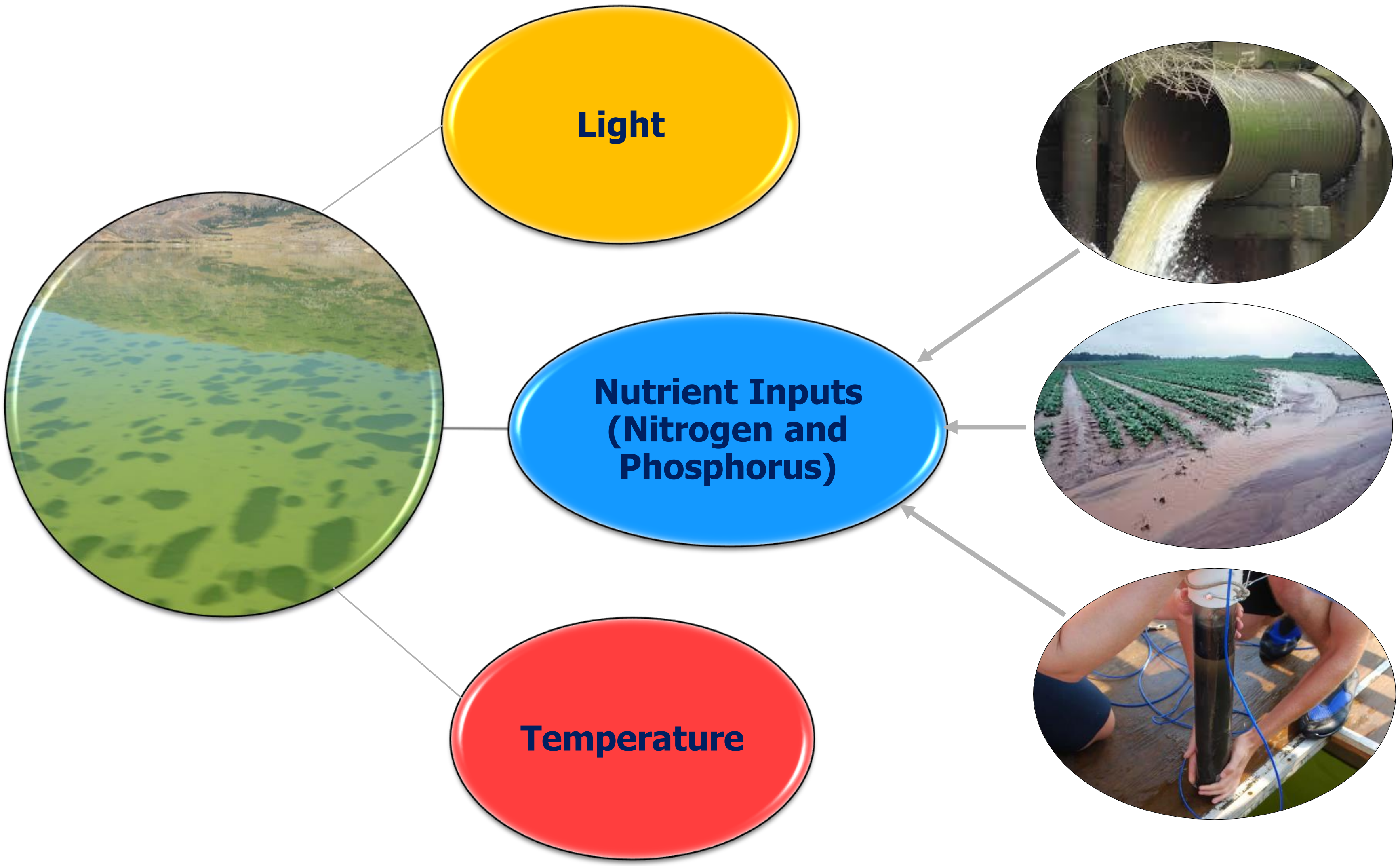
May 2020

Ibrahim Alameddine
Eliza S. Deutsch

Eutrophication: a Global Problem

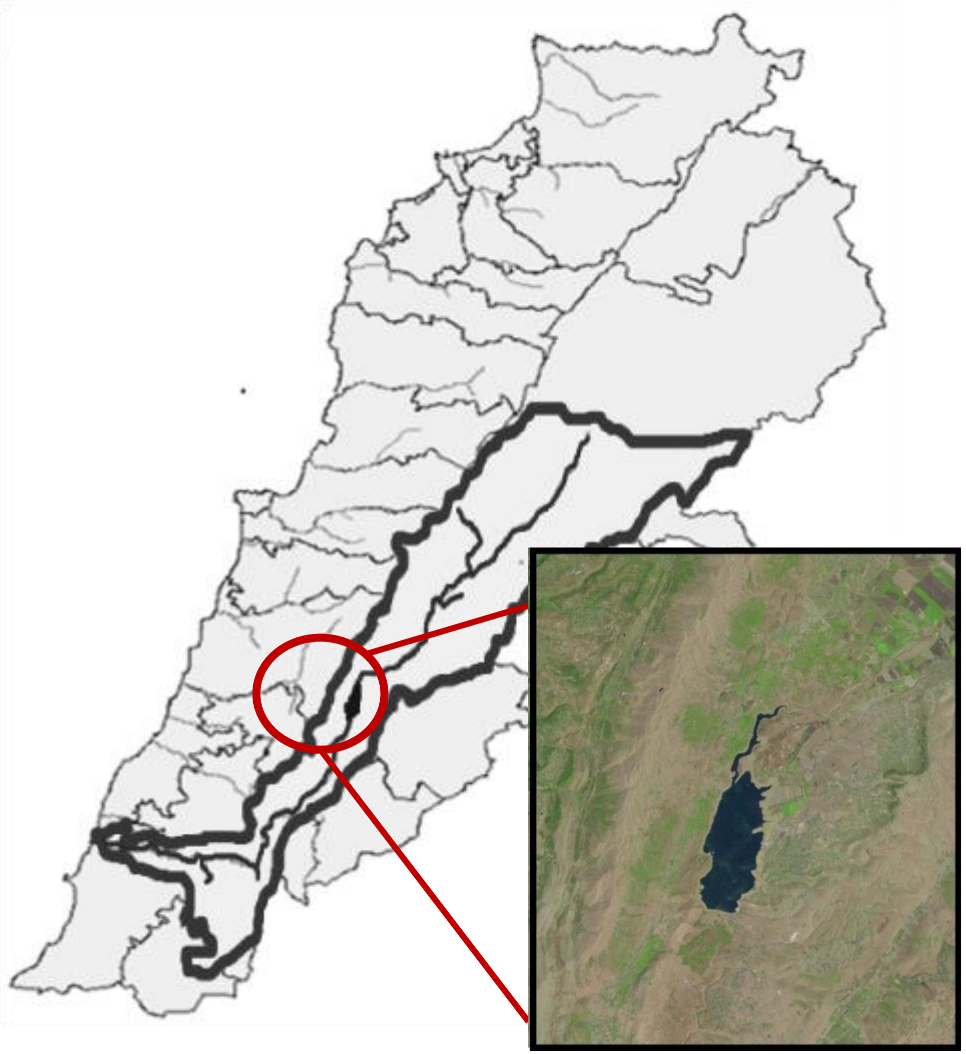


Drivers of Eutrophication



Case Study: Qaraoun Reservoir

- Dam completed in 1959
- Surface area: 4-11 km^2
- Depth near dam: >45 m
- Useful volume: 220 MCM
- Upstream catchment 1600 km^2



Uses:

- Hydropower generation
- Irrigation of 68,000 acres
- Some tourism
- Small fishing industry
- **Potential for domestic water supply**

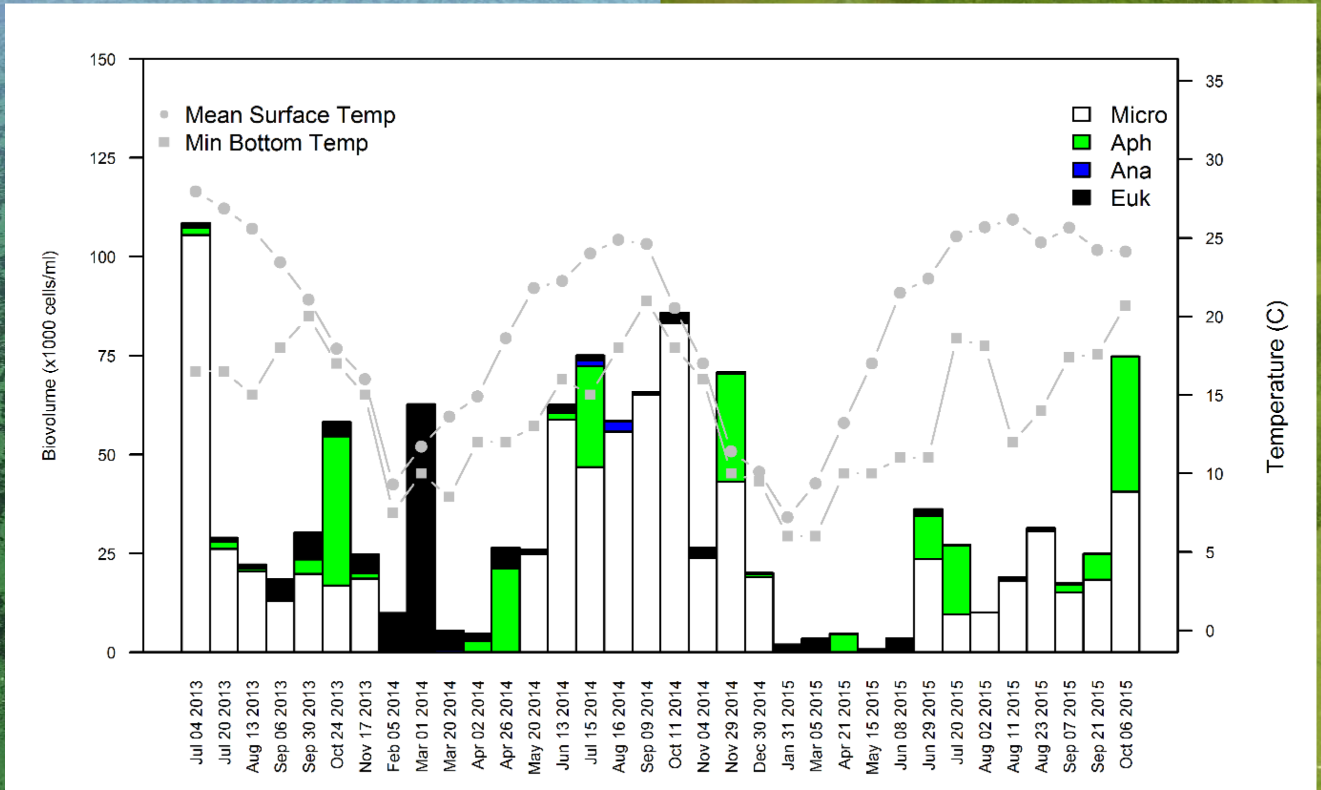
Qaraoun Reservoir: Water Quality

- Sporadic water quality studies:
  - Monitoring from 2013-2019: sampling ~ every 24 days
  - **High N and P external loads**
  - Little information on internal vs external loads
  - Reservoir is often **hypereutrophic**
  - Excessive ***Microcystis*** blooms during the growing season
  - **Microcystin** levels several times above the WHO recommended limits ➔ access to lake prohibited

Microcystis

Major player in Qaraoun Reservoir algae community/ eutrophication outcome

What drives *Microcystis* biovolume?



Microcystis aeruginosa

***Microcystis aeruginosa* Bloom Qaraoun Reservoir**

Can we develop a robust model to better understand and predict *Microcystis* blooms given our *prior* knowledge of the process and our limited data record?

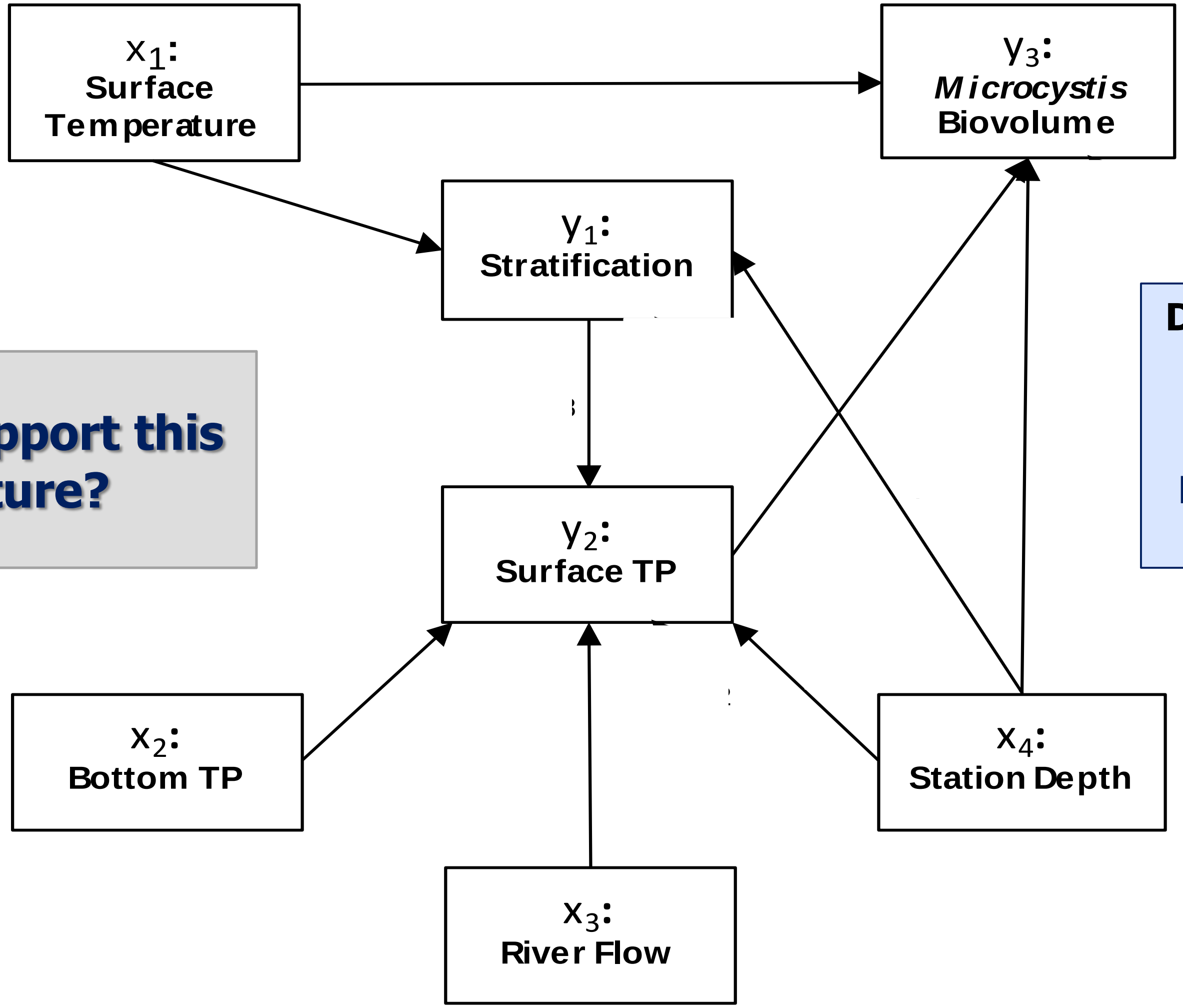
Modeling Approach

- Adopted **Structural Equation Modeling (SEM)** approach because it allows us to:
  - Use the data to assess our ***prior*** model structure and to calculate **modification indices ➔ assess model structure and propose changes**
  - Test theoretical/hypothesized relationships between variables
  - Account for direct and indirect pathways between variables
- ***A priori* method**: Develop conceptual model and test it against covariance structure in the data
  - Model fit by minimizing difference between model-covariances and data covariances (ML Estimator)
- **Null hypothesis**: covariance matrix implied in conceptual model is equal to observed covariance matrix of the data (p-value >0.05 = Good Model)

- Adopted **Bayesian Network (BN)** approach because:
  - Use data to "**learn**" the model structure (Necessary Path Condition algorithm), while assigning model structural constrains
  - Allow us to see if our data conflict with our *prior* model structure
  - Test theoretical/hypothesized relationships between variables
  - Assess the relative strength (Value of Information) between established parameter links
  - Identify the Markov Blanket of response variables
  - Use Bayes Theorem to generate predictions and propagate information both forwards and background ➔ a good **Decision Support System**

Will the data support our prior understanding of the system?
Will the two modeling methods converge on a similar model structure?
Will the strength of the pathways be similar between the two approaches?
Should we trust any of them?

Prior Model Structure: Literature and Expert Elicited



Several other important parameters were not measured

Data collected over 2 years:
(July 2013 – October 2015)

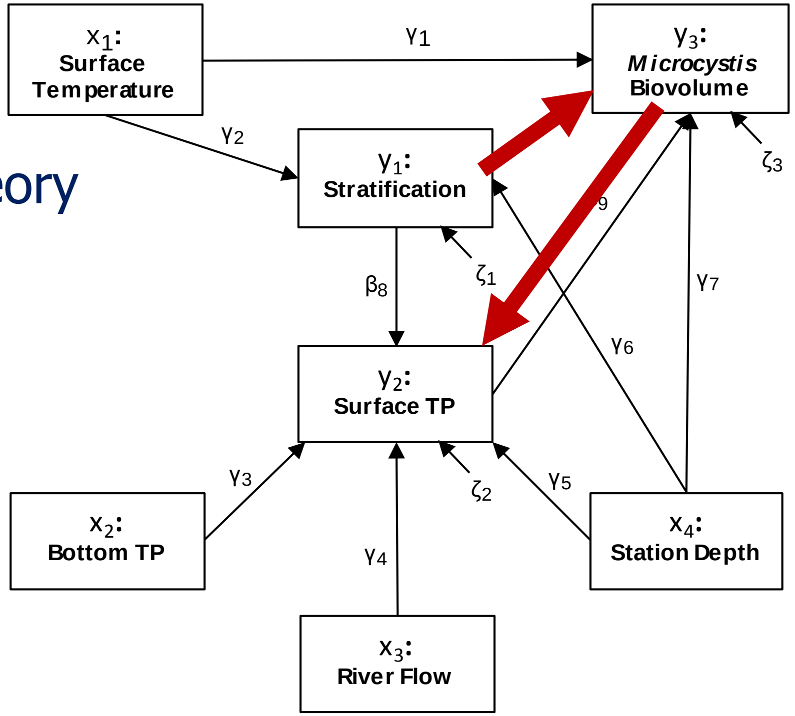
***Microcystis* biovolume**
Range = 0 – 154,000 cells/mL
Median = 17,000 cells/mL

Does the data support this model structure?

Initial SEM Results

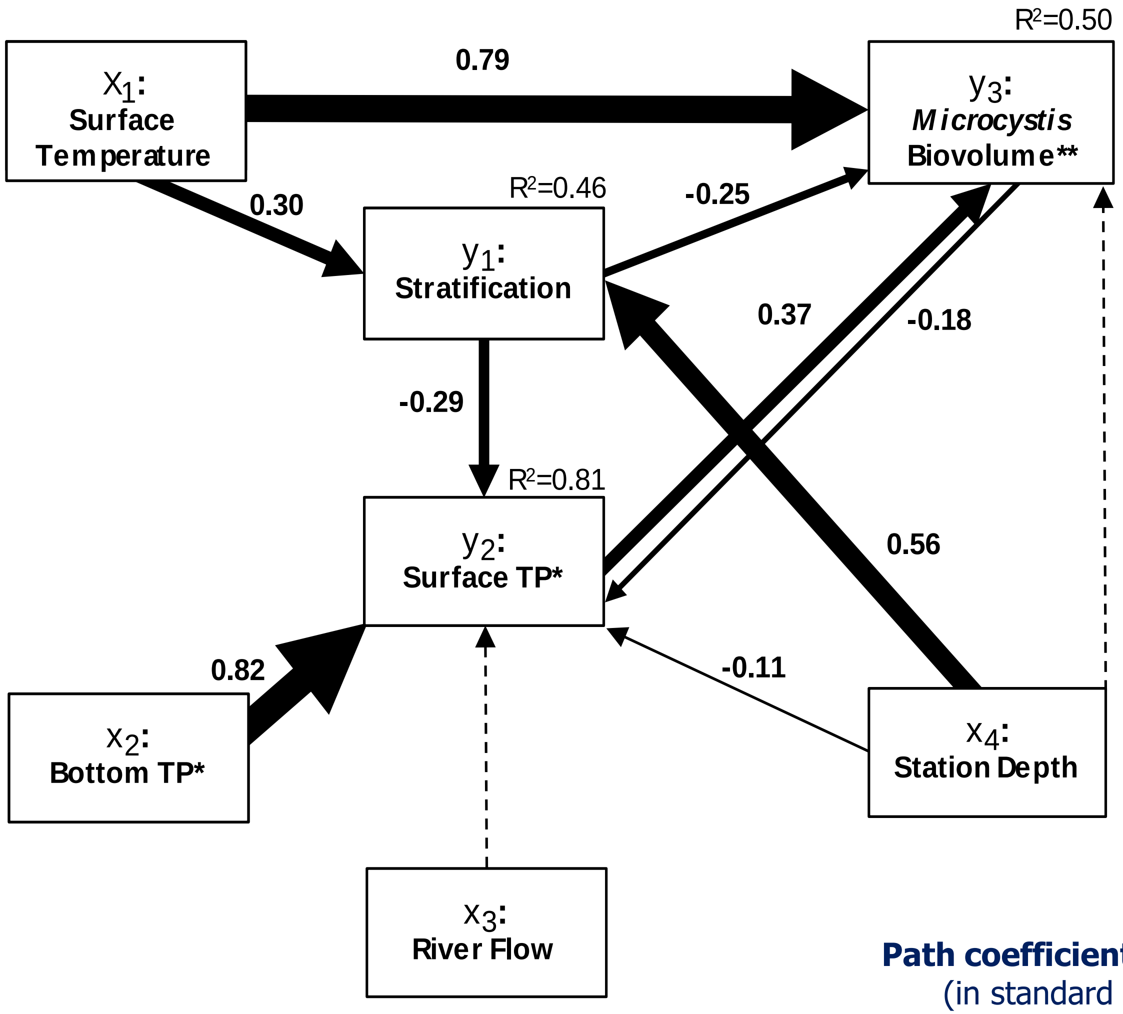
- Hypothesized model structure **DID NOT** concur with the covariance of collected data: **χ^2=30.7, df=6, p-value<0.0001**

- Used **modification indices** to improve model fit
  - Calculated based on expected decrease in χ^2 statistic given the addition of new model pathways
  - New pathways only considered if based on sound ecological theory
  - **2 proposed links and kept:**
    - **Stratification -> *Microcystis* Biovolume**
    - ***Microcystis* Biovolume -> Surface TP**



Final SEM

$\chi^2 = 7.85$; df = 4; p-value = 0.1



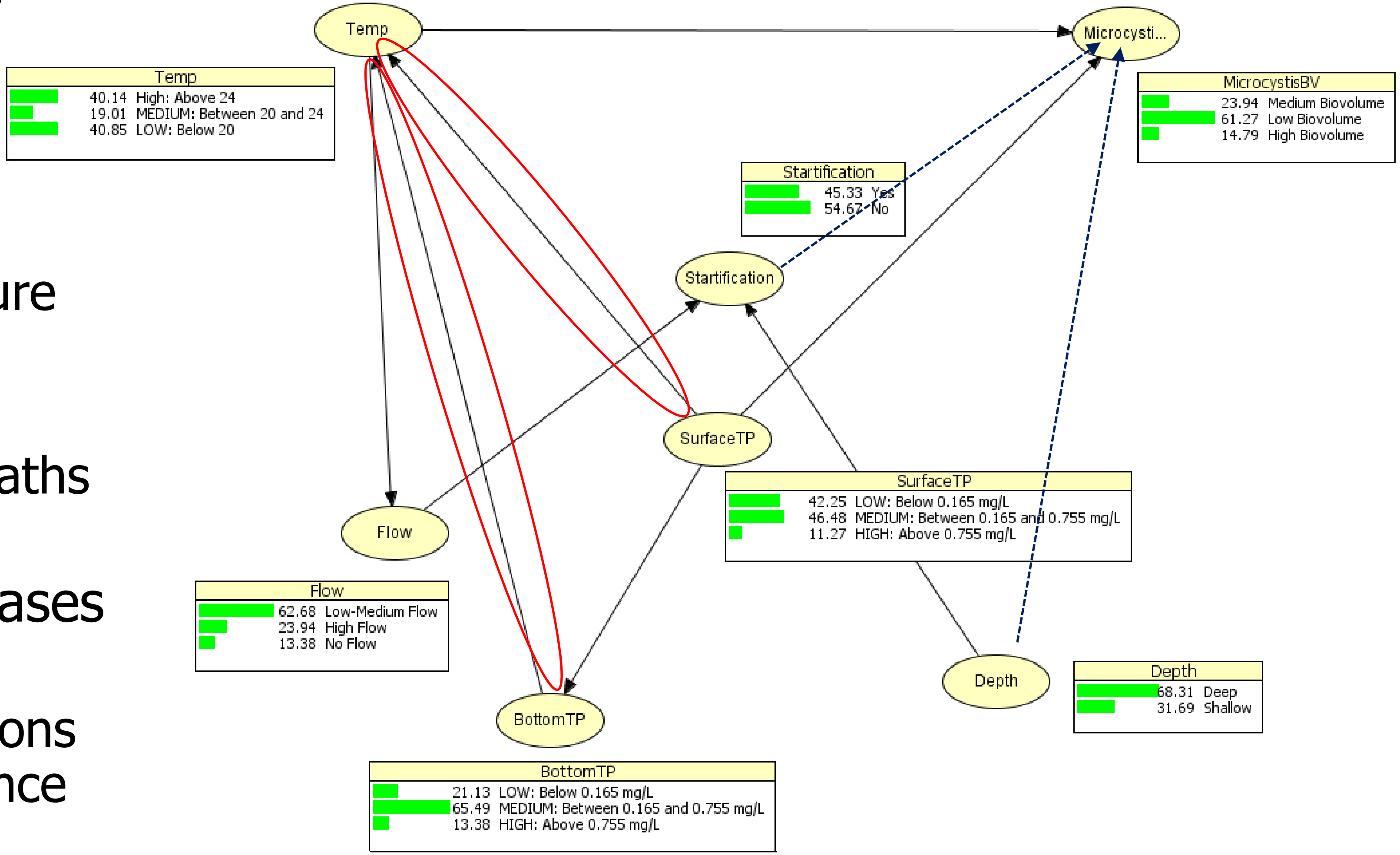
Path coefficients = standardized

(in standard deviation units)

$$standardized\ coefficient = unstandardized\ coefficient * \left(\frac{SDx}{SDy}\right)$$

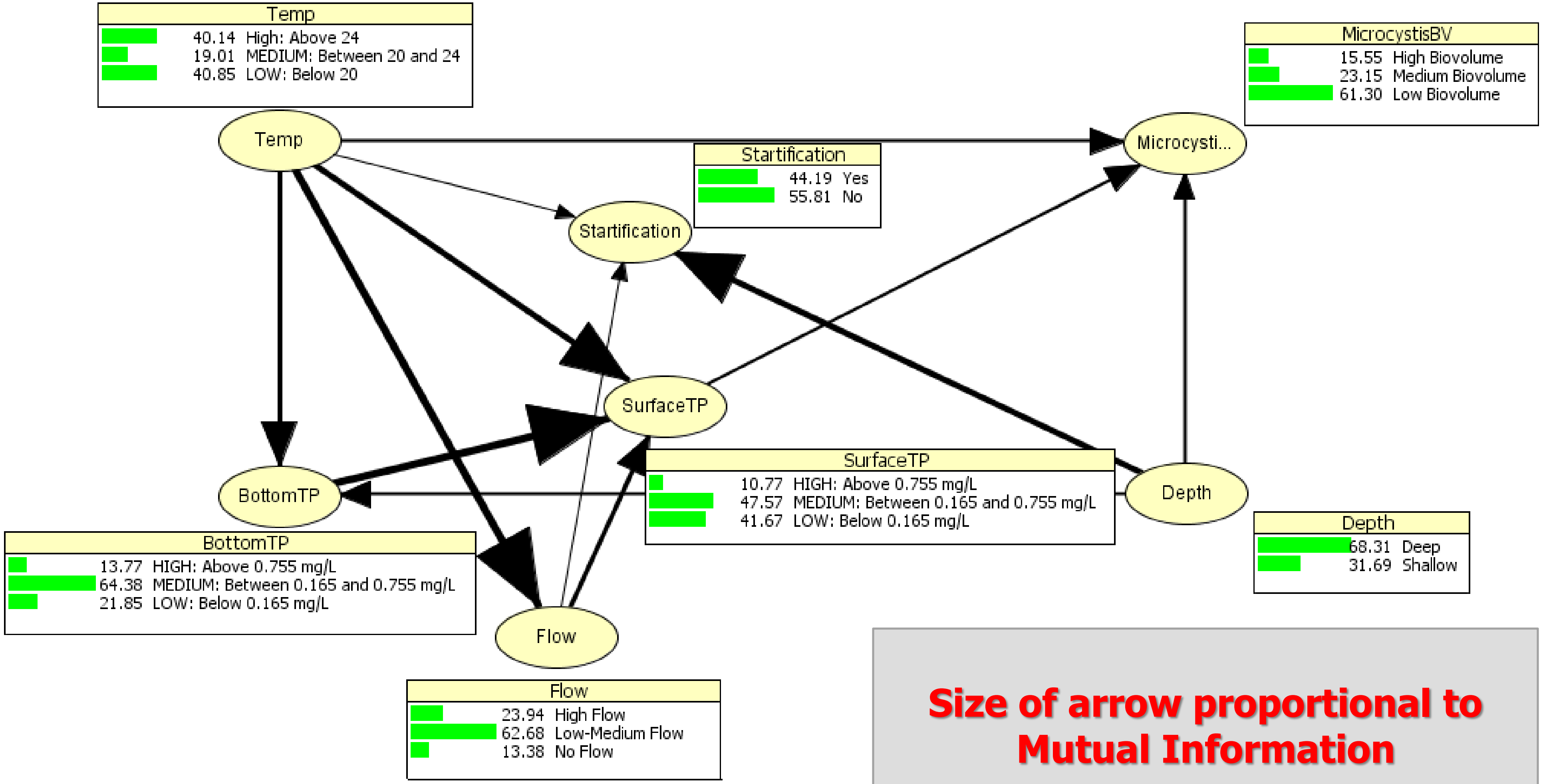
Initial BN Results

- Structure learning with the **NPC (Necessary Path Condition) algorithm** without any structural constraints resulted:
  - A reasonable model structure
  - Several reversed model pathways
  - Potentially a few missing paths (----)
- NPC is recommended for cases with limited data
  - Introduces ambiguous regions and asks for user interference



Final BN

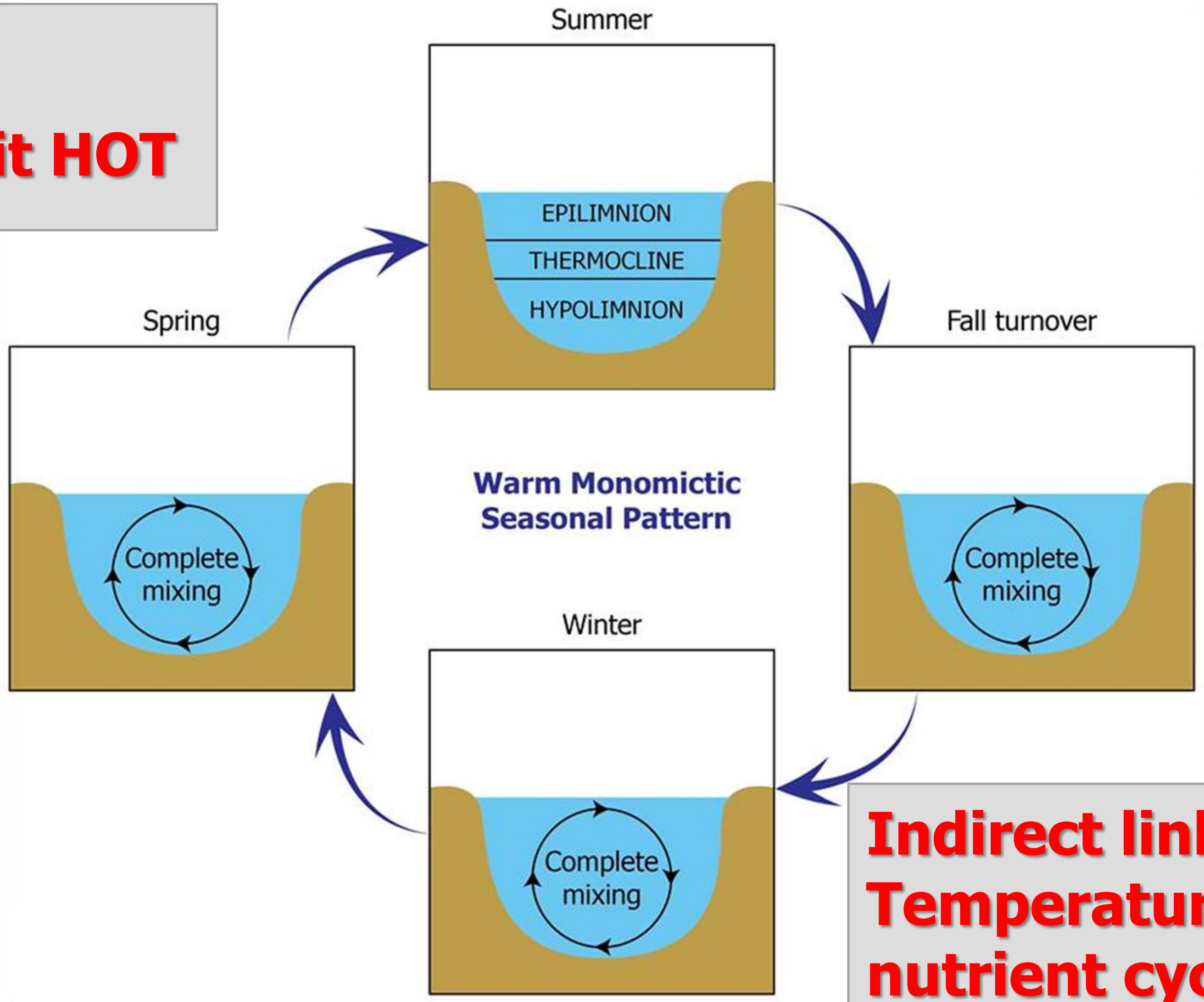
After adding Structural Constraints to the model and then learning the structure



Size of arrow proportional to Mutual Information

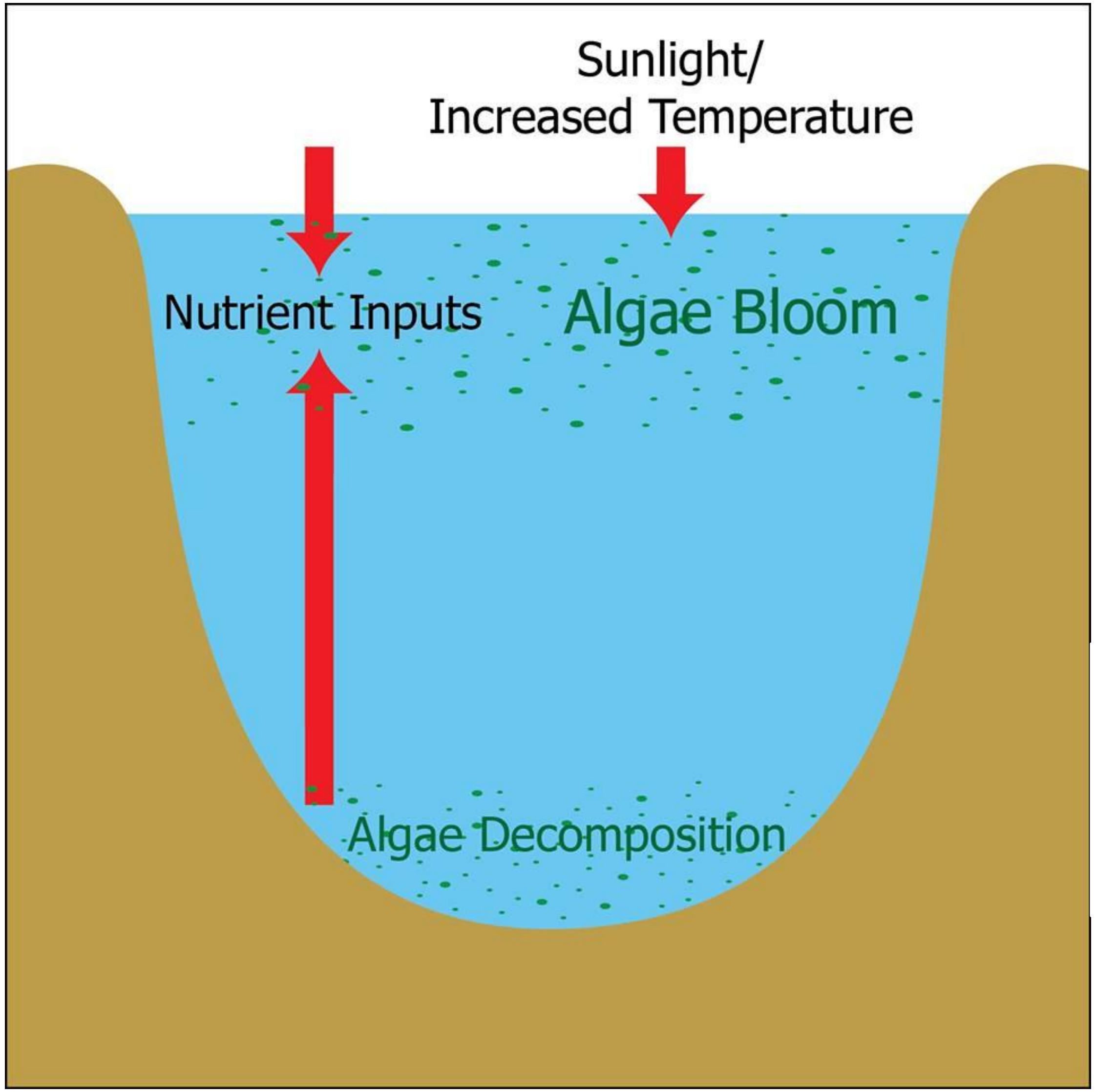
Question 1:
Direct vs Indirect Temperature Effects?

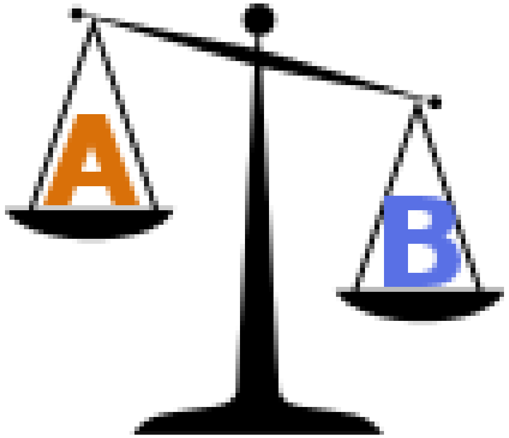
Direct link:
Cyanobacteria like it HOT



Indirect link:
Temperature affects nutrient cycling and mixing

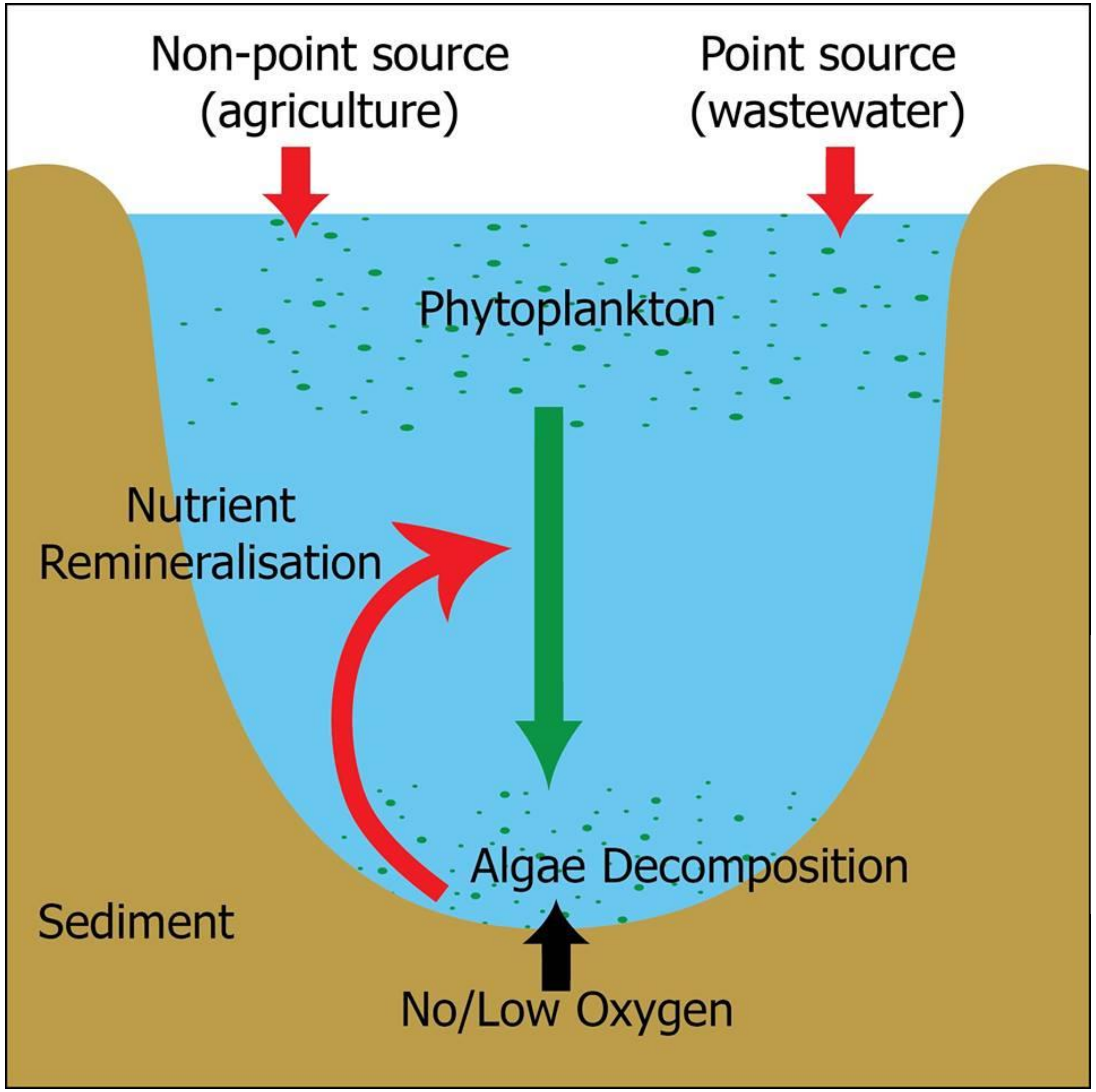
Question 2:
Temperature vs. Nutrients?

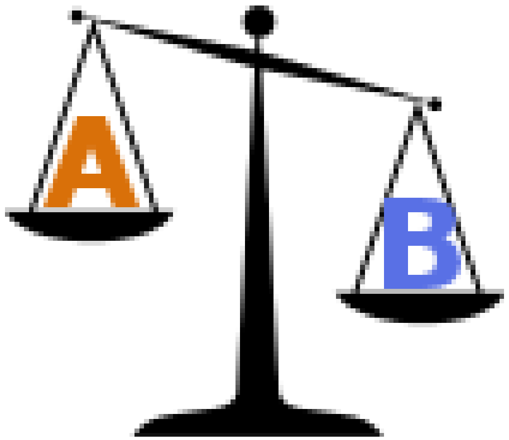




Relative importance

Question 3:
Internal vs. External Loads?





Relative importance

What do the two models tell us?

- **Direct temperature effects:**
  - **SEM:** Temp is the **MAJOR** promoter of *Microcystis* biovolume in Qaraoun given existing high nutrient concentrations
  - **BN:** Temp impact found to be less important than in SEM
- **Indirect temperature effects** in both models found to hinder *Microcystis* growth by limiting **nutrient** access and inducing **light** limitation (in the case of the SEM but not in the BN)
  - BN model included another pathway that meditated TP levels
- **Station depth** influenced stratification more strongly than temperature given drought conditions
  - **Shallow stations** have **higher nutrient concentrations** (internal loading) and tend to have **more blooms**
- **Surface Total Phosphorus levels:**
  - **SEM:** *Microcystis* pathway weaker than expected
    - Reservoir hypereutrophic
    - *Microcystis* strong competitors for nutrients
  - **BN:** Pathway found to be as important as Temp pathway
- **Internal loading** found to be much more important than external loading for promoting *Microcystis* blooms in both models
  - **External loads** during winter months expected to significantly contribute to high reservoir nutrient status & high summer internal loads

Predictions with SEM

$$\sqrt{Microcystis_i} = -1.65 + 0.40 * Temp_{surf_i} - 0.79 * Depth_i + 0.86 * ln(TP_{surf})_i - 0.22 * Stratification_i + \zeta_3 ; \qquad \zeta_3 \sim N(0,4.76)$$

where

$$ln(TP_{surf})_i = 0.28 + 0.89 * ln(TP_{bot})_i + 0.01 * Flow_i - 0.31 * Depth_i - 0.11 * Stratification_i - 0.08 * \sqrt{Microcystis_i} + \zeta_2 ; \qquad \zeta_2 \sim N(0,0.33)$$

and

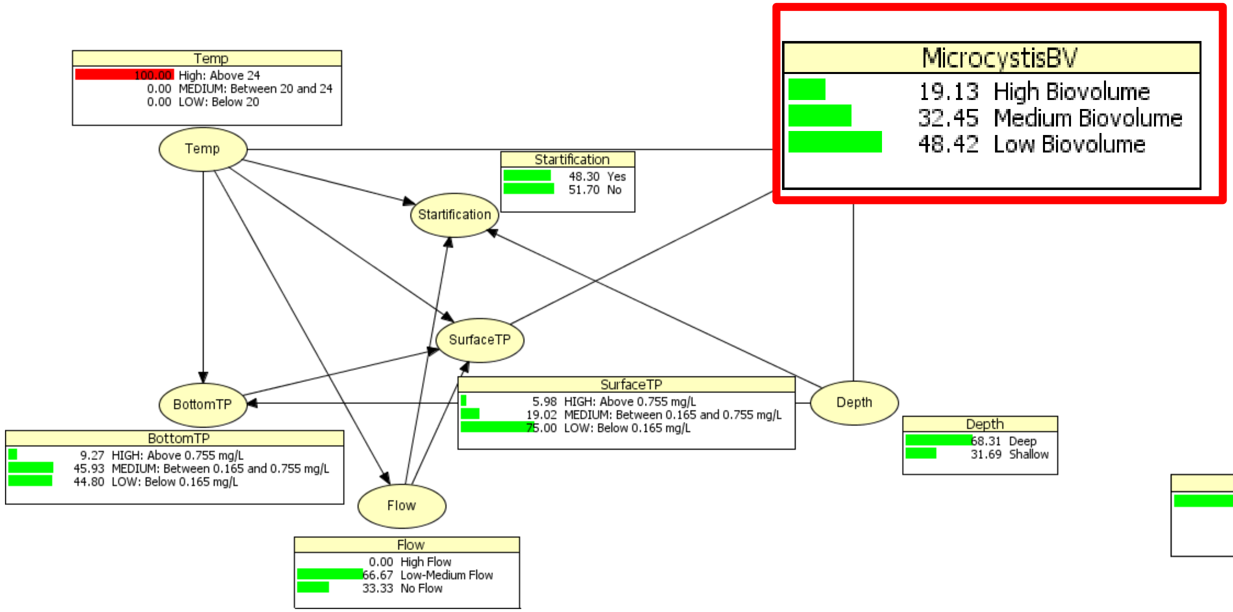
$$Stratification_i = -2.73 + 0.17 * Temp_{surf_i} + 4.17 * Depth_i + \zeta_1 ; \qquad \zeta_1 \sim N(0,6.62)$$

			Microcystis (organisms/mL)	
↑ Temp	↑ Strat	↑ TP_{surf}	Shallow	Deep
			27,000	19,000
√			51,000	40,000
√	√		-	28,000
		√	38,000	29,000
√		√	66,000	55,000

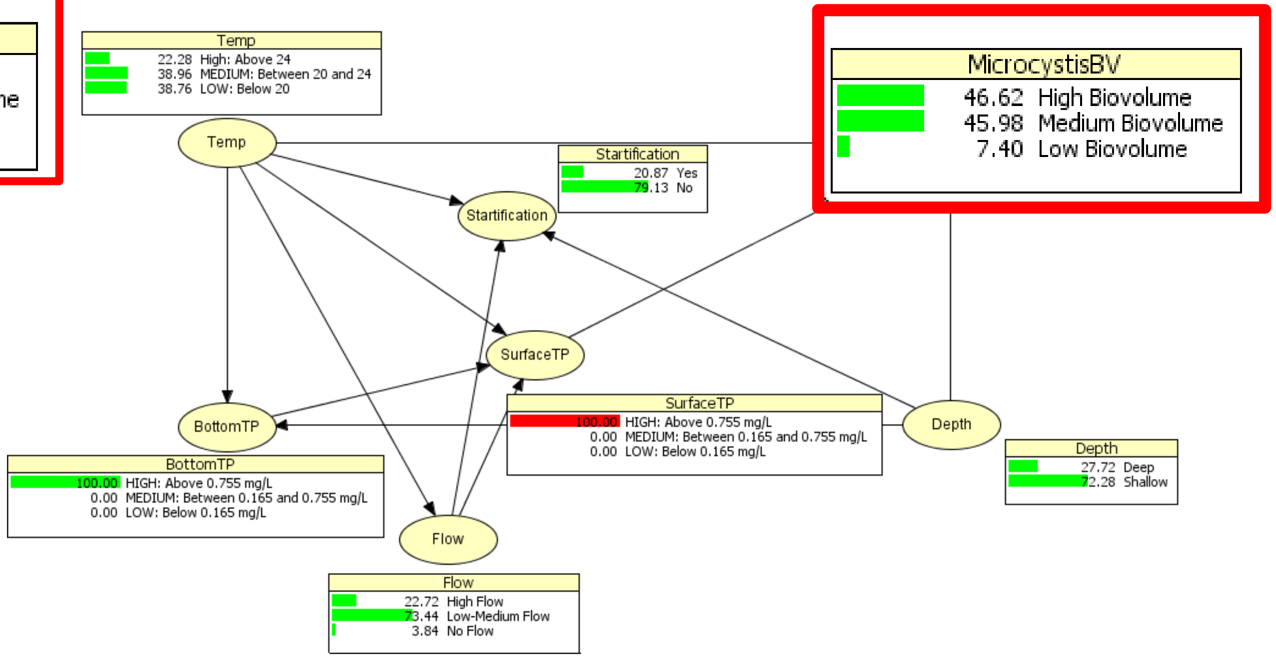
*Temp increased from 21.5°C to 26.5°C
*Strat increased from 2°C to 7°C
*TP_{surf} increased from 0.23 mg/L to 0.73 mg/L

Predictions with BN

Temp Alone

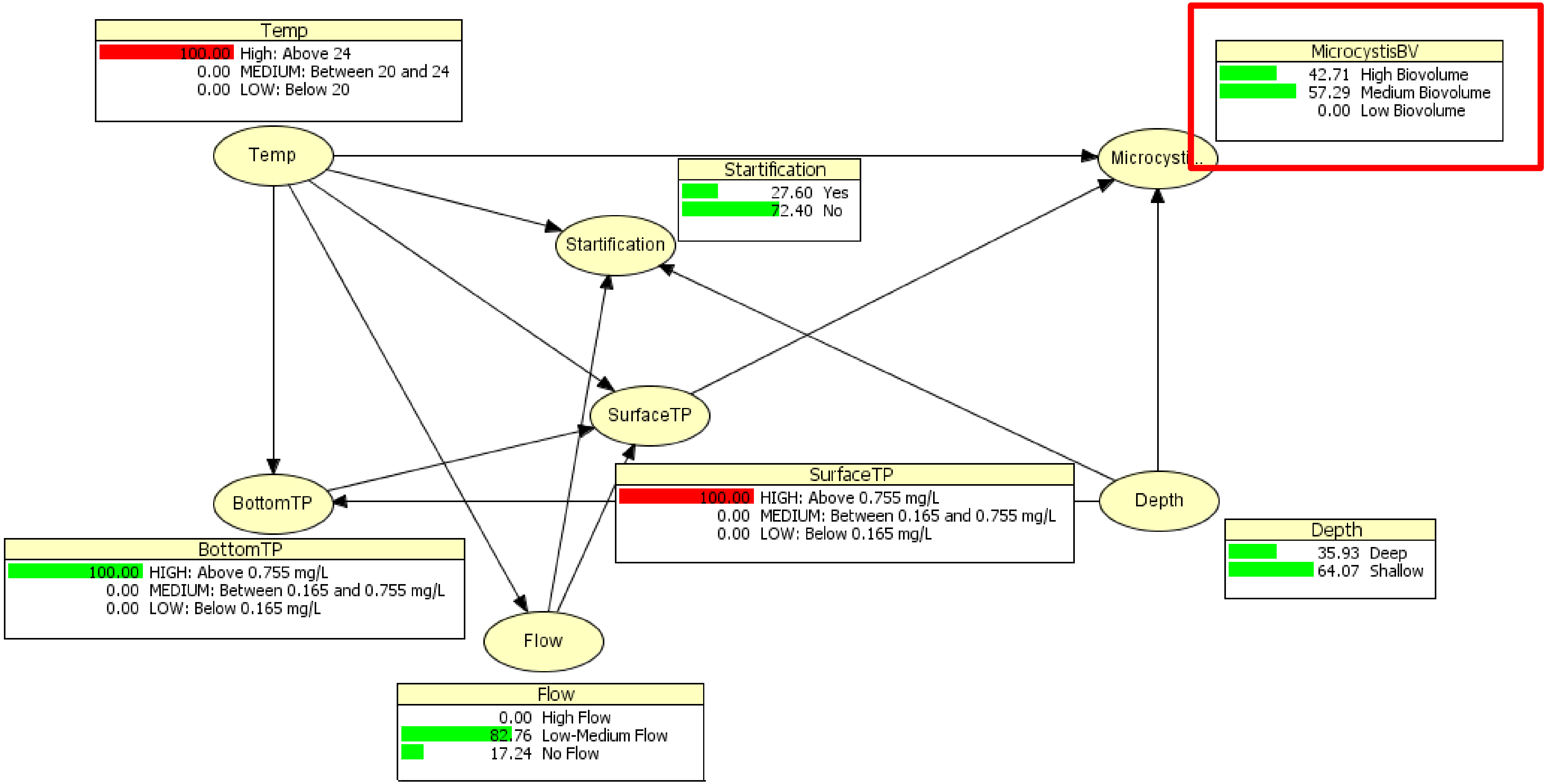


High TP



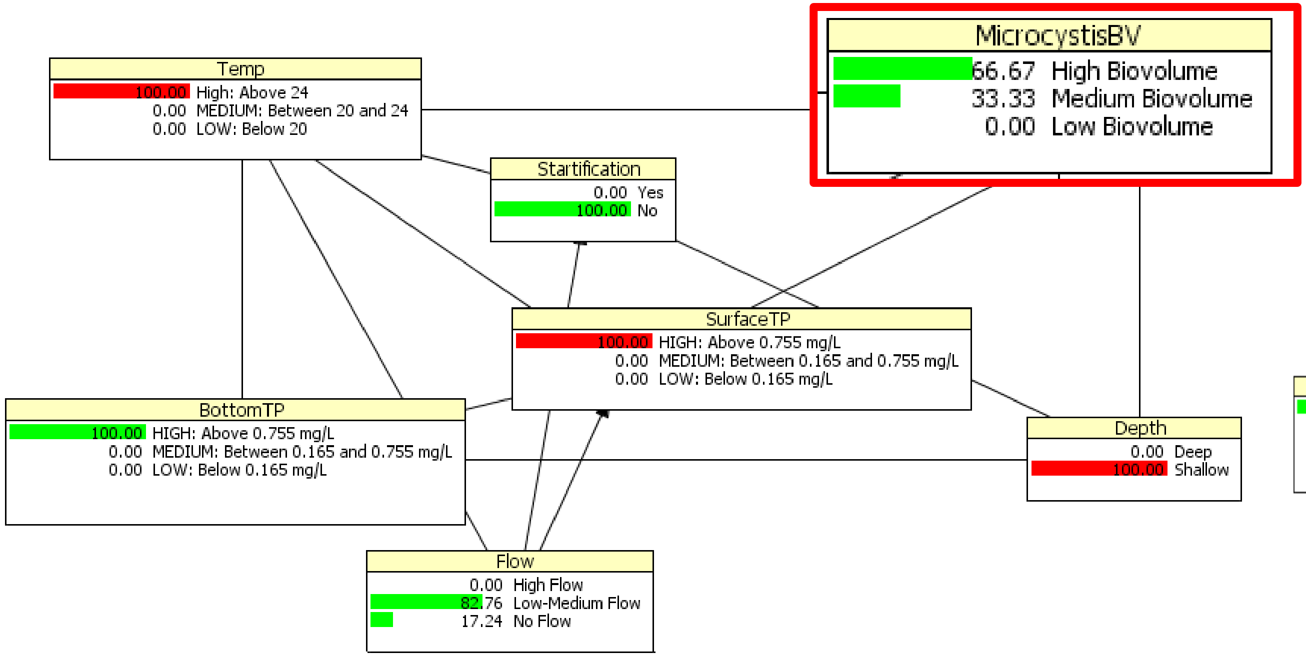
Cannot completely separate the two given the network
The sum of all pathways of Temp appears to be less important than TP

Predictions with BN (High Temp +High TP)

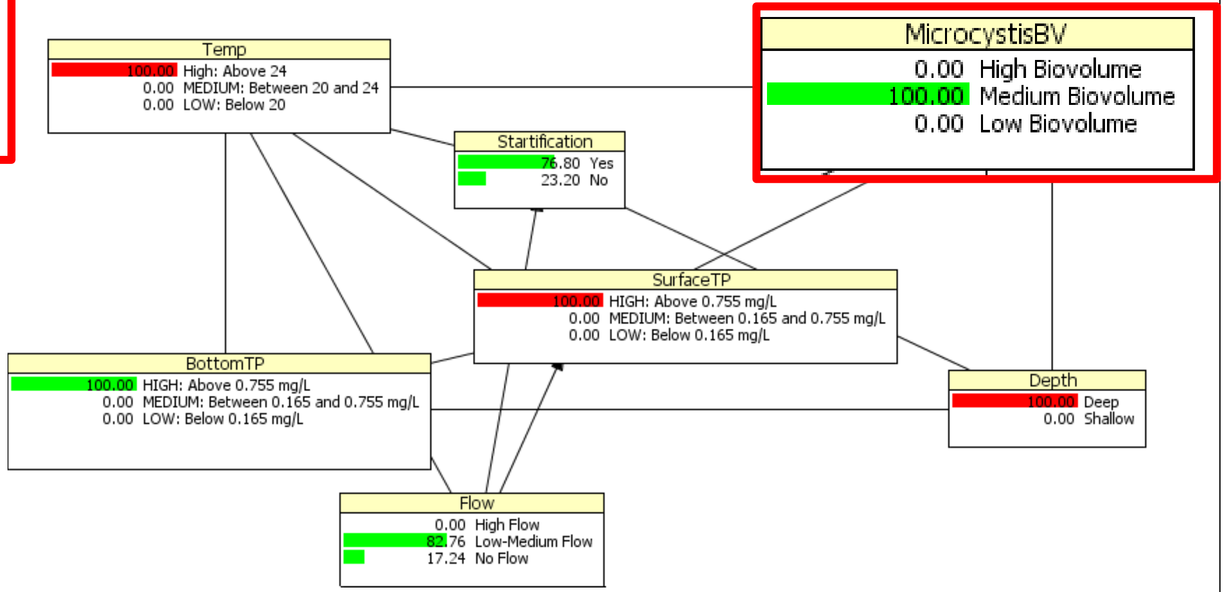


Predictions with BN (High Temp +High TP)

Shallow



Deep



Significant differences in Microcystis biovolume between shallow and deep parts
More pronounced in shallow areas
Stratification dynamics is captured by model

Conclusions

- Our ***prior*** expert elicited and literature based model structure was not supported by the data
- The modifications introduced by the **BN NPC** structure learning algorithm and the **SEM modification indices** were meaningful and accepted
- The two "learned" models structures were largely similar
  - Yet differed in estimating the relative importance of temperature versus surface TP levels
- With limited data, the validity of their structural features remains to be assessed

Acknowledgements

Funding provided through:

USAID
FROM THE AMERICAN PEOPLE

1866
AUB
American University of Beirut
الجامعة الأميركية في بيروت

NSERC
CRSNG