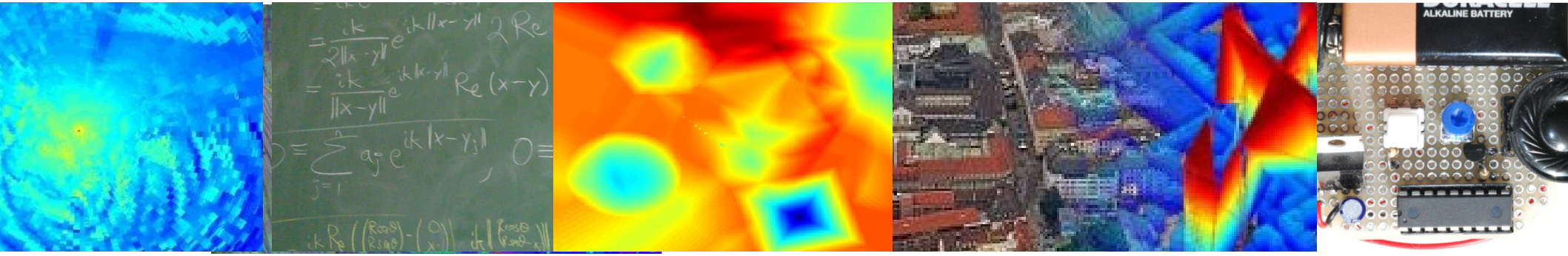
Topological Features
into
Complex Models



Michael Robinson

Acknowledgments

- Students:
  - Zander Memon (AU)
  - Harry Pham (AU)
  - Trixie Southwood (AU)
  - Maxwell Gualtieri (Northwestern)
- Collaborators:
  - Donna Dietz (AU)
  - Brian DiZio (NUWC Newport)
  - Michael Postol
  - Michael Szulczewksi (MITRE)
  - James Thorson (NOAA)
- Funding: Jason Chaytor (ONR) and PNNL KBASE

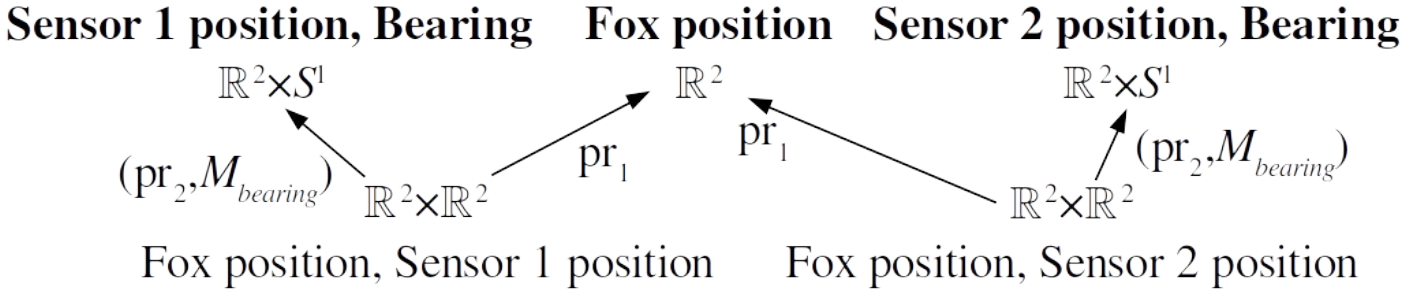
- Main references:

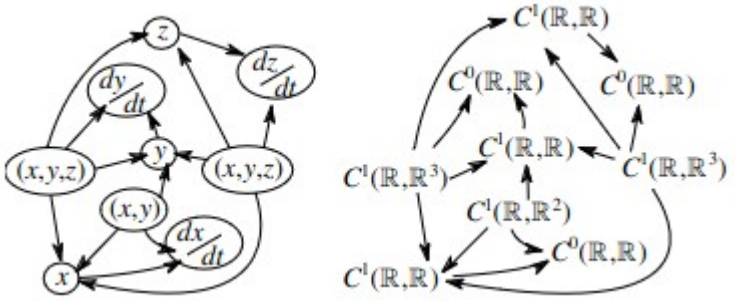
https://arxiv.org/abs/2511.04603

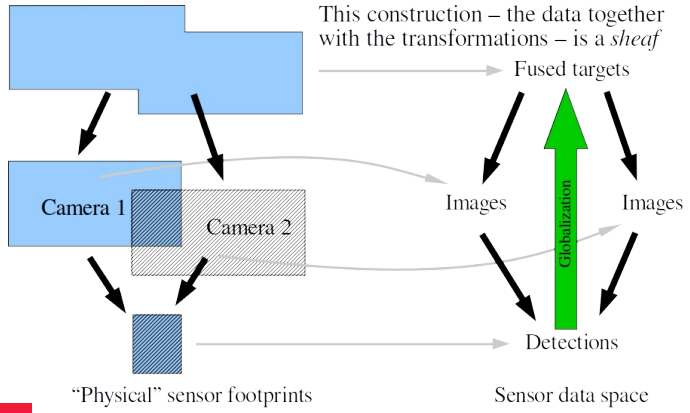
https://github.com/kb1dds/netlist-sheaf

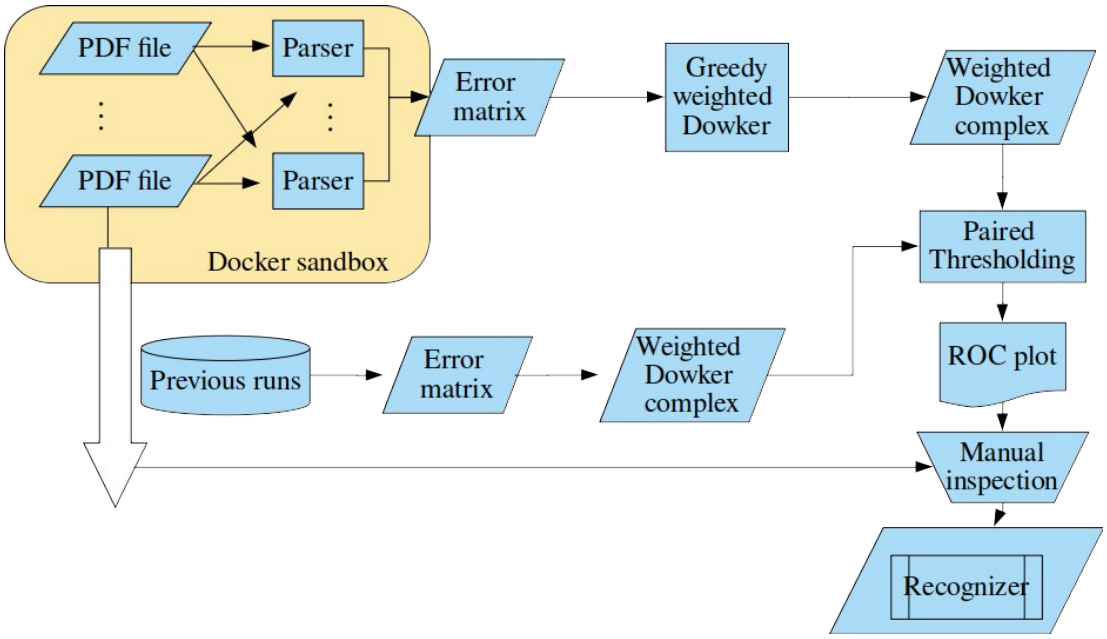
Thorson, J. T., Andrews, A. G., Essington, T., & Large, S. (2024). Dynamic structural equation models synthesize ecosystem dynamics constrained by ecological mechanisms. Methods in Ecology and Evolution 15(4): 744-755. https://doi.org/10.1111/2041-210X.14289

Systems of systems are diagrams...



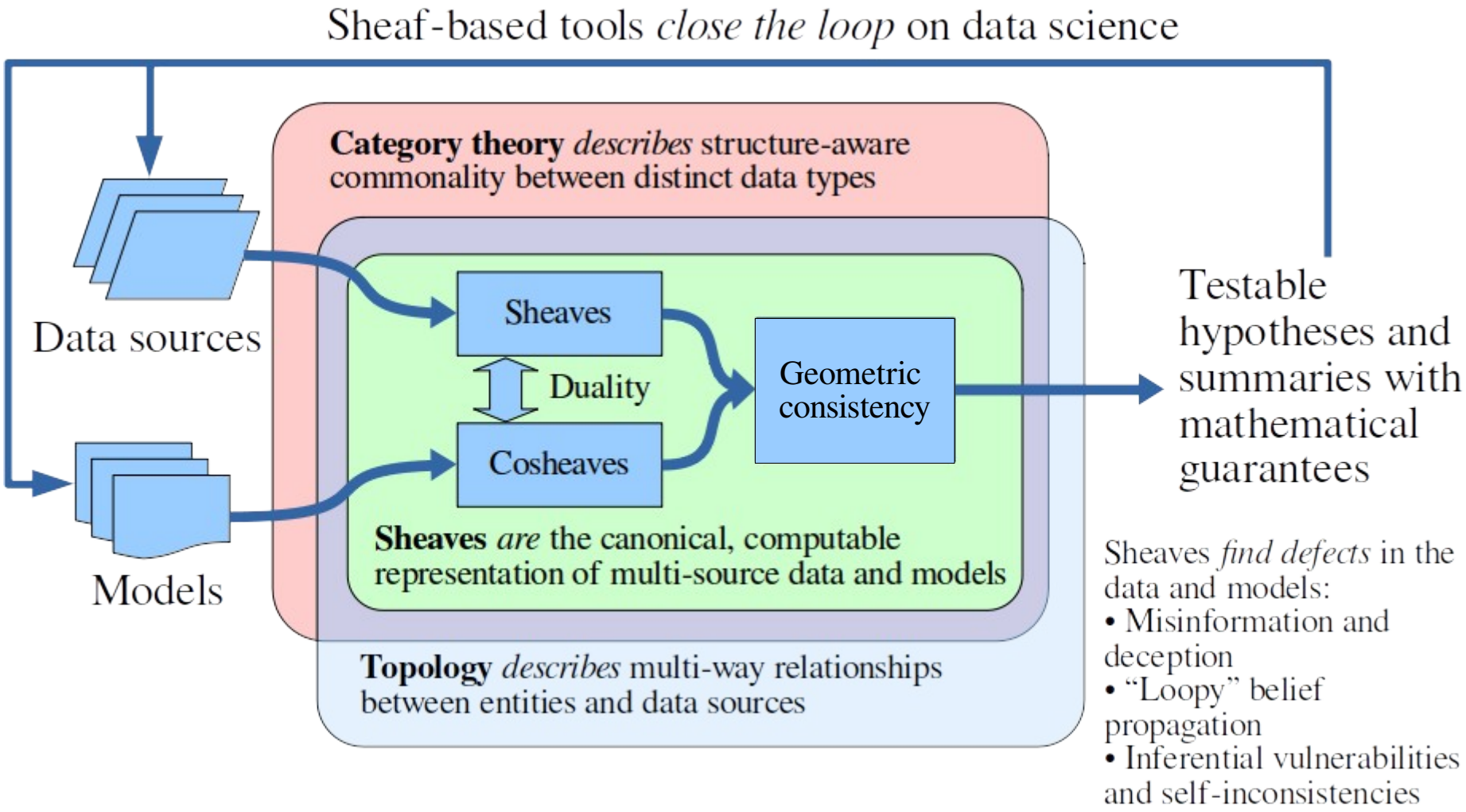






… so use mathematics suited to diagrams: *topology, sheaves,* and *categories*

… and the correct mathematical tools yield performance guarantees



Translating models into sheaves

Dynamic Structural Equation Model

SeaIce

ColdPool

Copepods

Krill

Diet_Cop

Diet_Krill

Spawners

Survival

Alaska pollock

Graphics from NOAA Fisheries

James Thorson Michael Szulczewski Michael Robinson

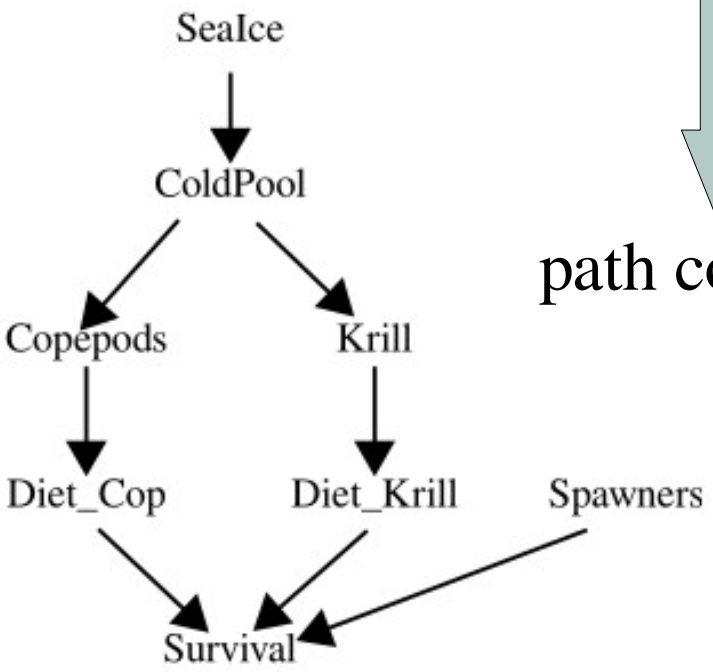
Dynamic Structural Equation Model

$$\frac{dx_k(\tau - t_\ell)}{d\tau} = \sum_{i=1}^{J}\sum_{j=1}^{T} \gamma_{k,\ell,i,j} x_i(\tau - t_j)$$

Time lags t_j not shown

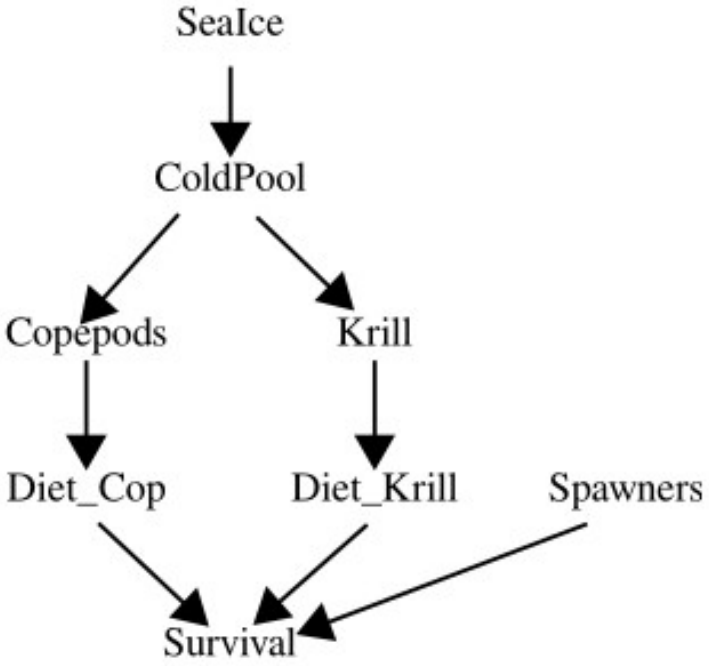
variables x_i are nodes

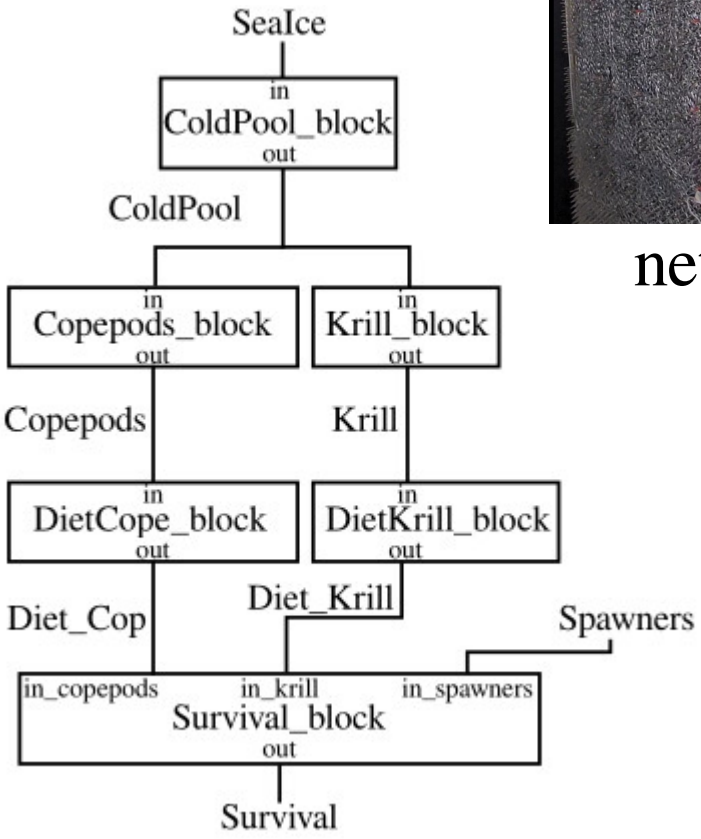
path coefficients γ are directed edges

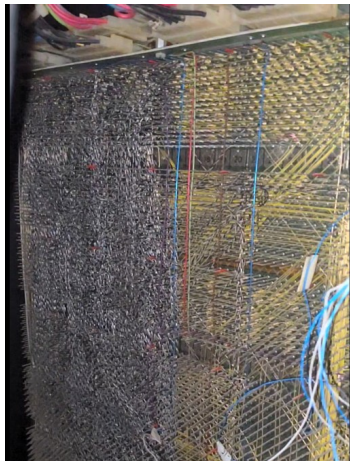


DSEM → netlist (circuit diagram)

$$\frac{dx_k(\tau - t_\ell)}{d\tau} = \sum_{i=1}^{J} \sum_{j=1}^{T} \gamma_{k,\ell,i,j} x_i(\tau - t_j)$$





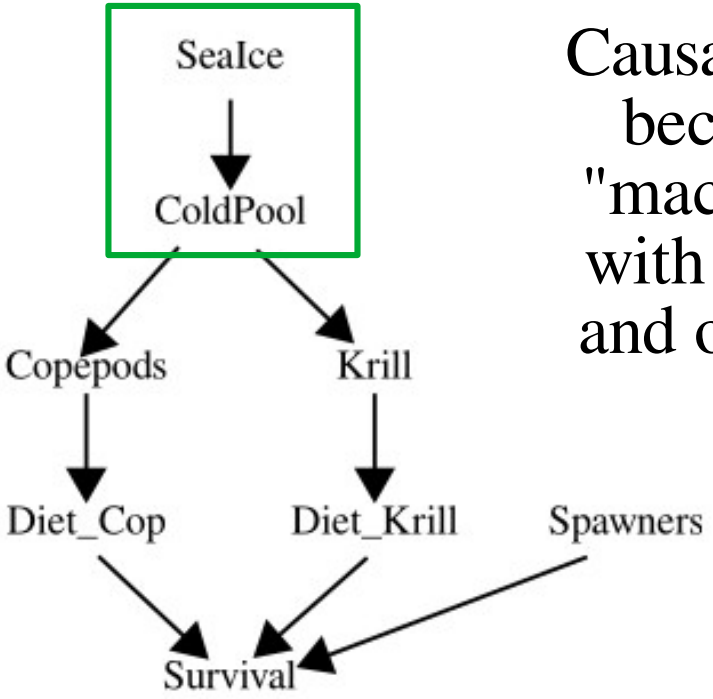


netlists!

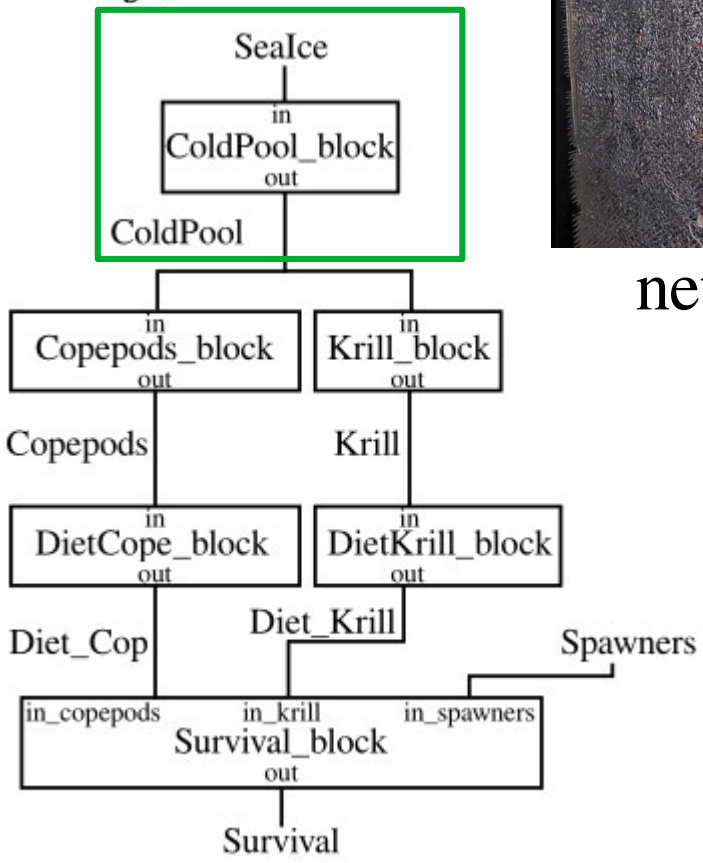
James Thorson Michael Szulczewski Michael Robinson

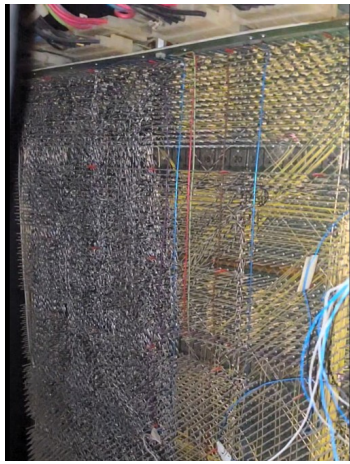
DSEM → netlist (circuit diagram)

$$\frac{dx_k(\tau - t_\ell)}{d\tau} = \sum_{i=1}^{J} \sum_{j=1}^{T} \gamma_{k,\ell,i,j} x_i(\tau - t_j)$$



Causal paths become "machines" with inputs and outputs

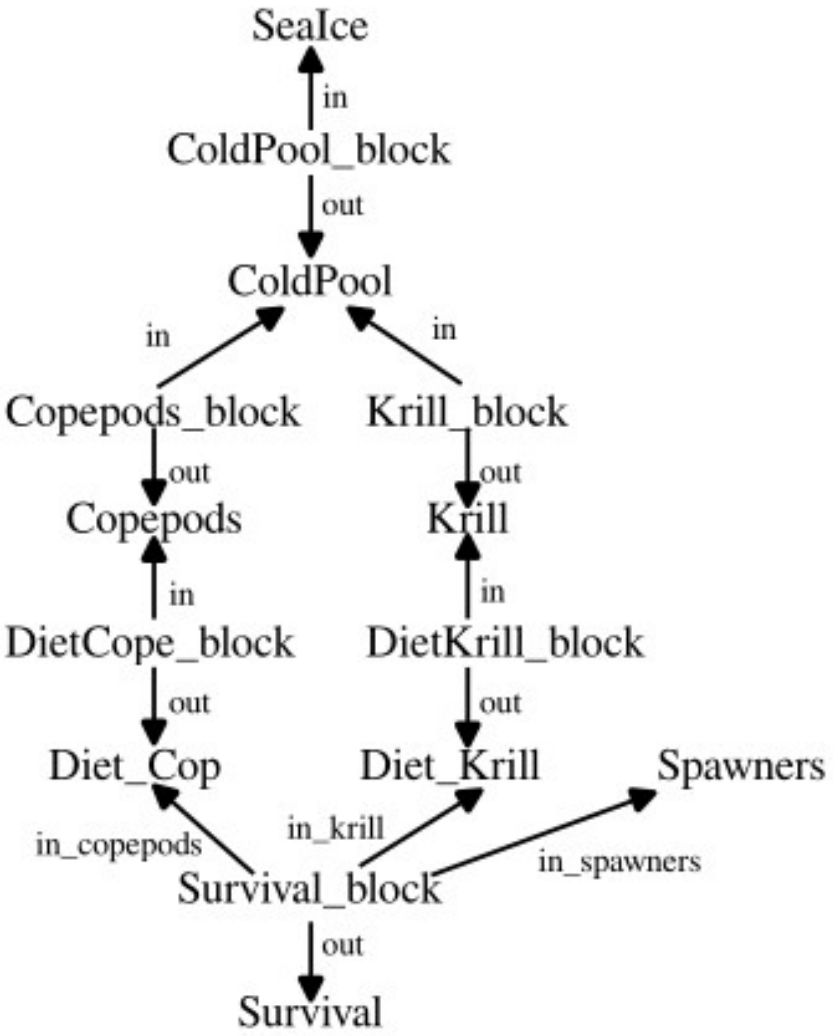


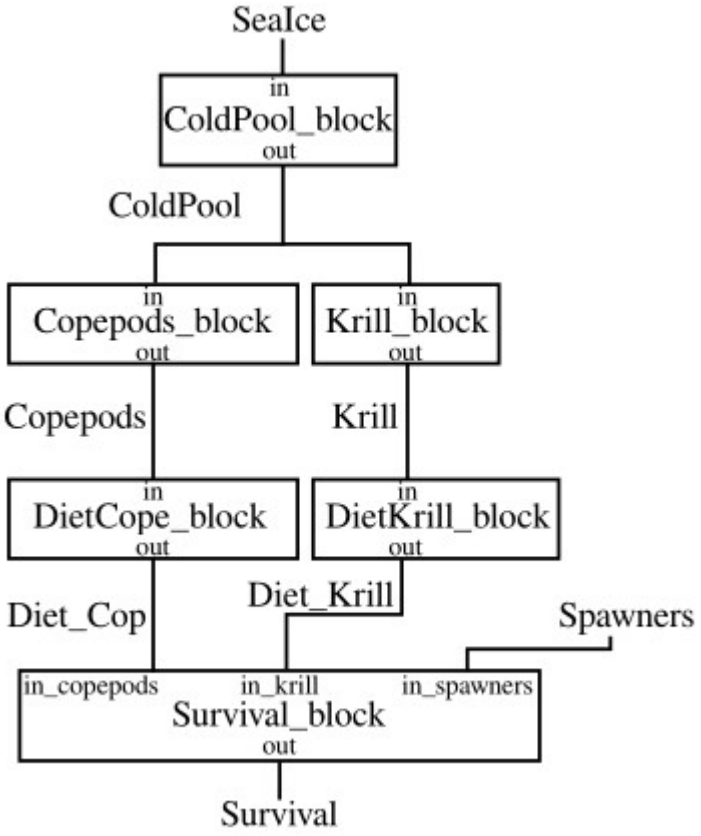


netlists!

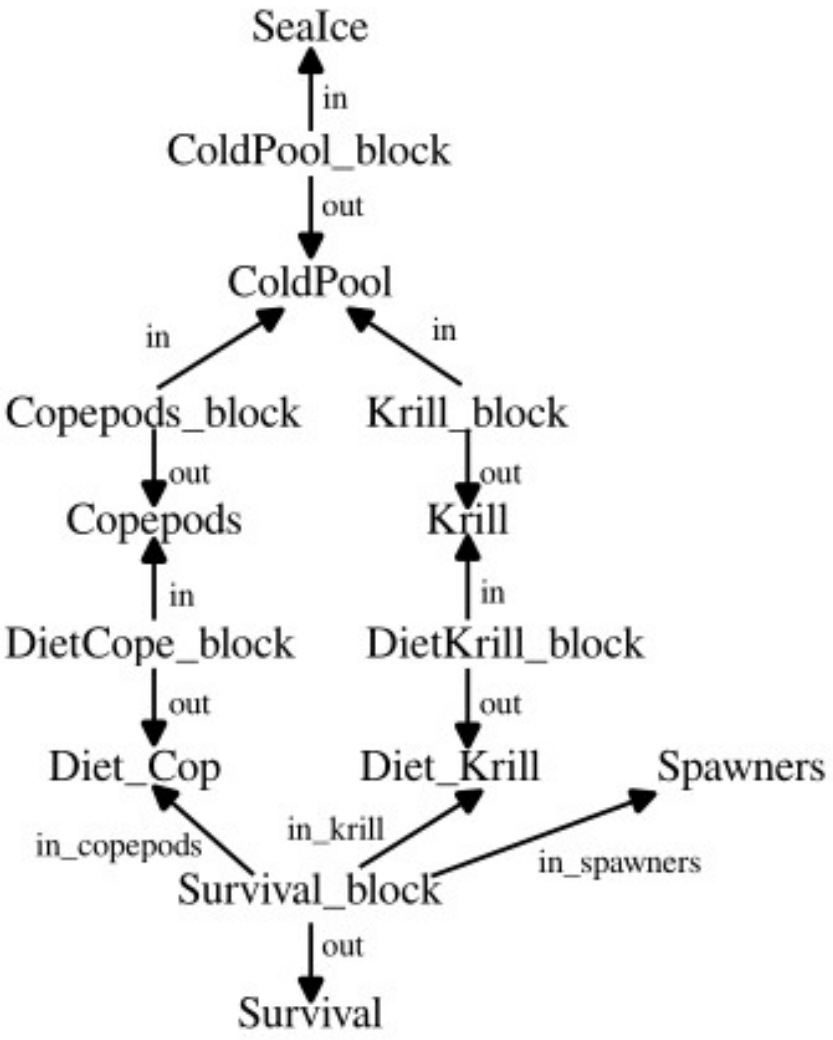
James Thorson Michael Szulczewski Michael Robinson

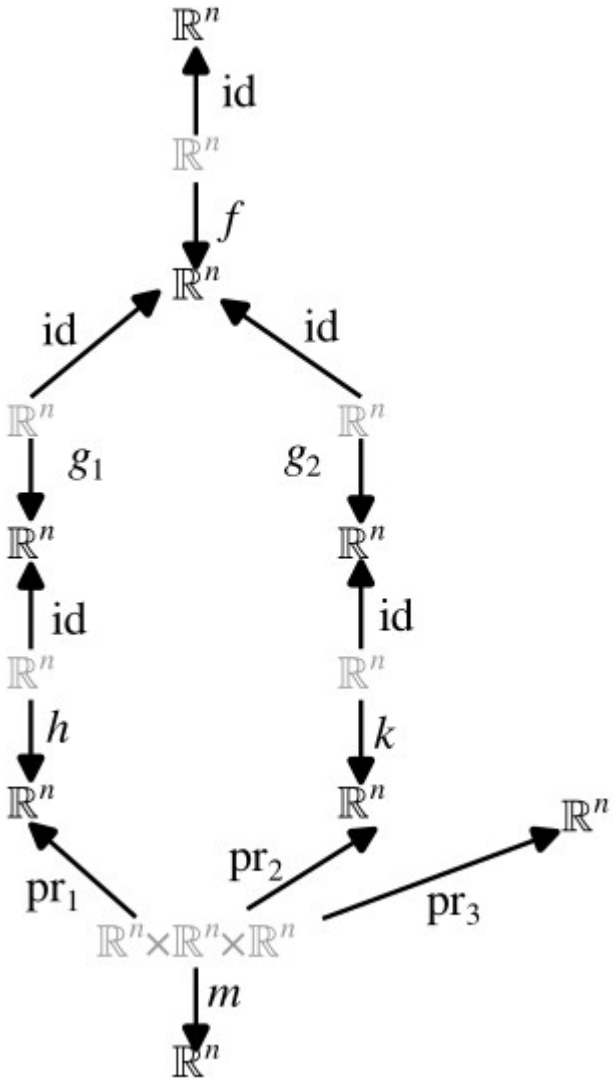
Netlist incidence graph



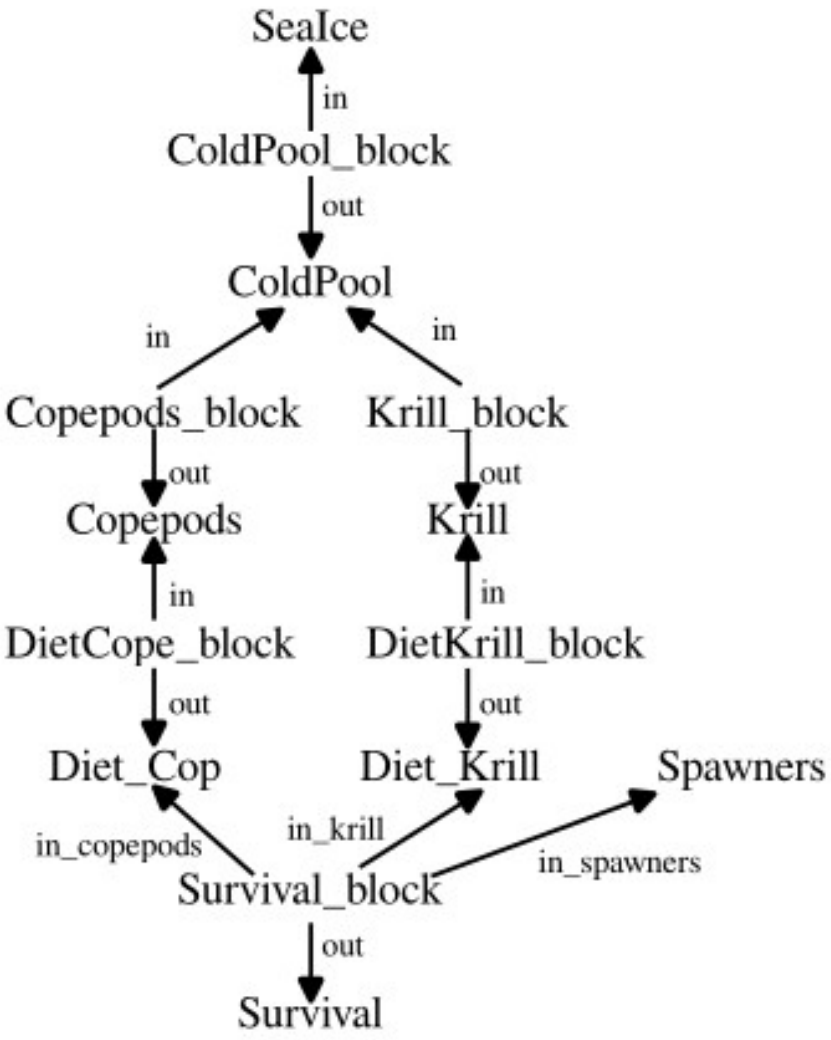


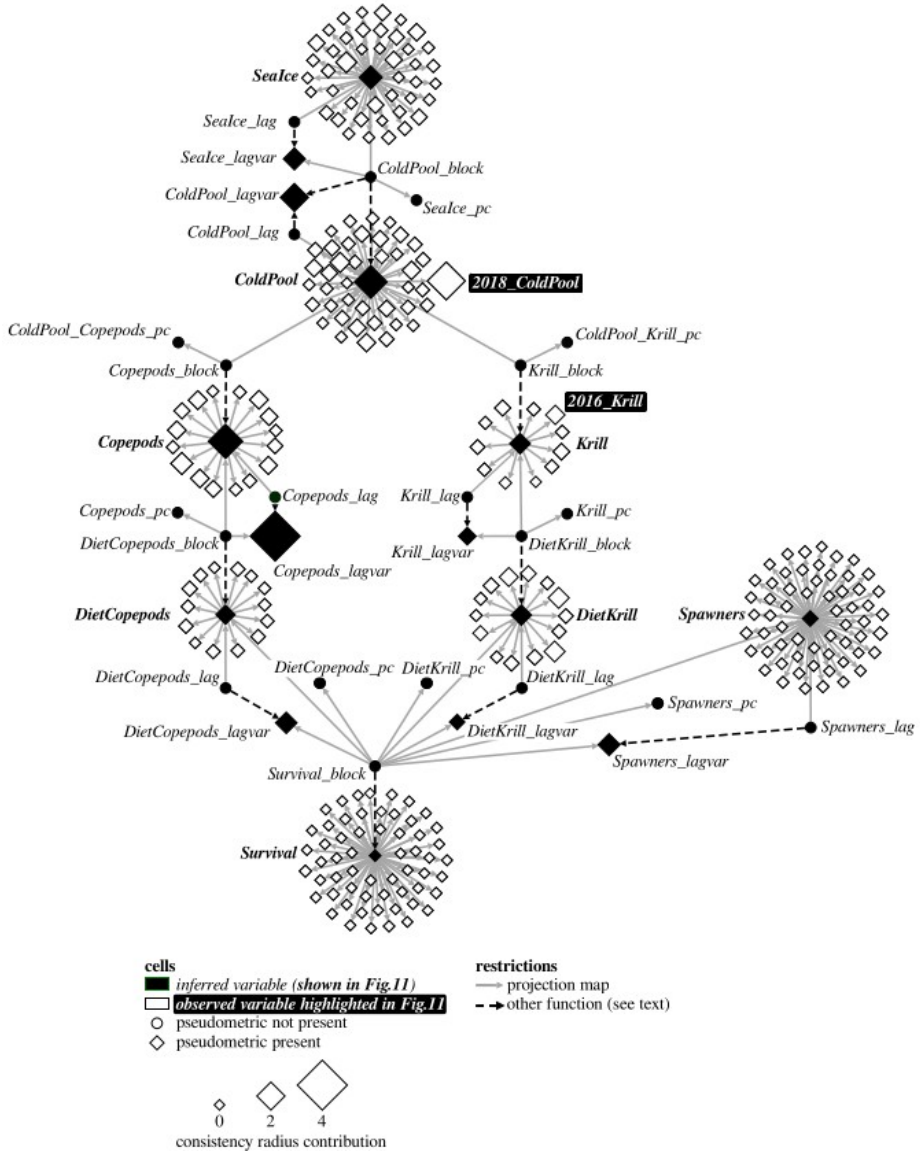
Incidence graph to sheaf





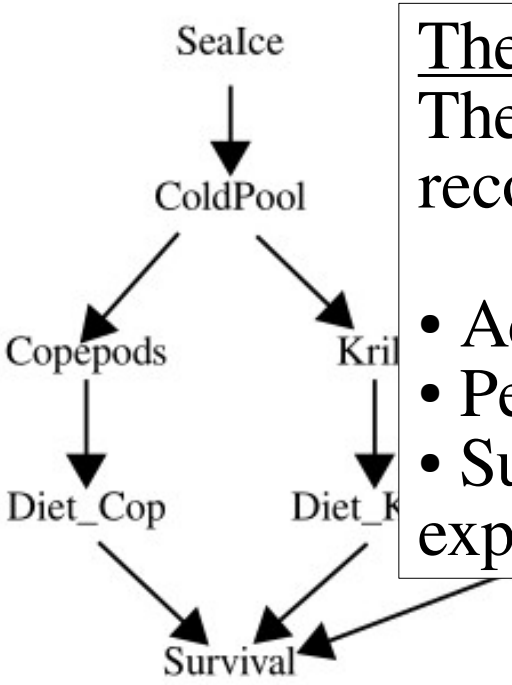
… expanded to handle missing data





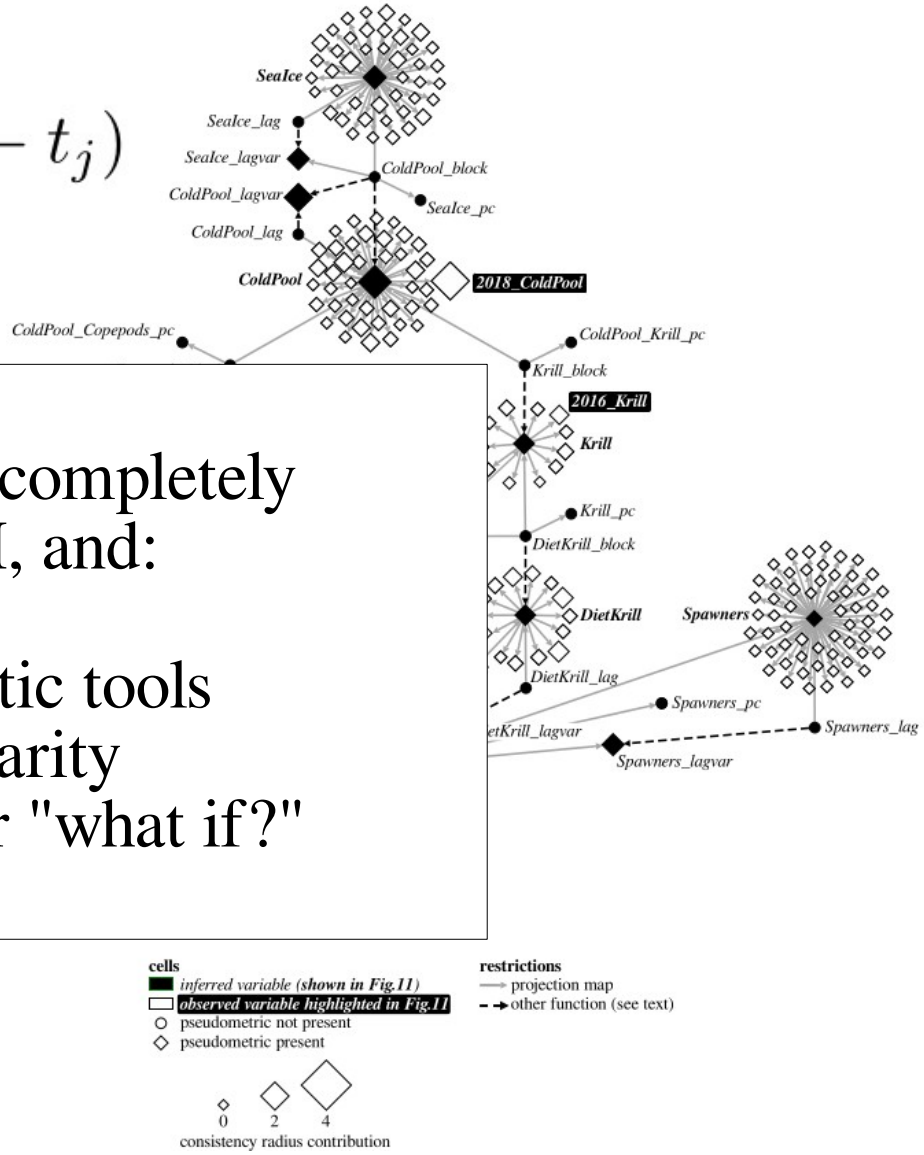
Sheaves generalize DSEM

$$\frac{dx_k(\tau - t_\ell)}{d\tau} = \sum_{i=1}^{J} \sum_{j=1}^{T} \gamma_{k,\ell,i,j} x_i(\tau - t_j)$$



Theorem:
The sheaf model completely recovers a DSEM, and:

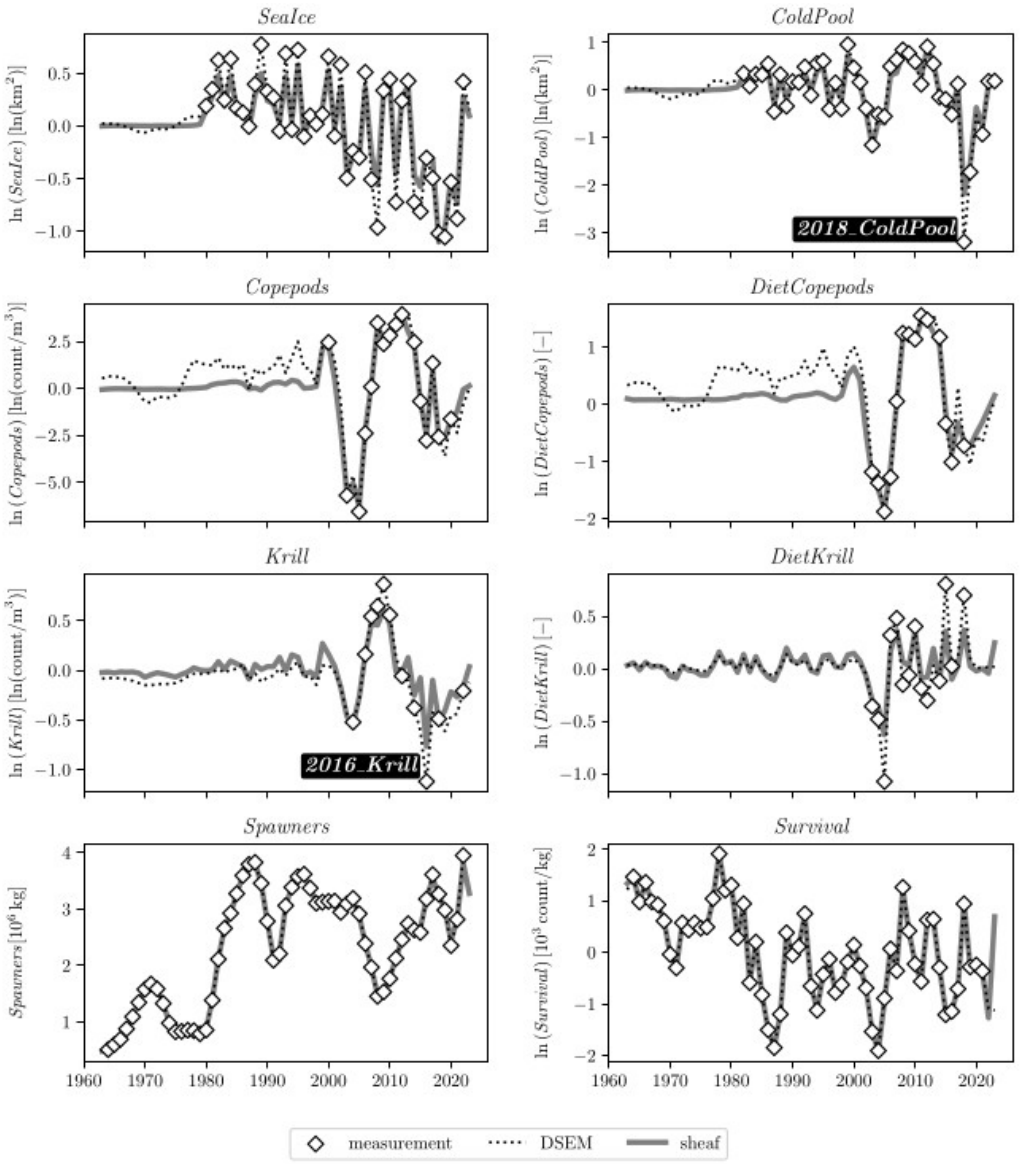
- Adds new analytic tools
- Permits nonlinearity
- Supports deeper "what if?" exploration



James Thorson Michael Szulczewski Michael Robinson

The sheaf recovers the DSEM

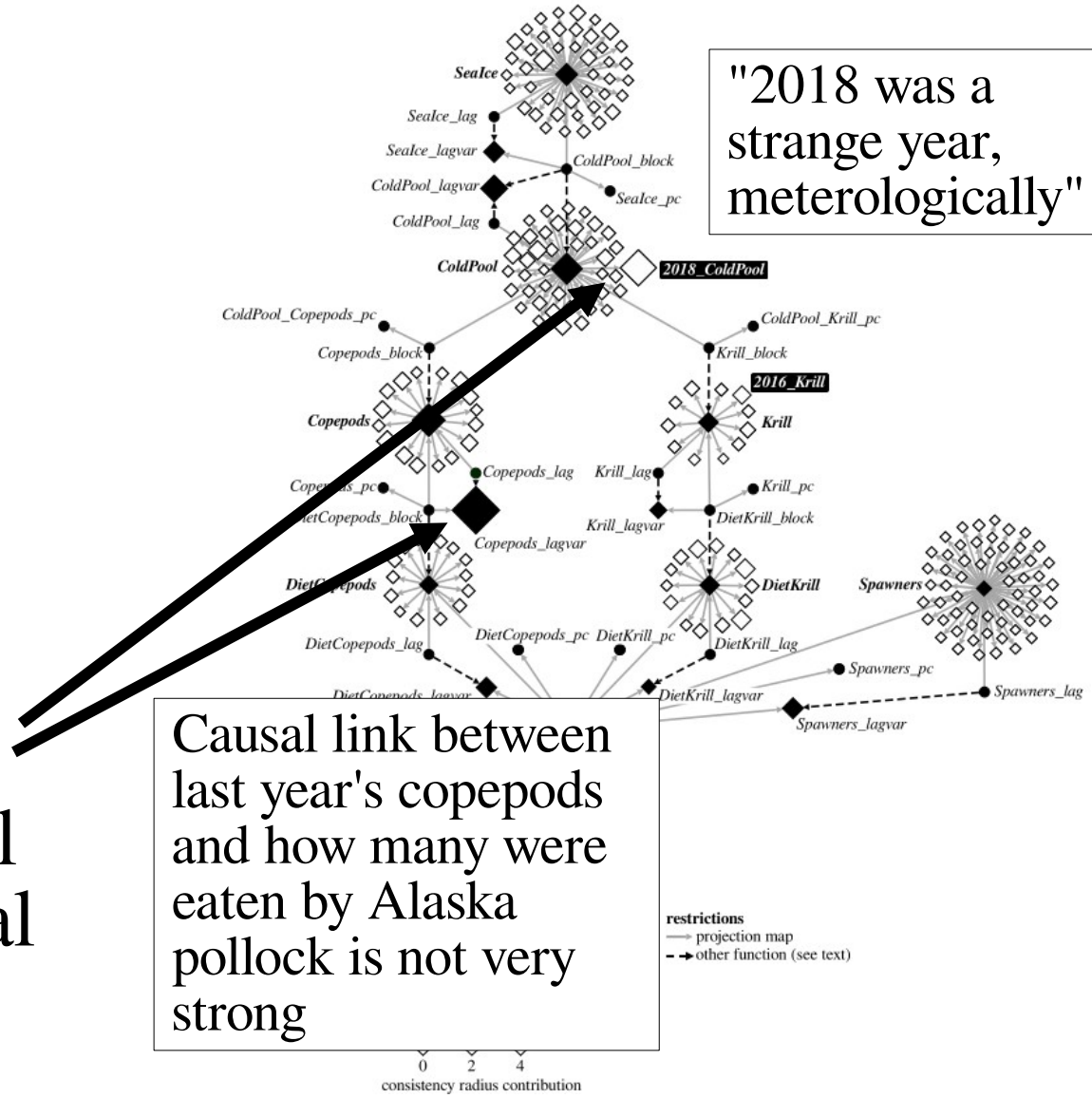
- Missing data are imputed "maximum consistency" instead of "maximum likelihood"
- Path coefficients can be inferred in the same way
- Local consistency radius allows unusual observations to causal links to be detected



James Thorson Michael Szulczewski Michael Robinson

The sheaf recovers the DSEM

- Missing data are imputed "maximum consistency" instead of "maximum likelihood"
- Path coefficients can be inferred in the same way
- Local consistency radius allows unusual observations to causal links to be detected



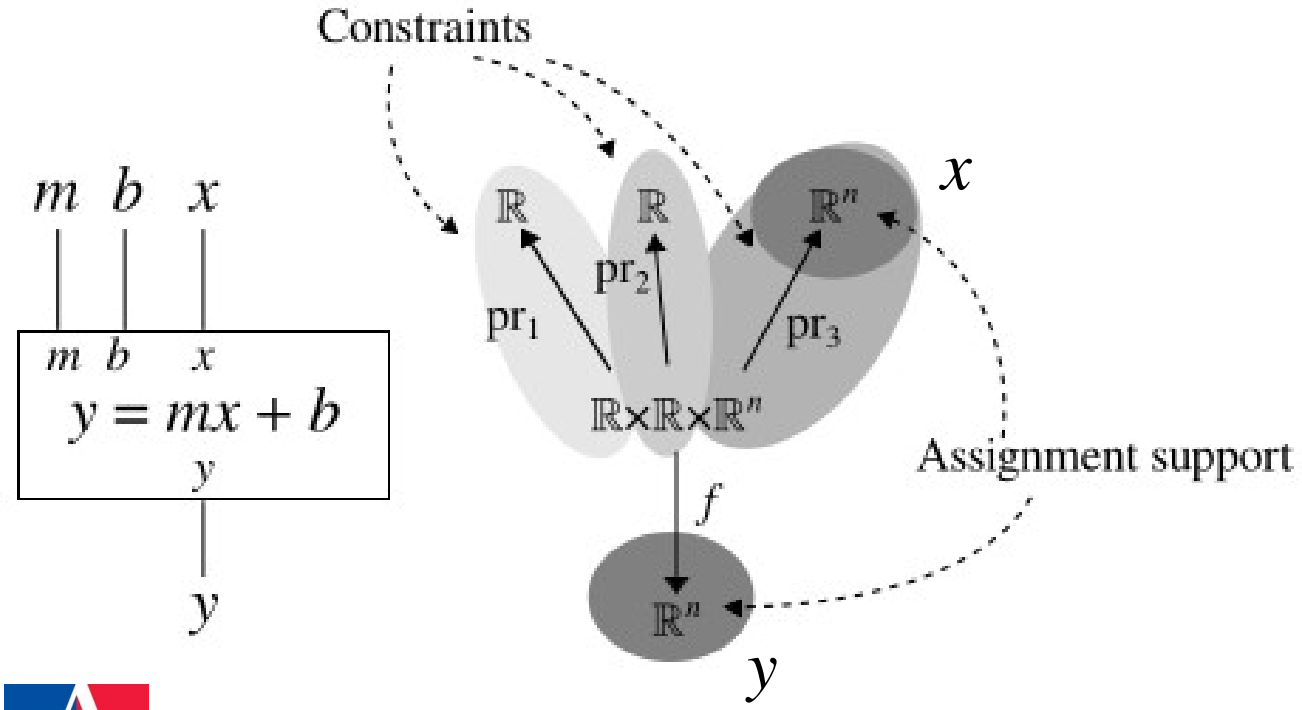
"2018 was a strange year, meterologically"

Causal link between last year's copepods and how many were eaten by Alaska pollock is not very strong

Sheaf analytics:
deeper math → better tools

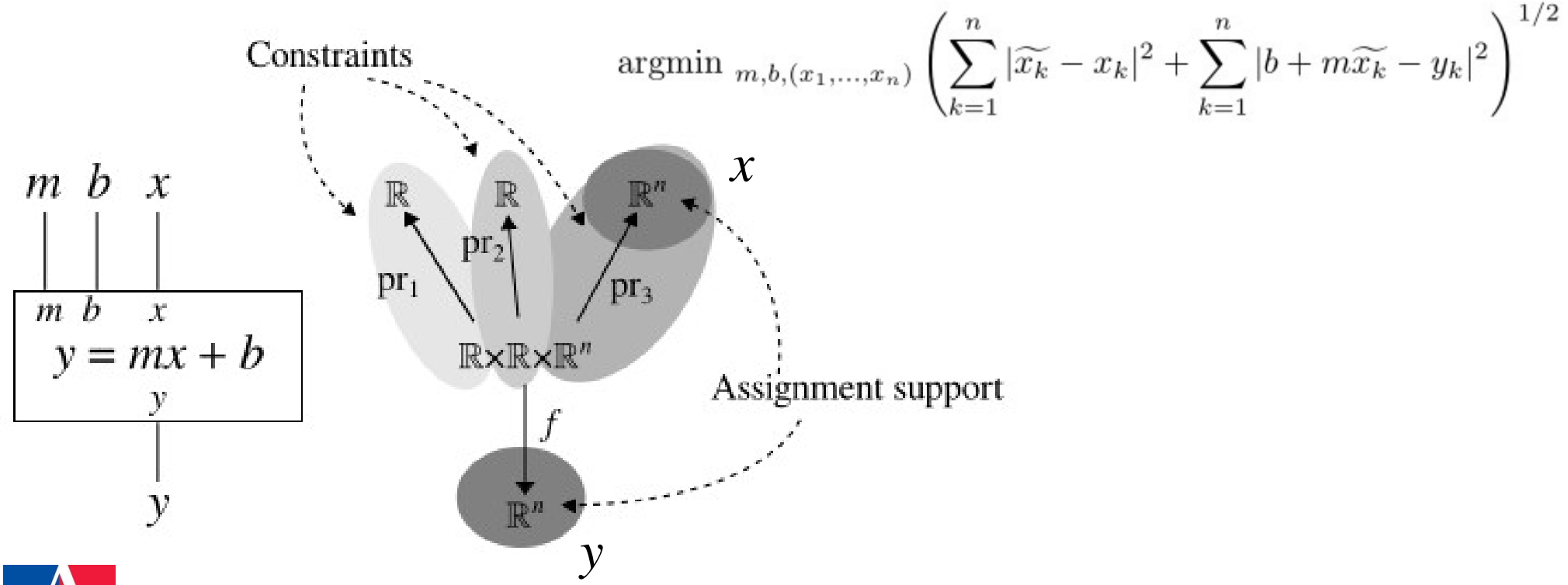
Linear regression as a sheaf

- Put observations on x and y (the assignment support)
- Constraints ensure the copies of parameters in the sheaf all agree



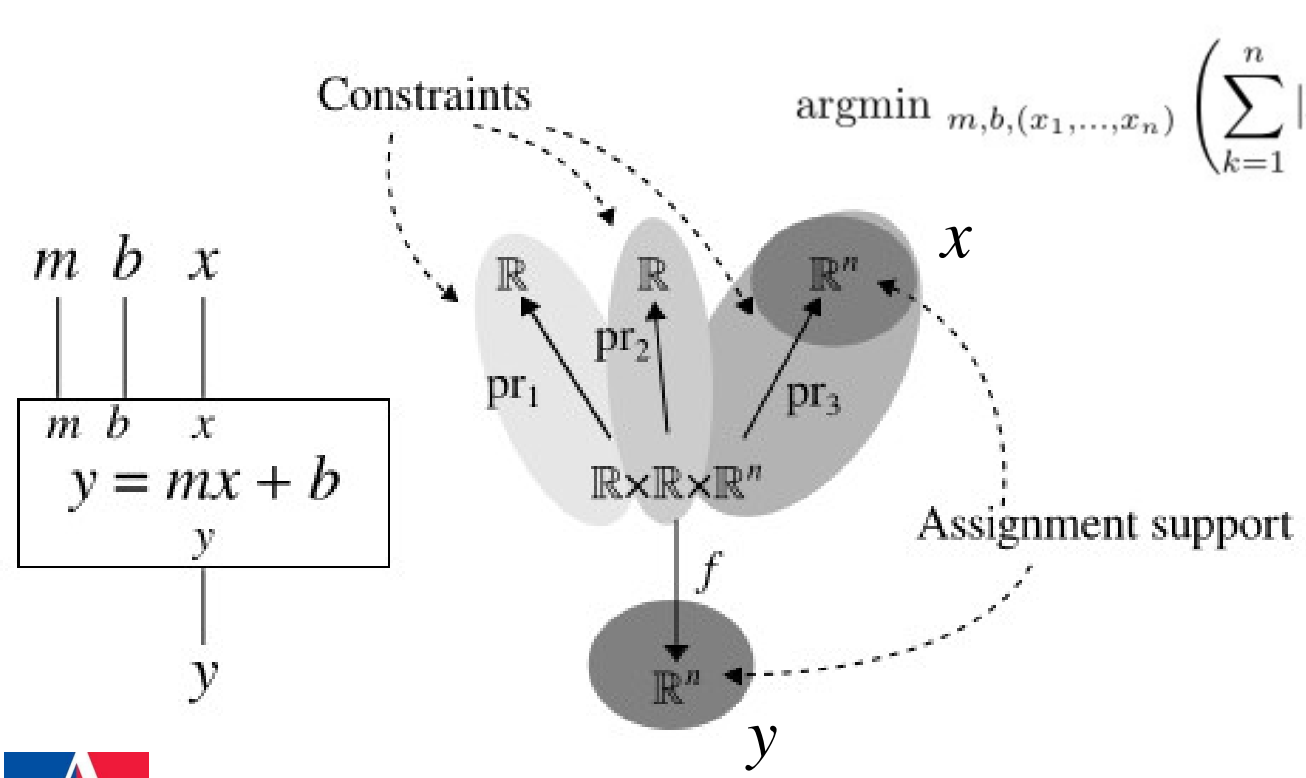
Consistency radius

- Put observations on x and y (the assignment support)
- What values of m and b will yield the least disagreement when applying the functions on the arrows?



Consistency radius

- Put observations on *x* and *y* (the assignment support)
- What values of *m* and *b* will yield the least disagreement when applying the functions on the arrows?

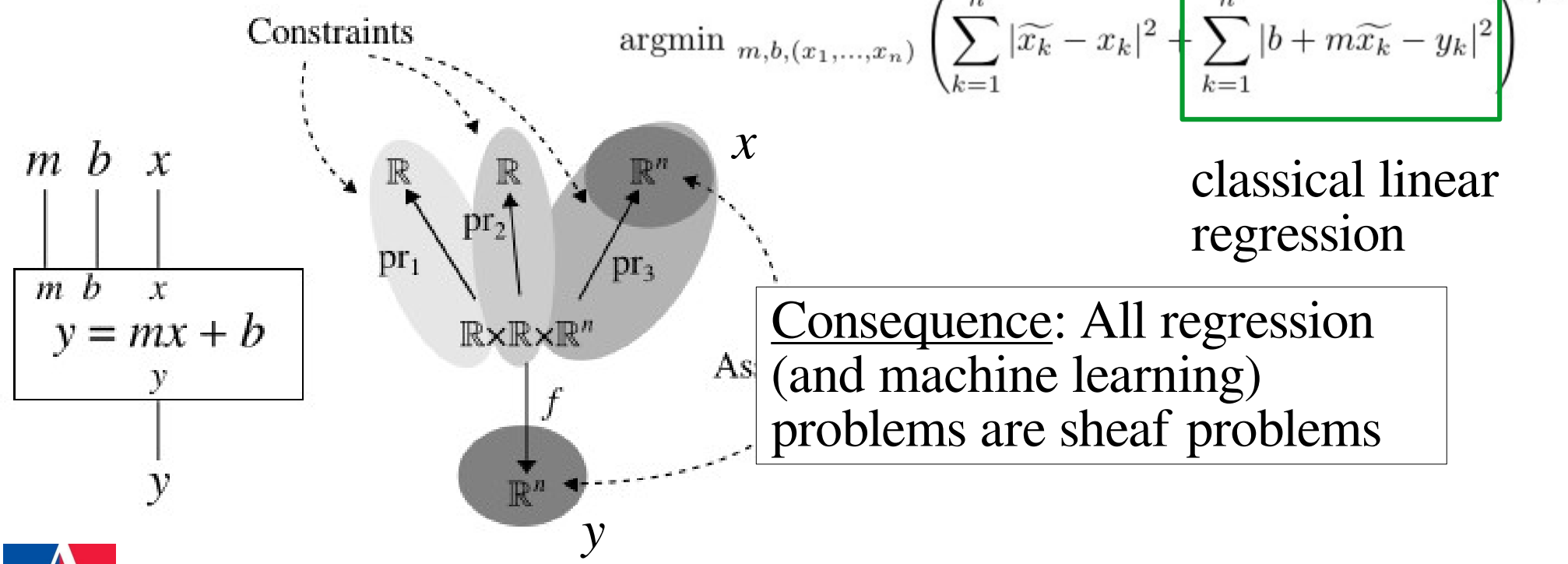


$$\operatorname{argmin}_{m,b,(x_1,\ldots,x_n)} \left(\sum_{k=1}^{n} |\widetilde{x_k} - x_k|^2 + \sum_{k=1}^{n} |b + m\widetilde{x_k} - y_k|^2 \right)^{1/2}$$

classical linear regression

Consistency radius

- Put observations on *x* and *y* (the assignment support)
- What values of *m* and *b* will yield the least disagreement when applying the functions on the arrows?



Sheaf modeling discipline

- **A *sheaf* is a hypothesis** about how a collection of variables interact
- **An *assignment* is a sample** of observations
- ***Consistency radius* is a test statistic** for the assignment being consistent with the sheaf
- ***Minimizing consistency radius* estimates values** of variables outside the sample

Next steps

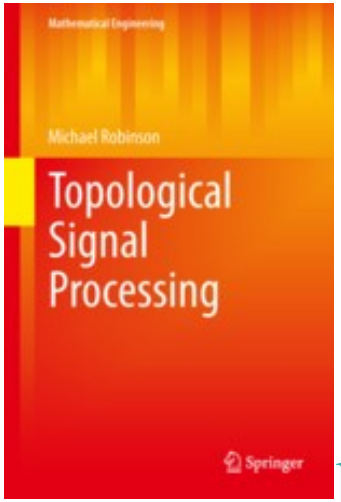
- Goal: Estimate the distribution of consistency radii under mild assumptions
  - Payoff: Formal hypothesis testing using sheaves
- Goal: Leverage topological acoustics models (found last year by our team) into sheaves
  - Payoff: Topological filters attuned to the acoustic environment
  - Payoff: New geometric/topological features may be highly robust to uncertainty, allowing operation in challenging environments

To learn more...

Michael Robinson

michaelr@american.edu

http://drmichaelrobinson.net



Relevant papers:

https://arxiv.org/abs/2511.04603

https://doi.org/10.1111/2041-210X.14289

Software:

https://github.com/kb1dds