Unraveling conditions associated with cyanobacteria blooms in Missisquoi Bay, Lake Champlain, USA, using a modified Self-Organizing Map

Andrea R. Pearce[*1,2], Mary C. Watzin[2,3], Gregory K. Druschel[3,4], Donna M. Rizzo[1]

(1) UVM School of Engineering
(2) UVM Rubenstein School of Environment and Natural Resources
(3) NCSU College of Natural Resources
(4) UVM Geology
(5) Indiana University-Purdue University Indianapolis

Bloom of the cyanobacteria Anabaena and Microcystis in Lake Sallie, Minnesota, U.S.A. Photo from old ASLO slide collection taken by D.F. Brakke, scanned by K.L. Schulz.

A.R. Pearce, M.C. Watzin, G.K. Druschel, D.M. Rizzo. (In Preparation), Unraveling conditions associated with cyanobacteria blooms in Missisquoi Bay, Lake Champlain, USA, using a modified Self-Organizing Map. *Environmental Science and Technology*

Unraveling conditions associated with cyanobacteria blooms in Missisquoi Bay, Lake Champlain, USA, using a modified Self-Organizing Map

Acknowledgements:

VT EPSCoR Grant NSF EPS #0701410
Kristen Hallock
Alison Pechenik
VT EPSCoR CSYS group
Rubenstein Lab Crew

Bloom of the cyanobacteria Anabaena and Microcystis in Lake Sallie, Minnesota, U.S.A. Photo from old ASLO slide collection taken by D.F. Brakke, scanned by K.L. Schulz.

Lake Champlain experiences cyanobacteria blooms

https://sharepoint.uvm.edu/sites/isi2008/Participant%20Photos

Human influence…

NABS (www.benthos.org)

http://www.dnr.state.oh.us/Home/species_a_to_z/SpeciesGuideIndex/whiteperch

?

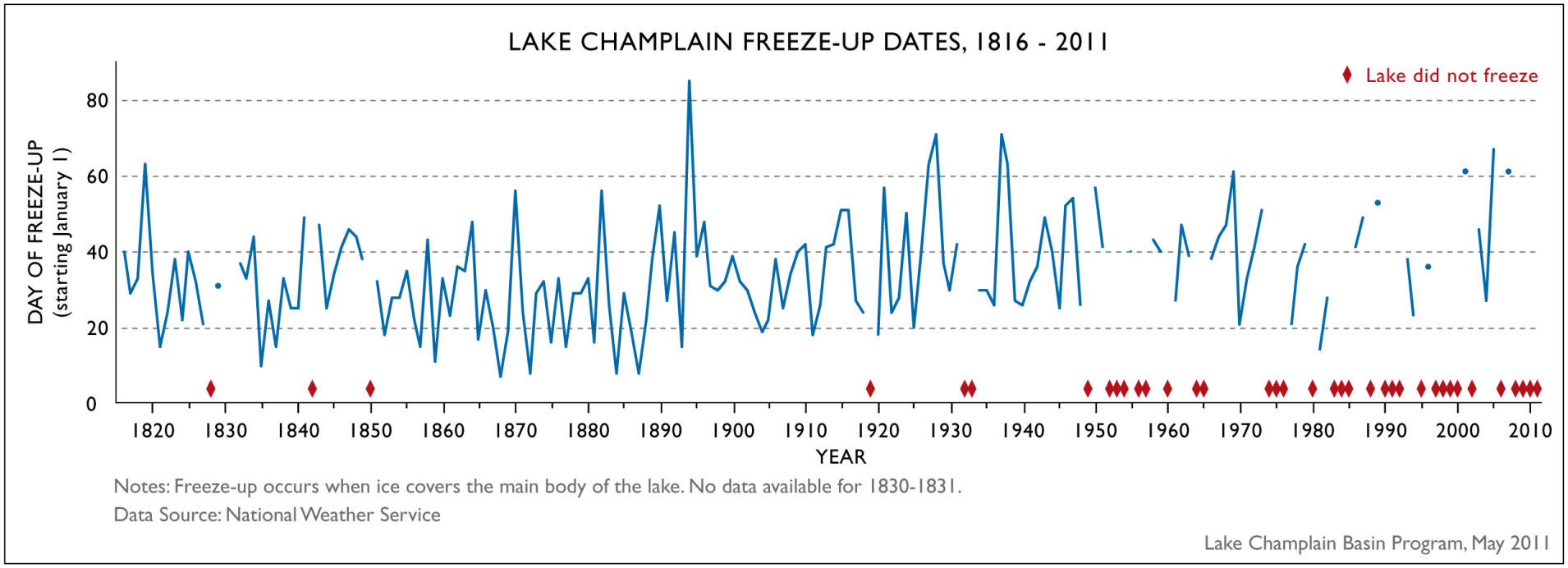
http://extension.usu.edu/waterquality/images/uploads/AG-WQ/manure/applyingmanure.jp1.jpg

http://www.mychamplain.net/threats-explained

http://www.ehow.com/how_8586143_fix-outboard-motor-fuel-filters.html

http://biosantee.com/pharmaceuticals

Human influence…

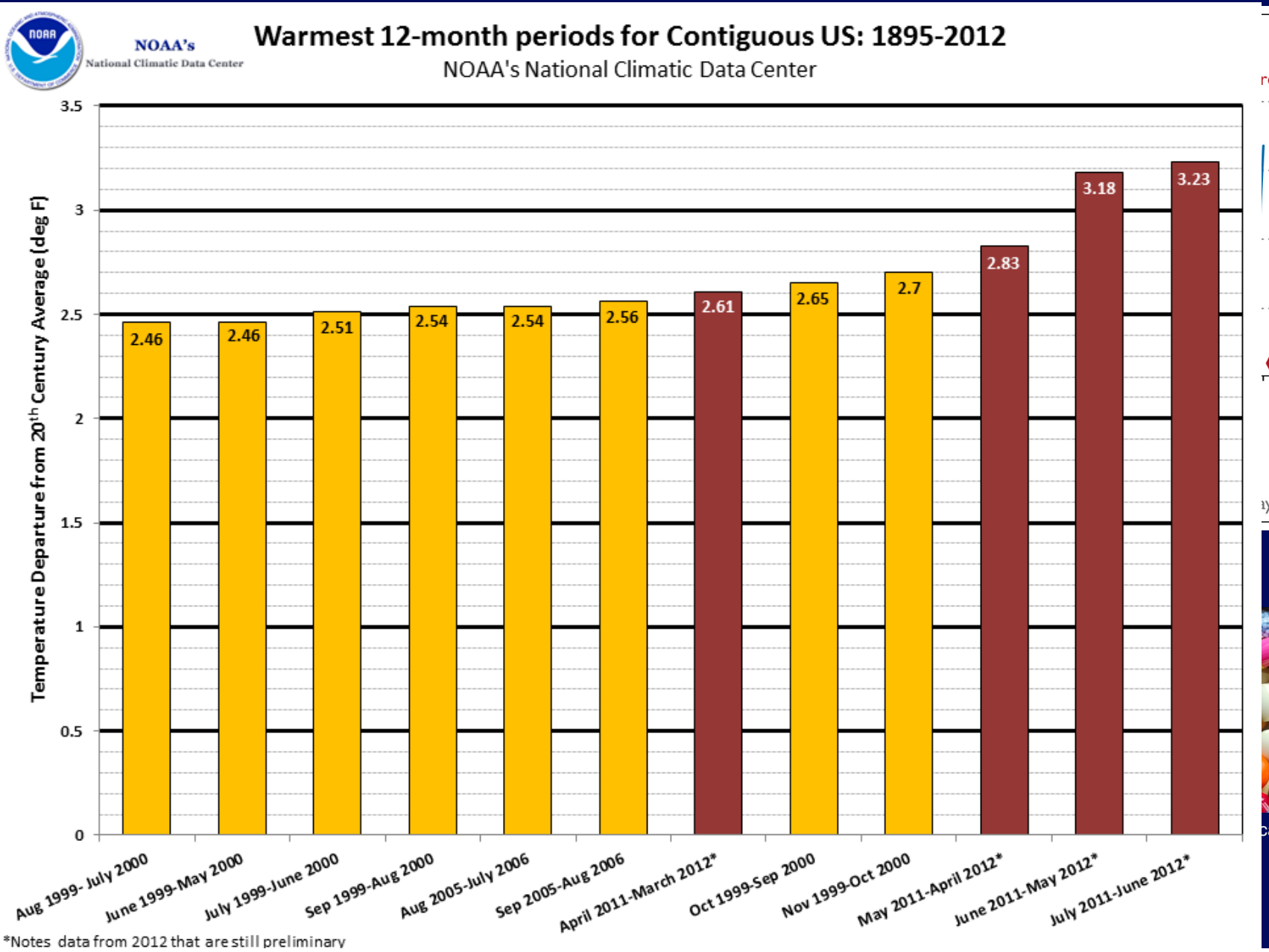


http://www.mychamplain.net/threats-explained

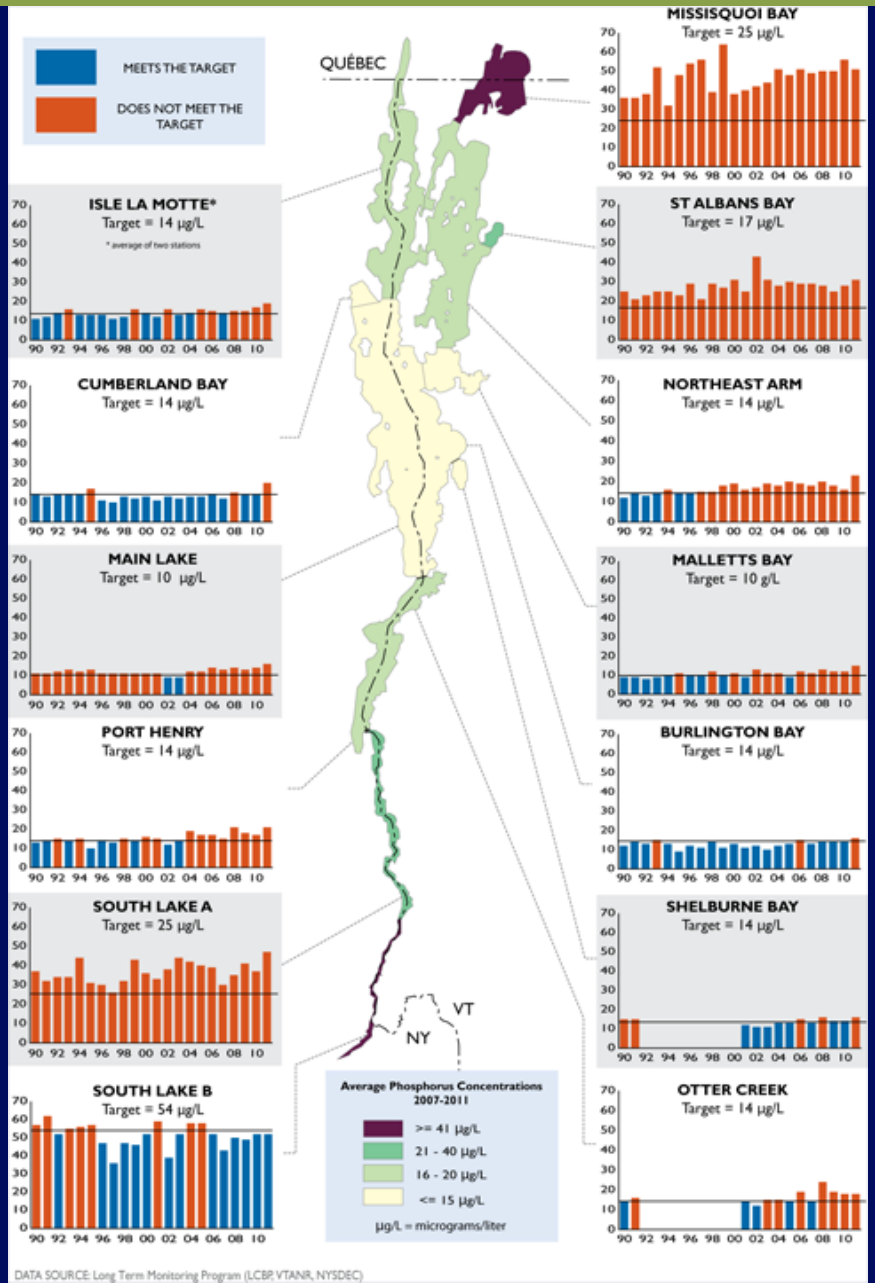
http://www.ehow.com/how_8586143_fix-outboard-motor-fuel-filters.html


http://biosantee.com/pharmaceuticals

Human influence…



Missisquoi Bay

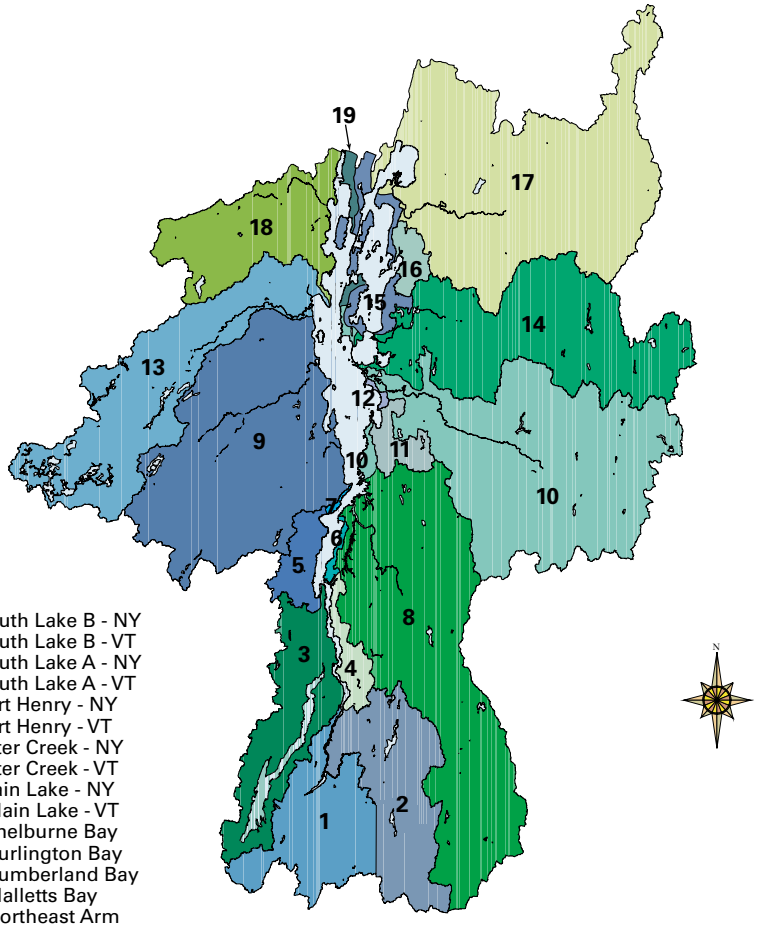


2012 State of the Lake Report

Missisquoi Bay

THE LAKE CHAMPLAIN BASIN ATLAS

Watersheds by Lake Segments

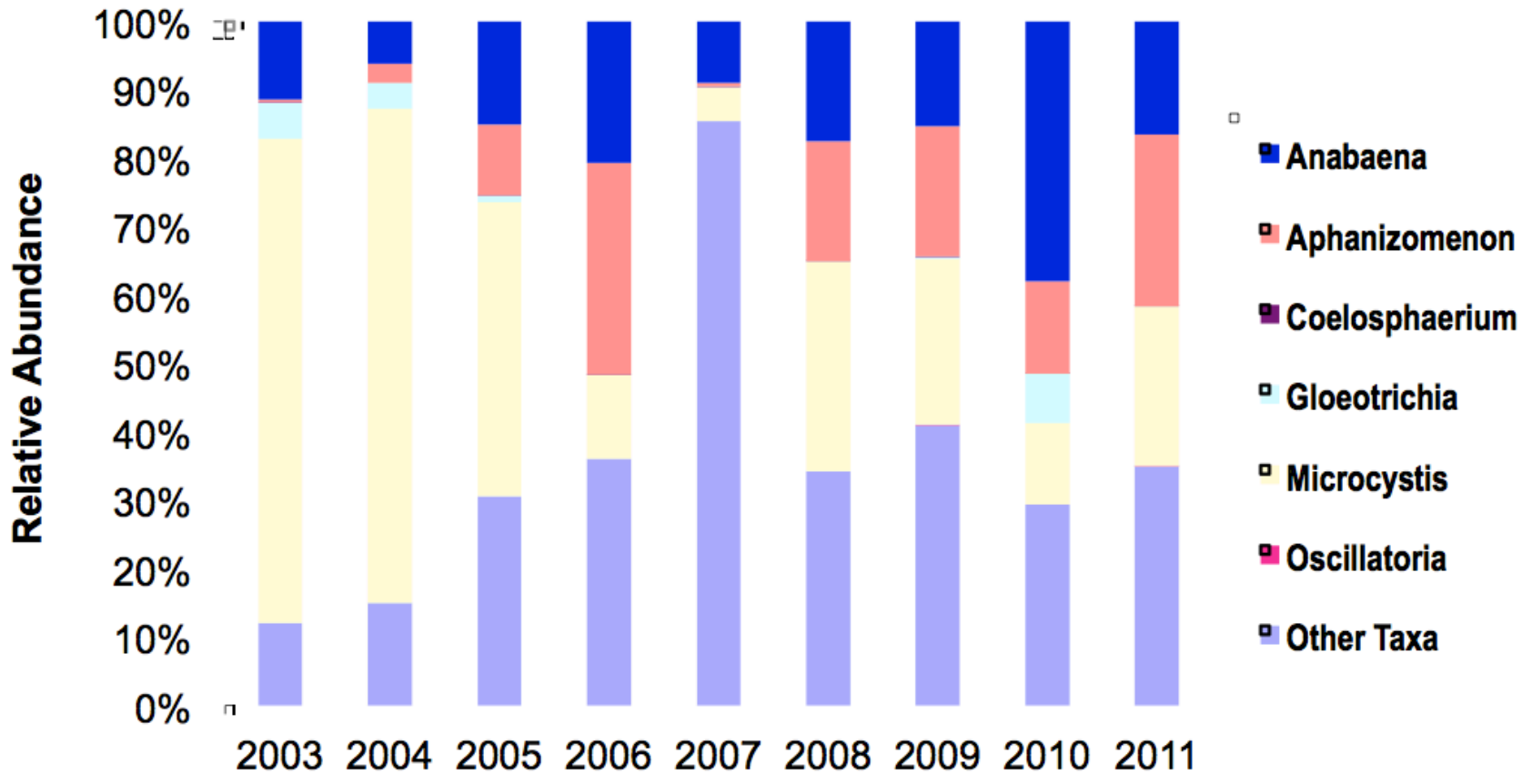


http://www.lcbp.org/atlas/index.htm

Map by Northern Cartographic

Mean depth = 2.8 m (max ~ 4m)
WS:Basin area = 40:1

Cyanobacteria



Seasonal mean percent generic composition of phytoplankton in Missisquoi Bay
(LCBP Technical Report #71 - Monitoring and Evaluation of Cyanobacteria in Lake Champlain)

www.noaanews.noaa.gov

Microcystis

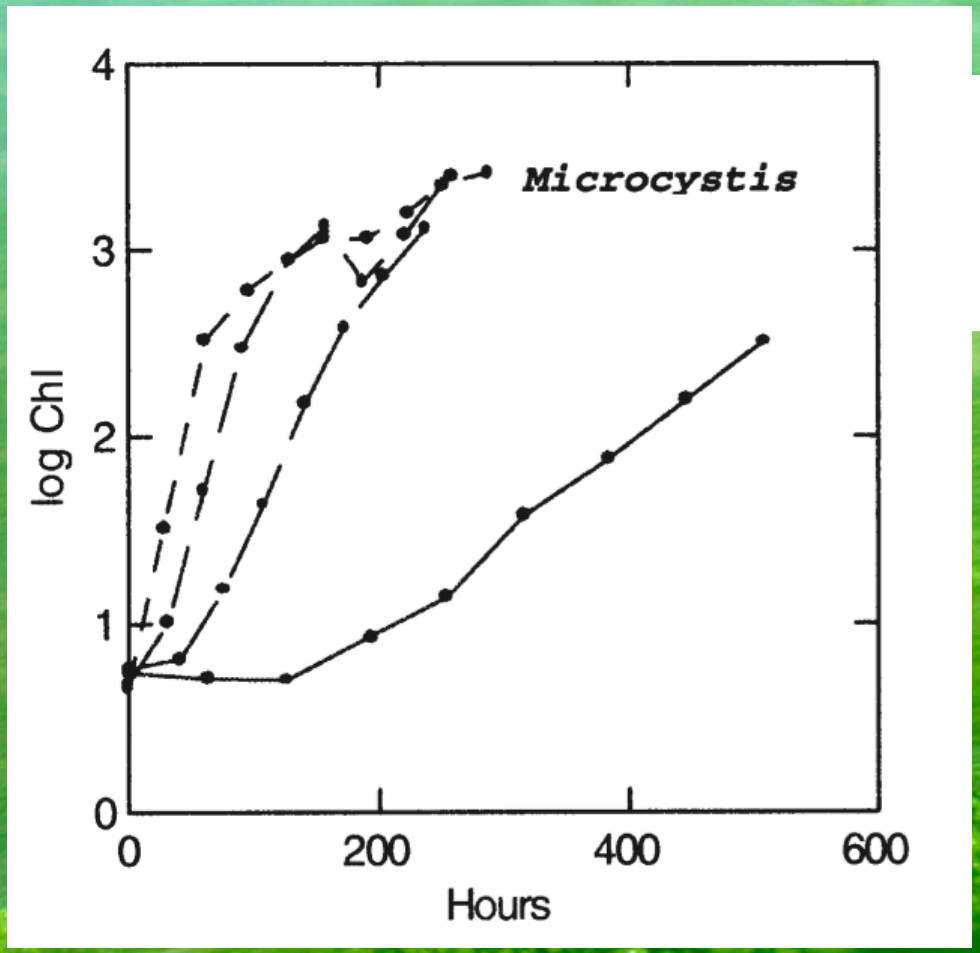
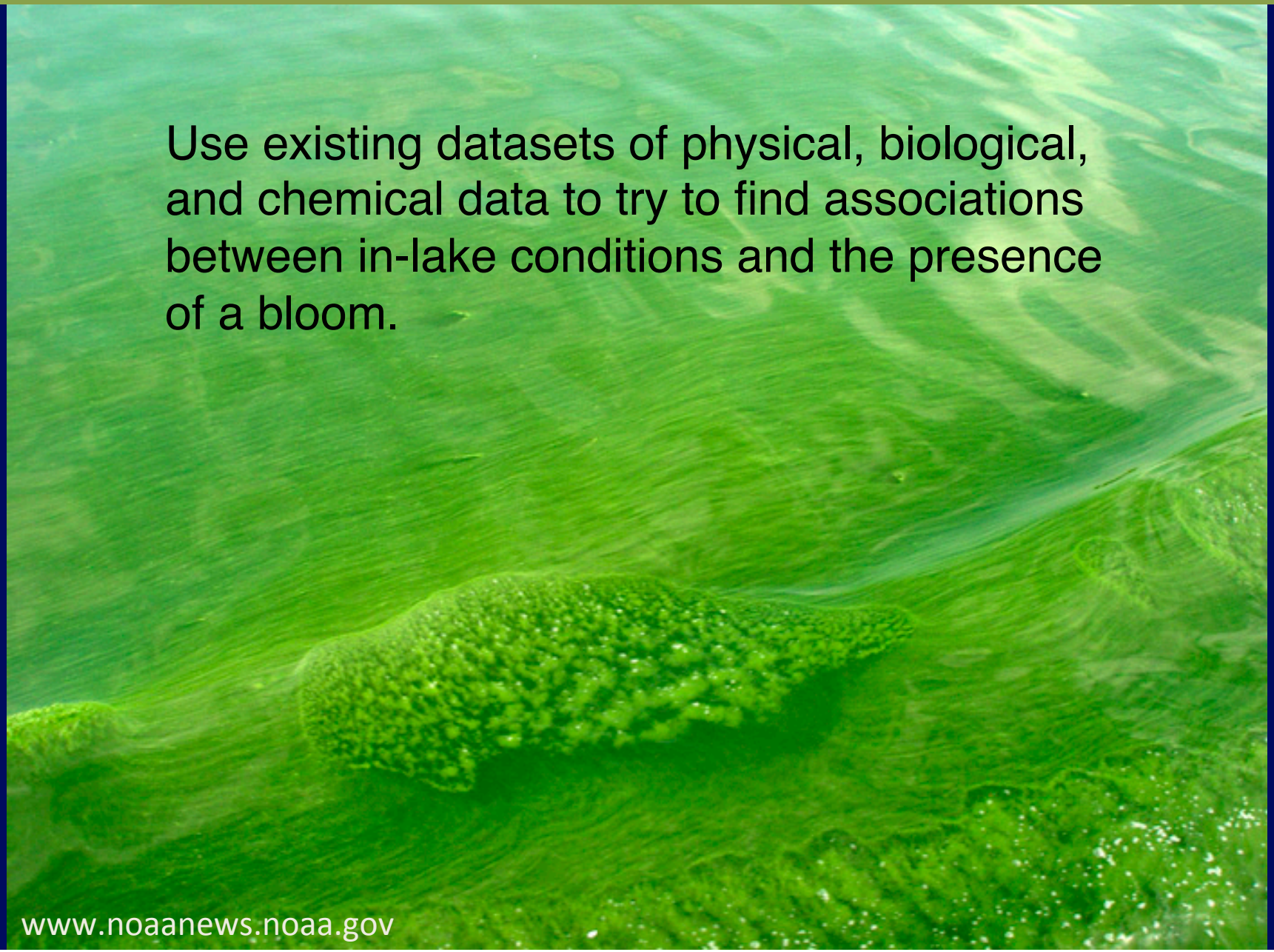


FIG. 2. Growth curves of $\log_{10}$ Chl *a* concentration ($\mu g \cdot L^{-1}$) at 15° C (solid lines), 20° C (long dashes), 25° C (medium dashes), and 30° C (short dashes). Hours represent the cumulative hours that cultures were subjected to light.

Coles and Jones (2000, Journal of Phycology)

www.noaanews.noaa.gov

Project Objective



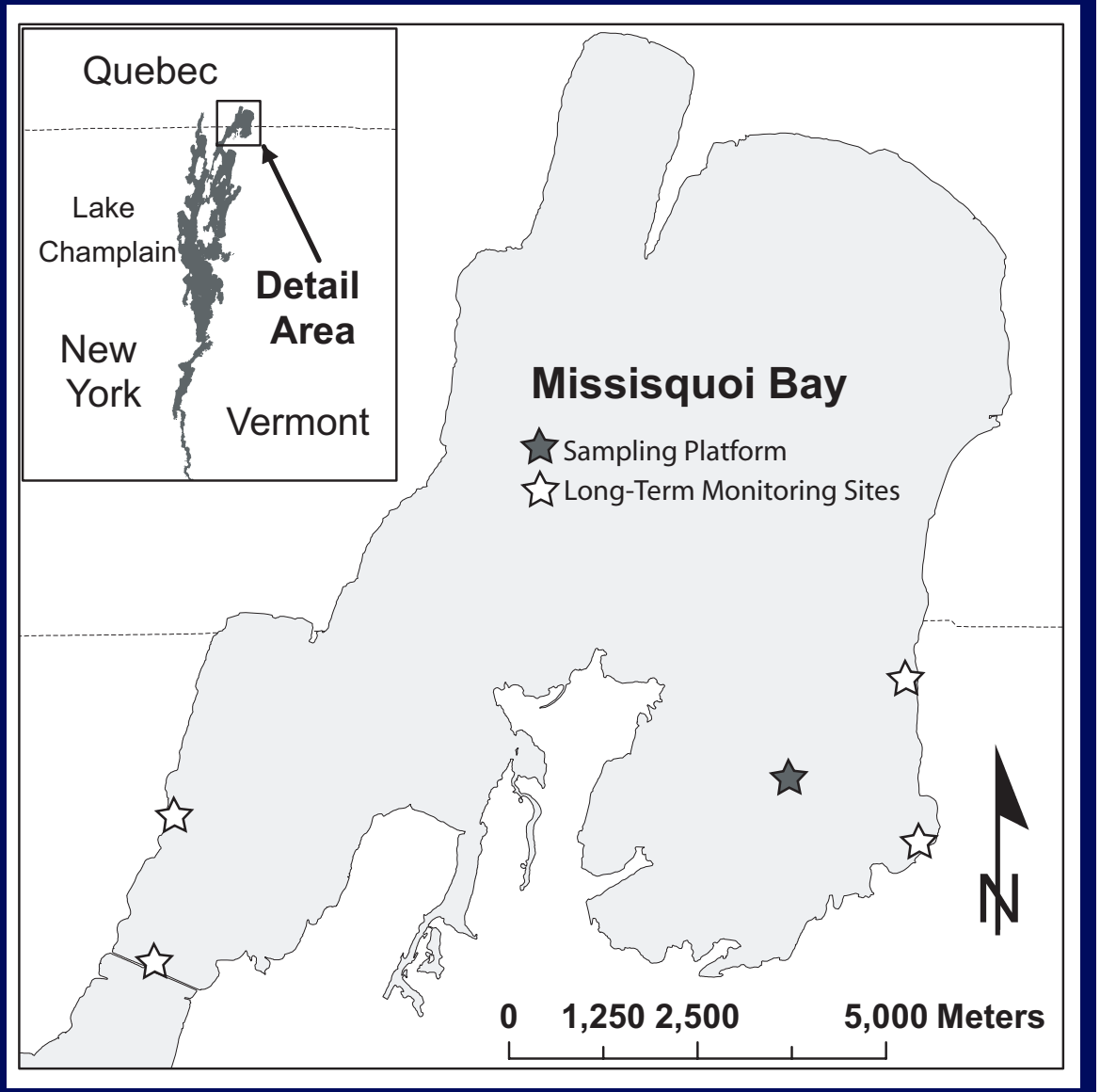
Use existing datasets of physical, biological, and chemical data to try to find associations between in-lake conditions and the presence of a bloom.

www.noaanews.noaa.gov

The Datasets

Emerging Threats Project

- 1 Sampling location
- 2007, 2008, 2009
- Multiple depths (5)
- Physical, Chemical and Biological Measurements
- Concurrent sediment analyses



The Datasets

Emerging Threats Project

- 1 Sampling location
- 2007, 2008, 2009
- Multiple depths (5)
- Physical, Chemical and Biological Measurements
- Concurrent sediment analyses

	Sampling Day and Time (EDT)						Sampling Depth (m)				
Day	6 h	10 h	12 h	13 h	18 h	19 h	0.0	0.6	1.5	2.4	3.5
12 Jun 03			X				X	X	X	X	X
9 Jul 03		X					X	X	X	X	X
13 Aug 03		X					X	X	X	X	X
17 Sep 03	X		X				X	X	X	X	X
24 Jun 04		X			X		X		X		X
28 Jul 04				X	X		X		X		X
29 Jul 04	X						X	X	X	X	X
26 Aug 04				X		X	X	X	X	X	X
27 Aug 04	X						X	X	X	X	
30 Sep 04			X		X		X		X		X
10 Aug 05			X			X	X	X	X	X	X
11 Aug 05	X						X	X	X	X	X

The Datasets

	Parameter	Units	Minimum	Mean	Maximum	Standard Deviation
Nutrients	Total Phosphorus *	µg/L	16.7	49	266	43.5
	Total Nitrogen *	mg/L	0.327	0.579	0.997	0.144
	Dissolved Nitrogen *	mg/L	0.29	0.445	0.997	0.15
	Soluble Reactive Phosphorus *	µg/L	0.5	5.99	33.26	5.46
-	Microcystin	µg/L	0.002	2.62	18.5	4.23
Phytoplankton	Bacillariophyceae *	cells/mL	0	306	1694	402
	Chlorophyceae *	cells/mL	23	898	3455	886
	Anabaena *	cells/mL	0	2649	17563	3878
	Aphanizomenon *	cells/mL	0	1027	9919	1960
	Microcystis *	cells/mL	0	25381	319804	58425
Physical Parameters	Temperature *	°C	16.45	21.58	24.44	2.4
	Conductivity	µS/cm	82.15	115.7	127.87	11.09
	Dissolved Oxygen	mg/L	5.98	8.67	11.13	1.25
	PAR (Irradiance)	W/m^2	0.025	212.4	1170	365.6
	Fluorescence	mg/m^3	0.98	7.78	18.86	3.4
	Turbidity	FTU	3.03	10.94	84.01	12.79
Sediment	Dissolved Oxygen *	µM	0	47.4	150	43.8
	Oxic Boundary***	mm	-2.5	-0.66	0.5	0.94
	Mn(II) Redox Boundary***	mm	-26	-1.18	10	7.57

* Also measured as part of the long-term monitoring dataset

** Measured at the sediment water interface by voltametric microelectrode

*** Relative to the sediment-water interface

The Datasets

Nutrients

Green Algae & Diatoms

Physical Parameters

Sediment Redox Chemistry

	Parameter	Units	Minimum	Mean	Maximum	Standard Deviation
Nutrients	Total Phosphorus *	µg/L	16.7	49	266	43.5
	Total Nitrogen *	mg/L	0.327	0.579	0.997	0.144
	Dissolved Nitrogen *	mg/L	0.29	0.445	0.997	0.15
	Soluble Reactive Phosphorus *	µg/L	0.5	5.99	33.26	5.46
	Microcystin	µg/L	0.002	2.62	18.5	4.23
Phytoplankton	Bacillariophyceae *	cells/mL	0	306	1694	402
	Chlorophyceae *	cells/mL	23	898	3455	886
	Anabaena *	cells/mL	0	2649	17563	3878
	Aphanizomenon *	cells/mL	0	1027	9919	1960
	Microcystis *	cells/mL	0	25381	319804	58425
Physical Parameters	Temperature *	°C	16.45	21.58	24.44	2.4
	Conductivity	µS/cm	82.15	115.7	127.87	11.09
	Dissolved Oxygen	mg/L	5.98	8.67	11.13	1.25
	PAR (Irradiance)	W/m^2	0.025	212.4	1170	365.6
	Fluorescence	mg/m^3	0.98	7.78	18.86	3.4
	Turbidity	FTU	3.03	10.94	84.01	12.79
Sediment	Dissolved Oxygen **	µM	0	47.4	150	43.8
	Oxic Boundary***	mm	-2.5	-0.66	0.5	0.94
	Mn(II) Redox Boundary***	mm	-26	-1.18	10	7.57

Toxin

Cyanobacteria

* Also measured as part of the long-term monitoring dataset

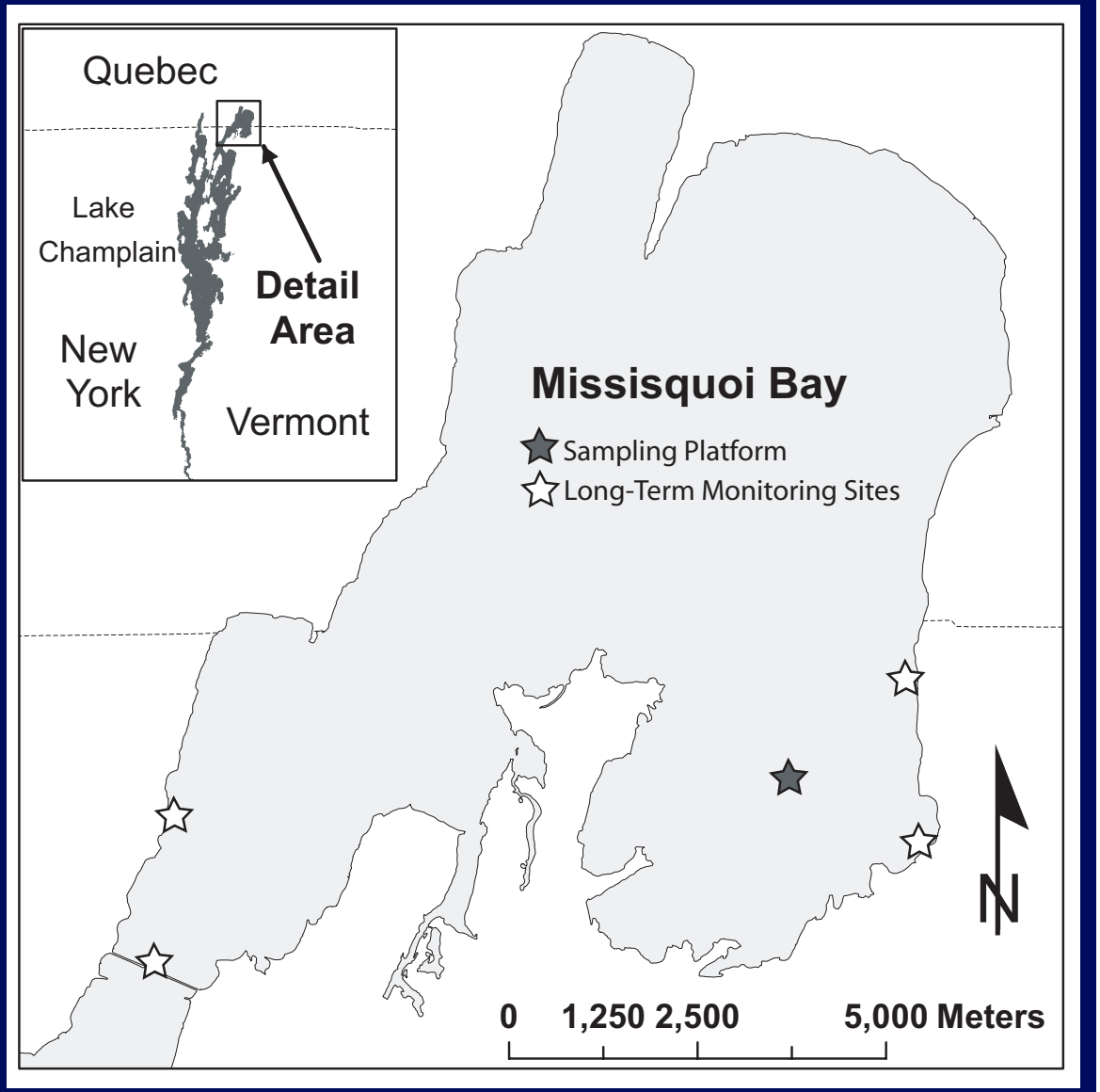
** Measured at the sediment water interface by voltametric microelectrode

*** Relative to the sediment-water interface

The Datasets

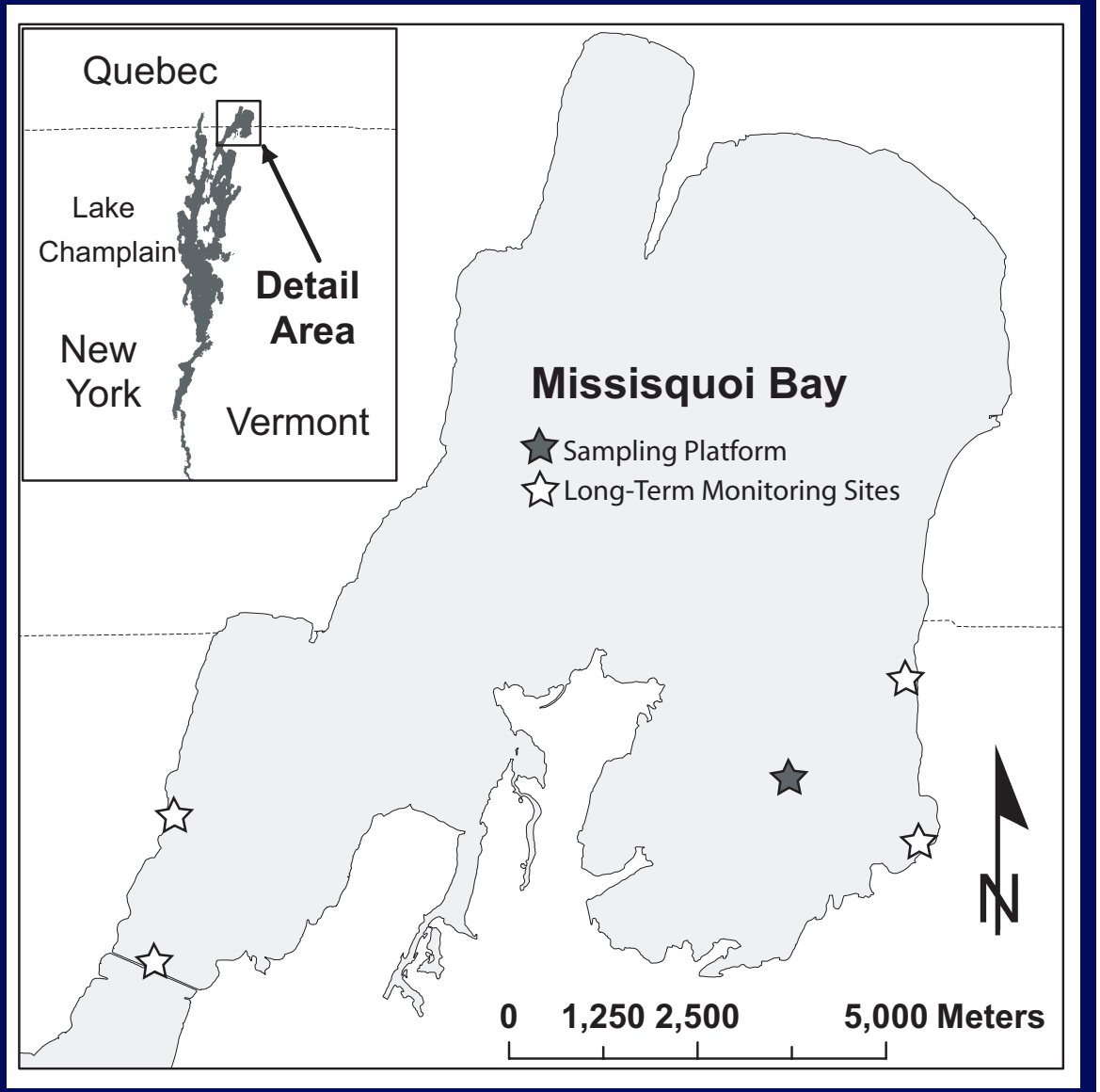
Long-Term Monitoring

- 5 Sampling locations
- 2005-2009
- Surface Grab Samples
- Limited analysis
- More frequent sampling

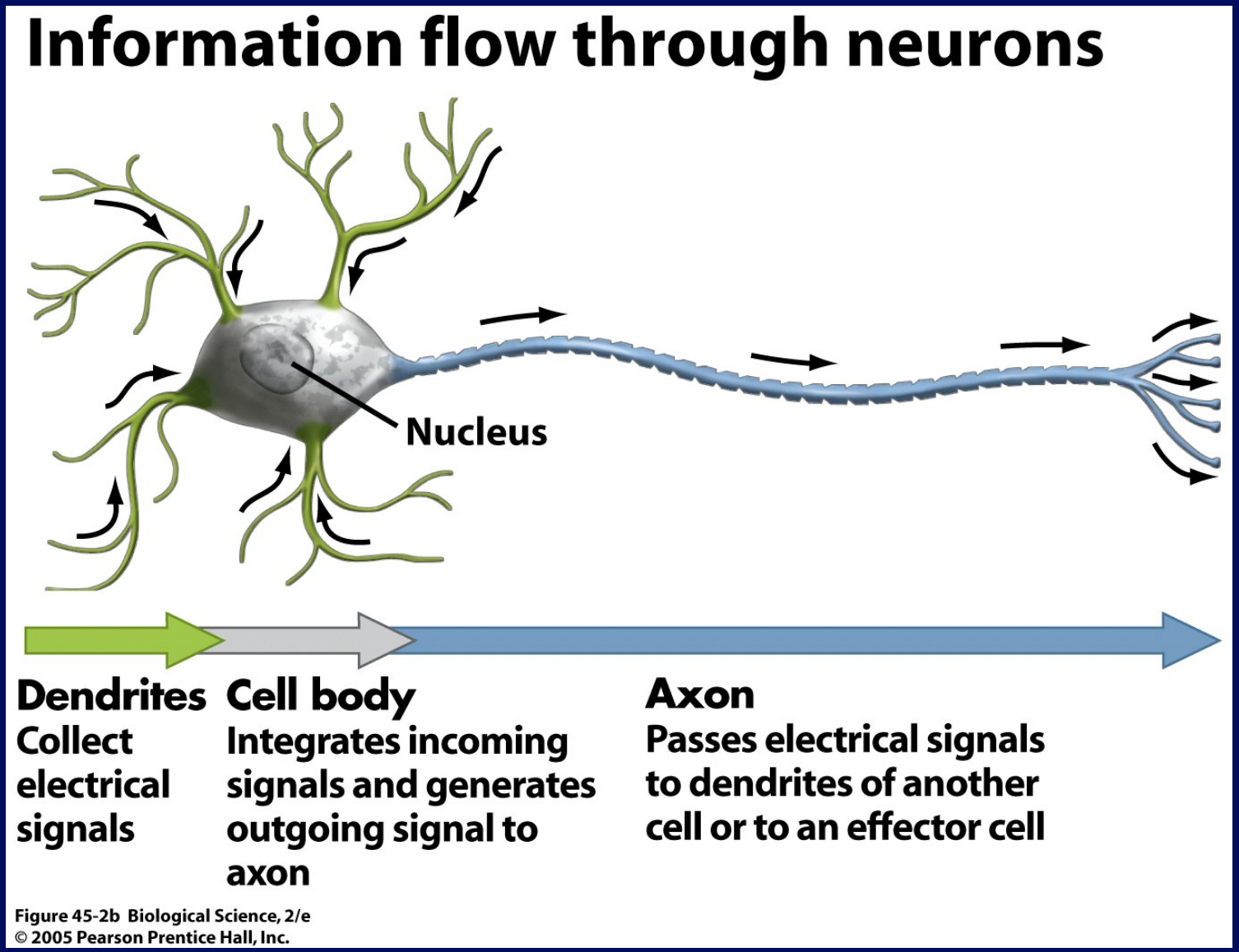


The Datasets - Limitations

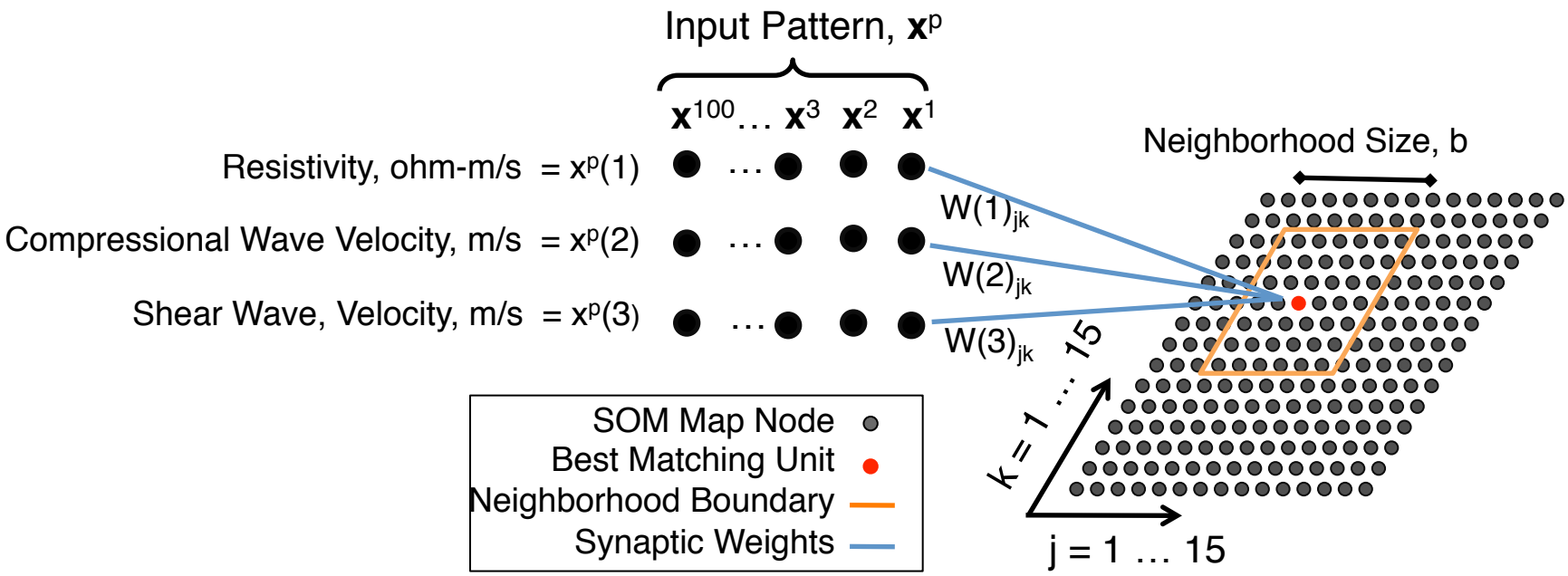
- Small n – ET dataset
- Correlated Varaibles
- Auto-correlated data
- Other artifacts of sampling?



Why an Artificial Neural Network?



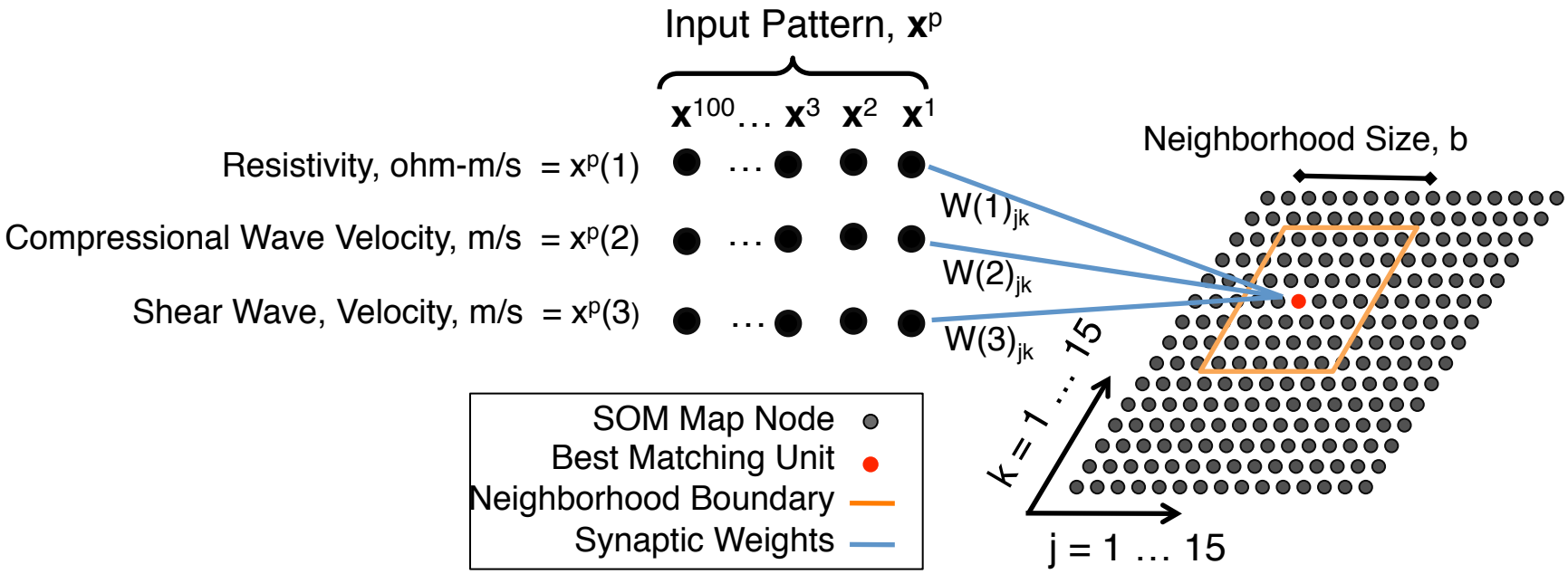
Self-Organizing Map (SOM) Network Architecture



Kohonen, T. (1990), The self-organizing map, *Proceedings of the IEEE, 78(9), 1464- 1480.*

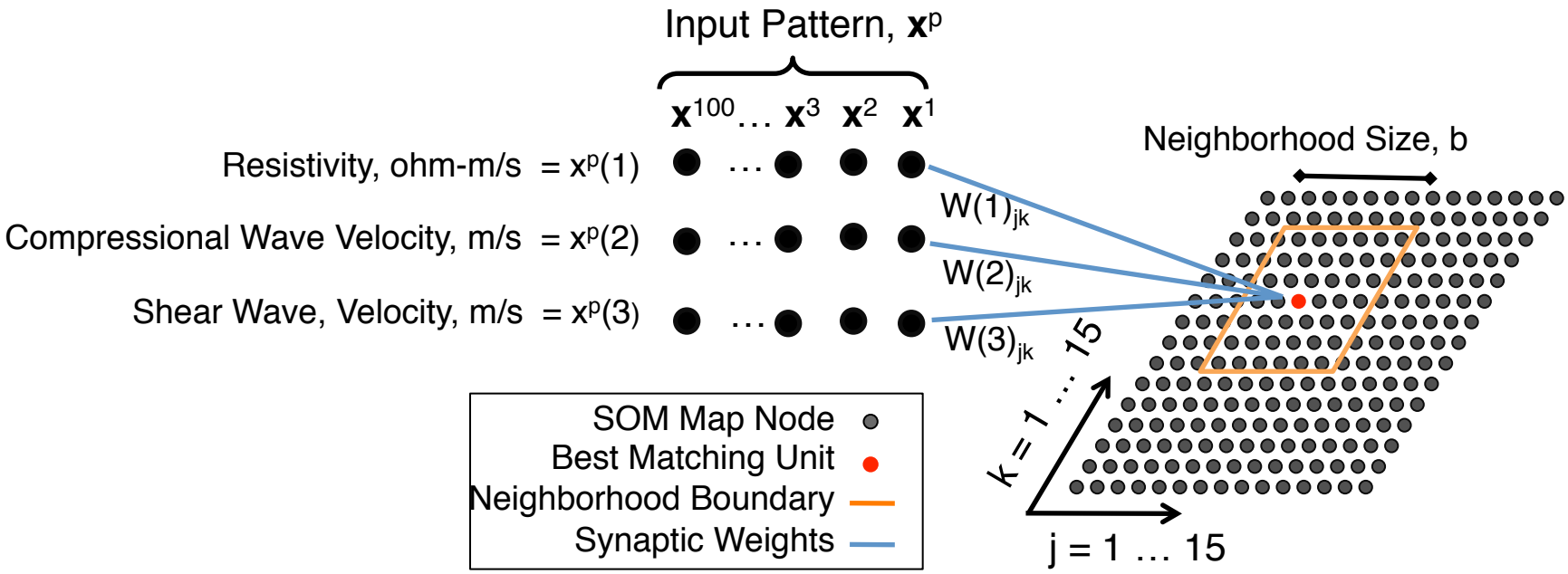
Step 1: Normalize input variables between 0 and 1

$$norm(\mathbf{x}_i) = \frac{(\mathbf{x}_i - \min(\mathbf{x}_i))}{(\max(\mathbf{x}_i) - \min(\mathbf{x}_i))}$$



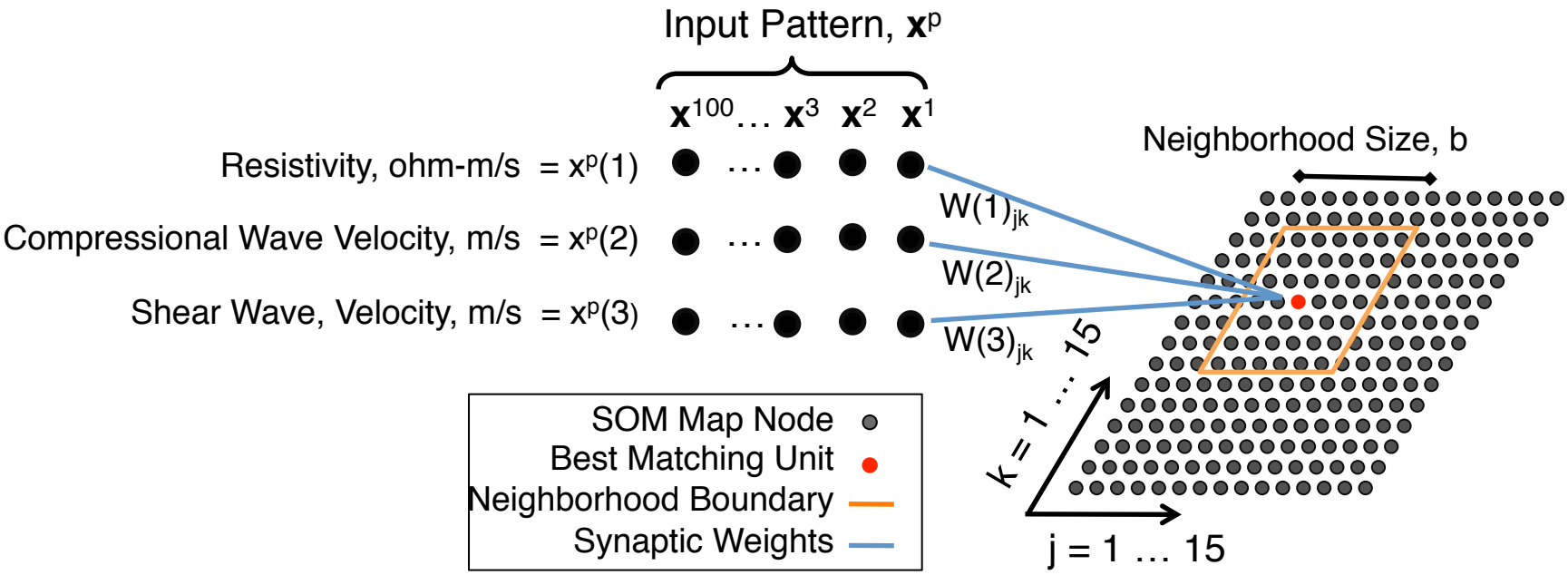
Step 2:Initialize weights

$$w(i)_{init} = \text{mean}(x_i) \pm \varepsilon \qquad \varepsilon <= 0.1 \times \text{mean}(x_i)$$



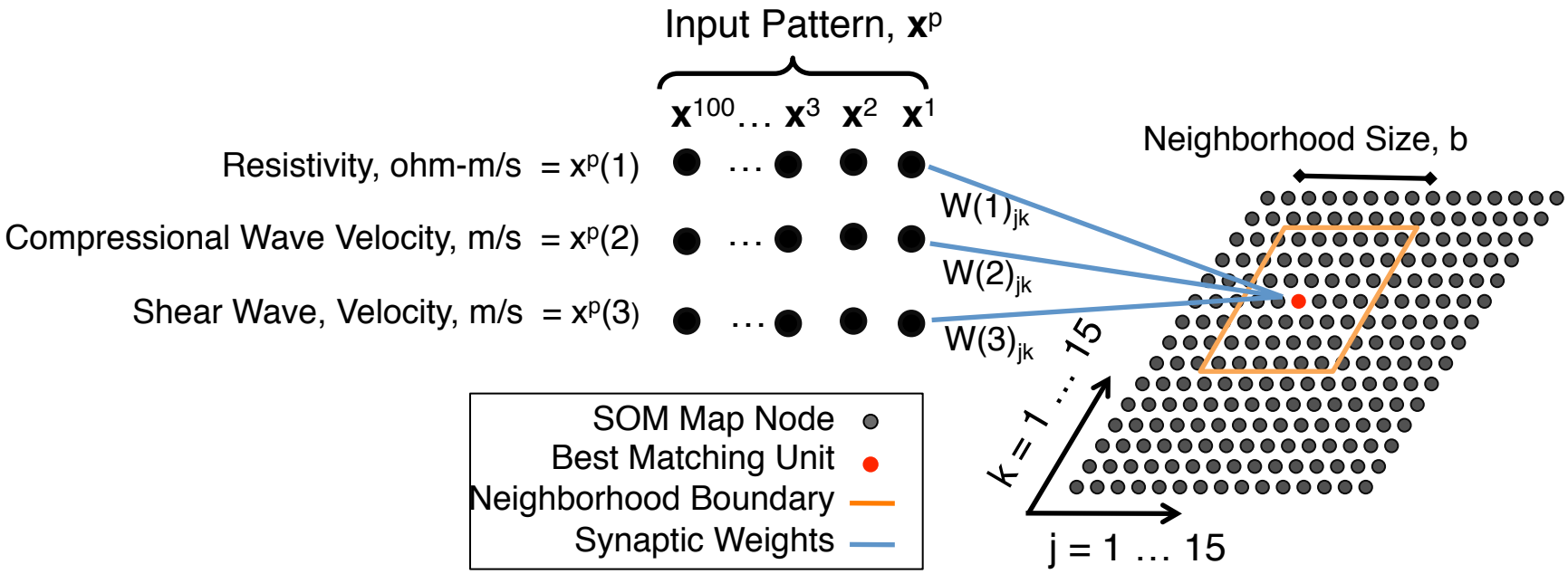
Step 3: Find Best Matching Unit (BMU)

$$BMU = \min\left[\sum_{i=1}^{I}\left(x(i) - w(i)_{jk}\right)^2\right]^{0.5}$$



Step 4: Update Weights

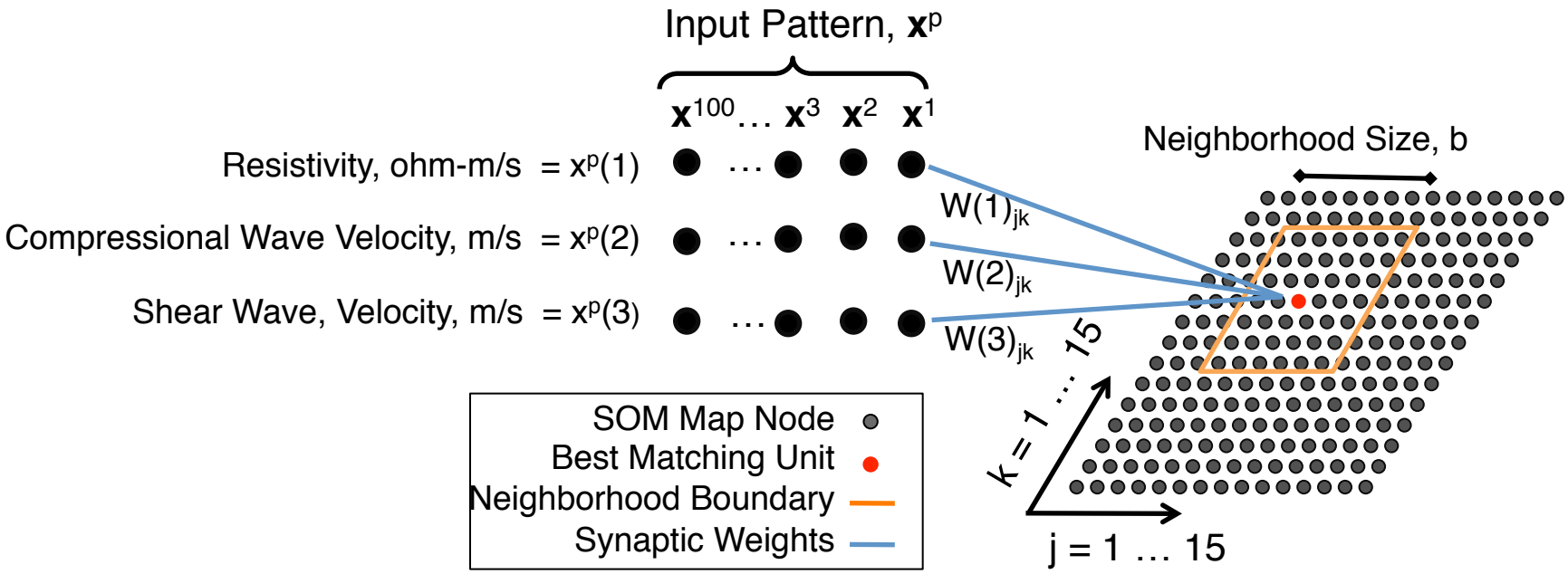
$$\mathbf{w}_{jk}^{new} = \mathbf{w}_{jk}^{old} + \alpha\left(\mathbf{x}^{p} - \mathbf{w}_{jk}^{old}\right) \qquad 0 < \alpha < 1$$



Repeat Step 3 and 4 Find BMU & Update Weights

Predefined # of iterations through the entire dataset

α and b decrease through training



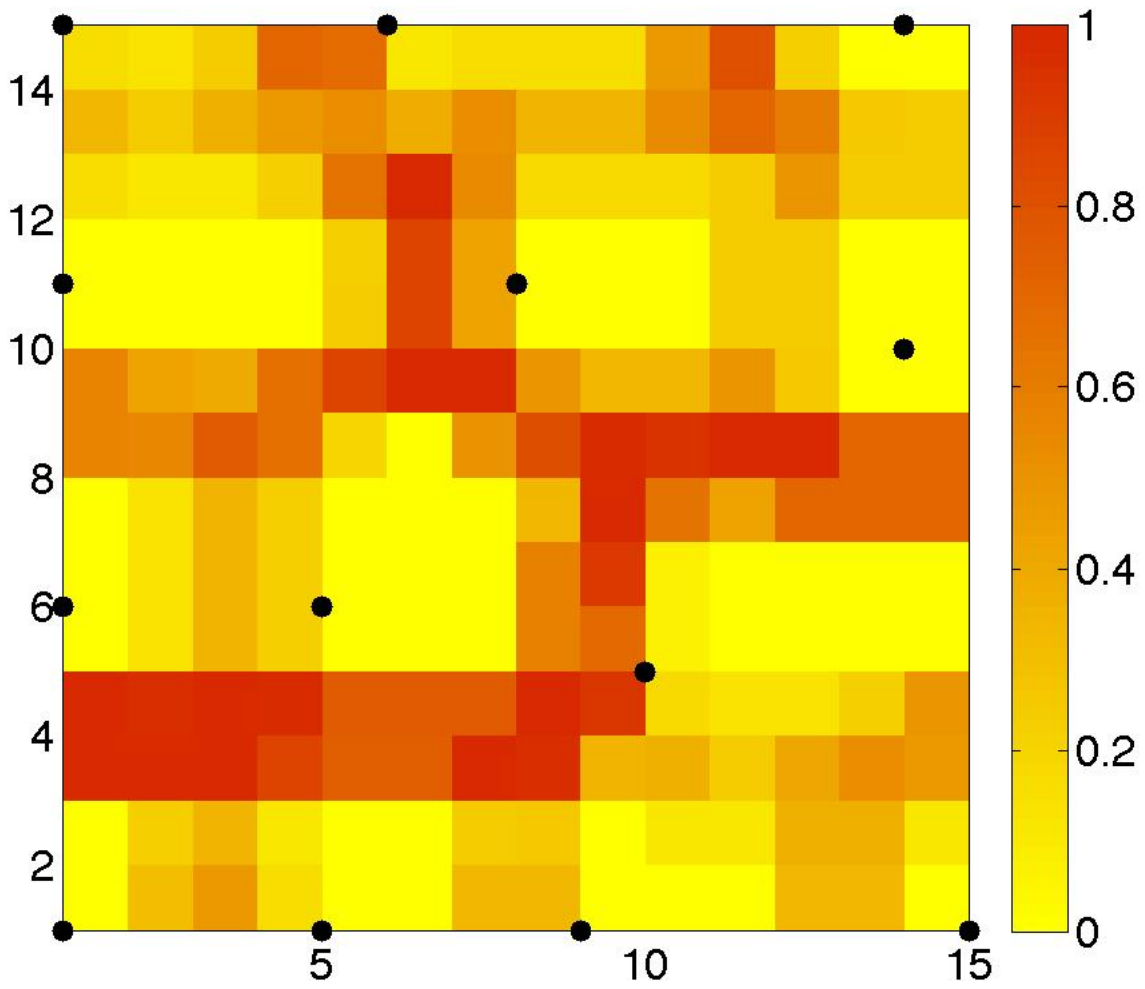
Introduction to the SOM: Kohonen Animal Example

	small	medium	big	2 legs	4 legs	hair	hooves	mane	feathers	hunt	run	fly	swim
dove	1	0	0	1	0	0	0	0	1	0	0	1	0
hen	1	0	0	1	0	0	0	0	1	0	0	0	0
duck	1	0	0	1	0	0	0	0	1	0	0	1	1
goose	1	0	0	1	0	0	0	0	1	0	0	1	1
owl	1	0	0	1	0	0	0	0	1	1	0	1	0
hawk	1	0	0	1	0	0	0	0	1	1	0	1	0
eagle	0	1	0	1	0	0	0	0	1	1	0	1	0
fox	0	1	0	0	1	1	0	0	0	1	0	0	0
dog	0	1	0	0	1	1	0	0	0	0	1	0	0
wolf	0	1	0	0	1	1	0	0	0	1	1	0	0
cat	1	0	0	0	1	1	0	0	0	1	0	0	0
tiger	0	0	1	0	1	1	0	0	0	1	1	0	0
lion	0	0	1	0	1	1	0	1	0	1	1	0	0
horse	0	0	1	0	1	1	1	1	0	0	1	0	0
zebra	0	0	1	0	1	1	1	1	0	0	1	0	0
cow	0	0	1	0	1	1	1	0	0	0	0	0	0

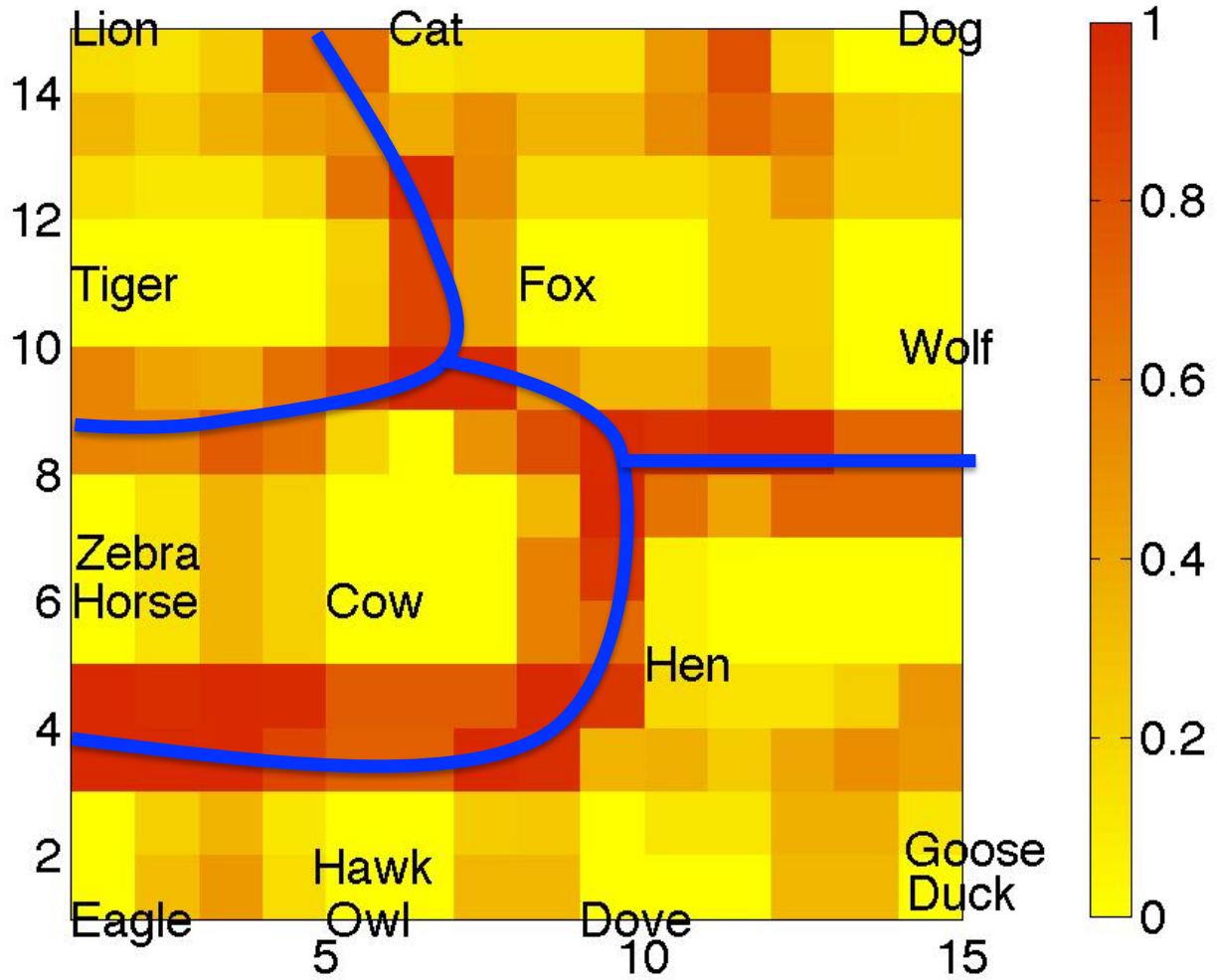
Introduction to the SOM: Kohonen Animal Example

Sample	small	medium	big	2 legs	4 legs	hair	hooves	mane	feathers	hunt	run	fly	swim
1	1	0	0	1	0	0	0	0	1	0	0	1	0
2	1	0	0	1	0	0	0	0	1	0	0	0	0
3	1	0	0	1	0	0	0	0	1	0	0	1	1
4	1	0	0	1	0	0	0	0	1	0	0	1	1
5	1	0	0	1	0	0	0	0	1	1	0	1	0
6	1	0	0	1	0	0	0	0	1	1	0	1	0
7	0	1	0	1	0	0	0	0	1	1	0	1	0
8	0	1	0	0	1	1	0	0	0	1	0	0	0
9	0	1	0	0	1	1	0	0	0	0	1	0	0
10	0	1	0	0	1	1	0	0	0	1	1	0	0
11	1	0	0	0	1	1	0	0	0	1	0	0	0
12	0	0	1	0	1	1	0	0	0	1	1	0	0
13	0	0	1	0	1	1	0	1	0	1	1	0	0
14	0	0	1	0	1	1	1	1	0	0	1	0	0
15	0	0	1	0	1	1	1	1	0	0	1	0	0
16	0	0	1	0	1	1	1	0	0	0	0	0	0

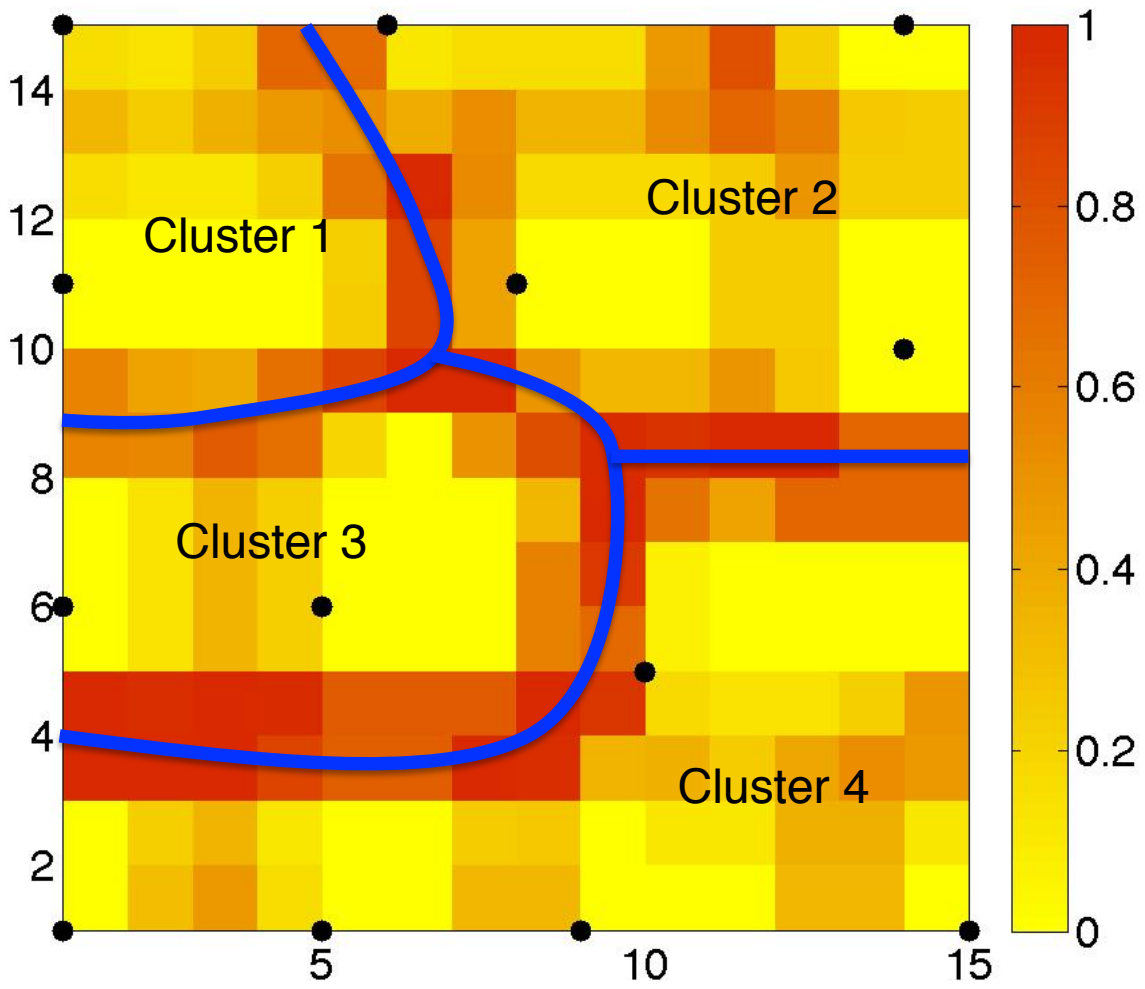
Kohonen Animal Example: U-Matrix



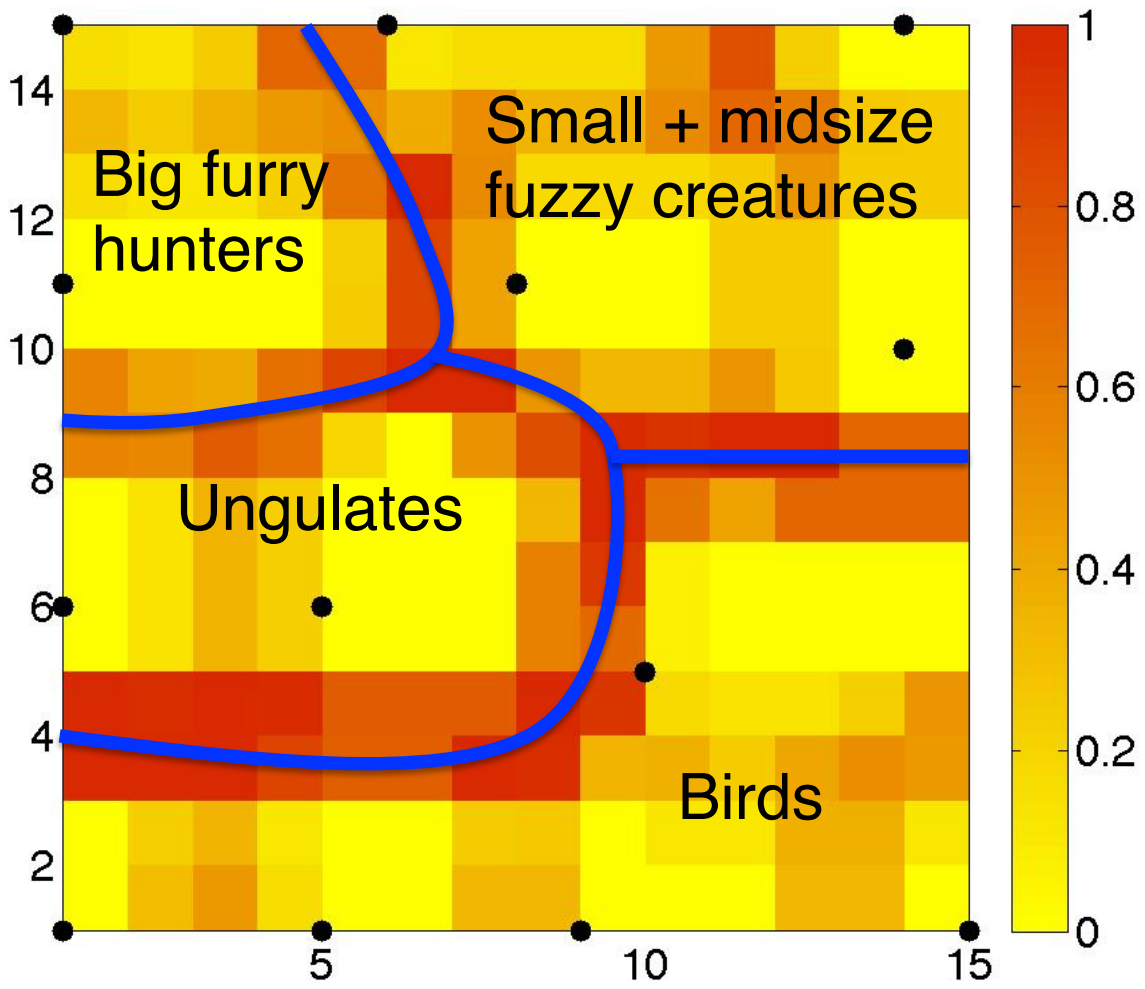
Kohonen Animal Example: U-Matrix



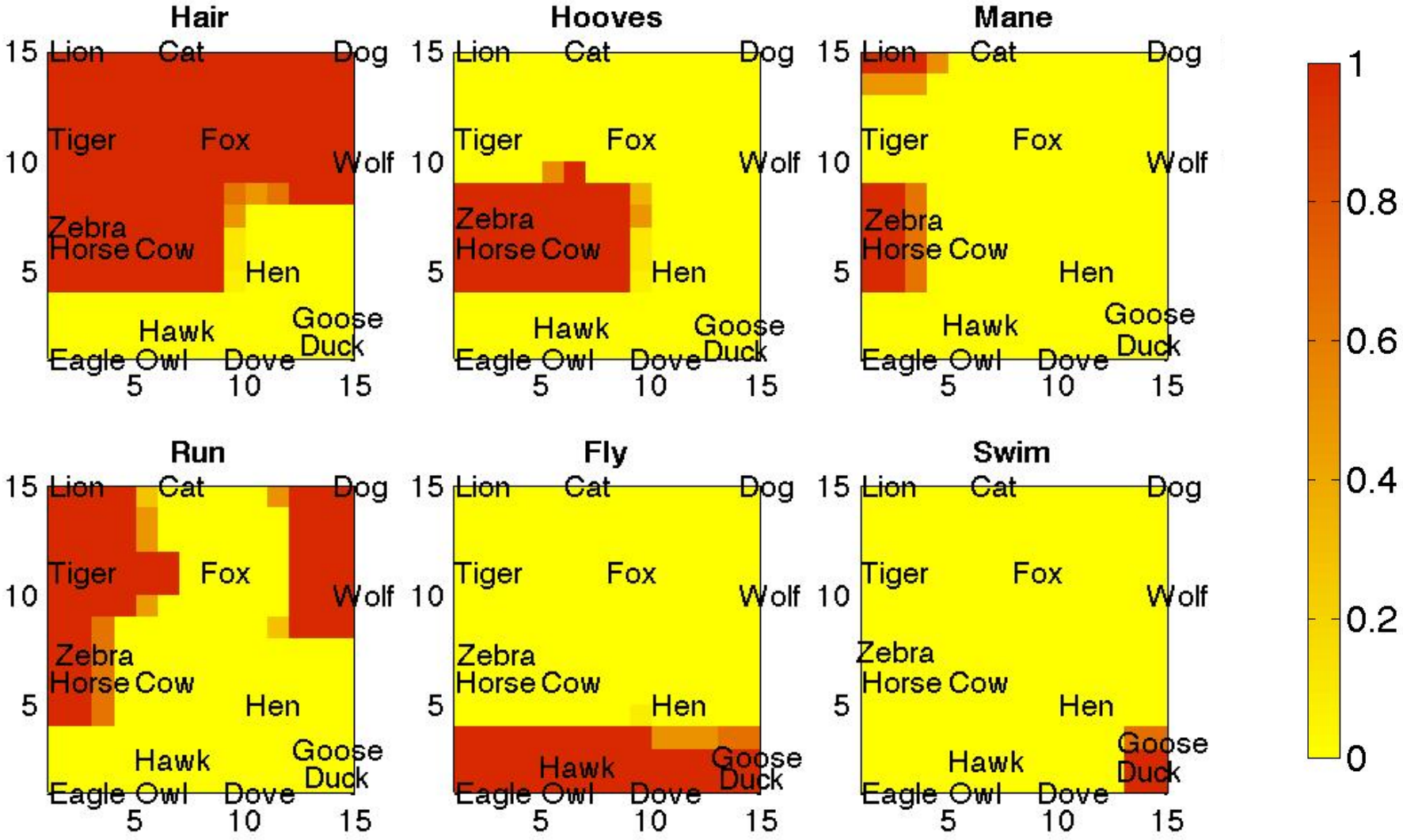
Kohonen Animal Example: U-Matrix



Kohonen Animal Example: U-Matrix

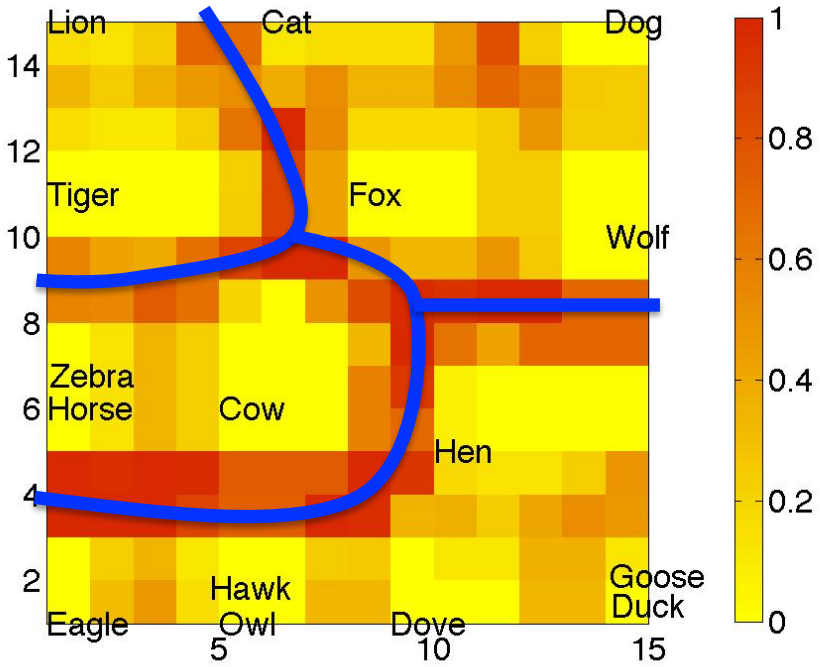


Kohonen Animal Example: Component Planes

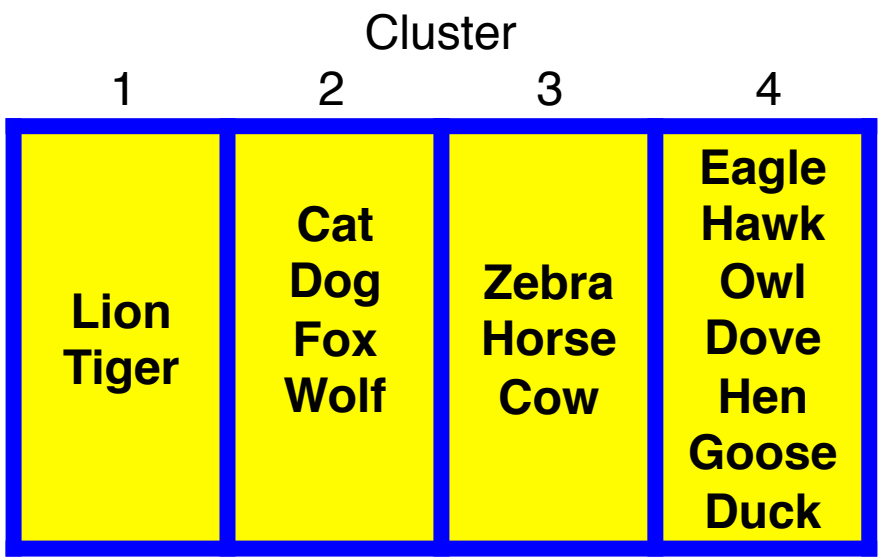


It's not always this easy with real data…

Many methods for clustering the SOM,
and the approach depends on the size of the dataset



15 x 15 node output map



1 x 4 node output map

Mangiameli, P., S. Chen, and D. West (1996), A comparison of SOM neural network and hierarchical clustering methods, *European Journal of Operational Research, 93(2), 402-417.*

SOM Input – Emerging Threats Dataset

Input to the SOM:

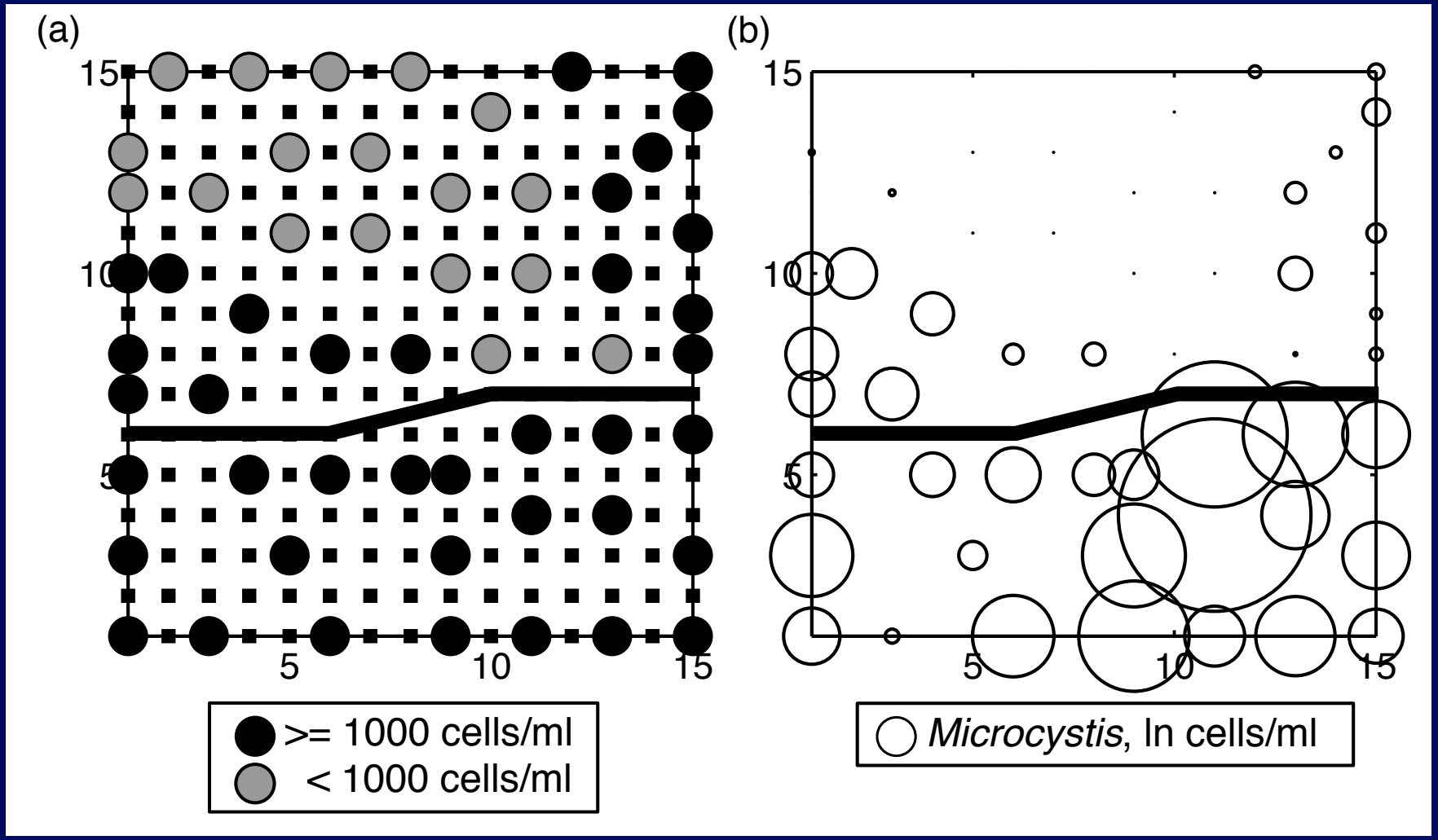
1. ln(**DN:SRP**)

2. **D.O**., mg/l
3. **Conductivity**, μS/cm
4. **Temperature**, deg C
5. **Turbidity**, ln(FTU)
6. **Chlorophyll Fluorescence**, ln(mg/m^3)
7. **Irradiance**, ln(W/m^2)

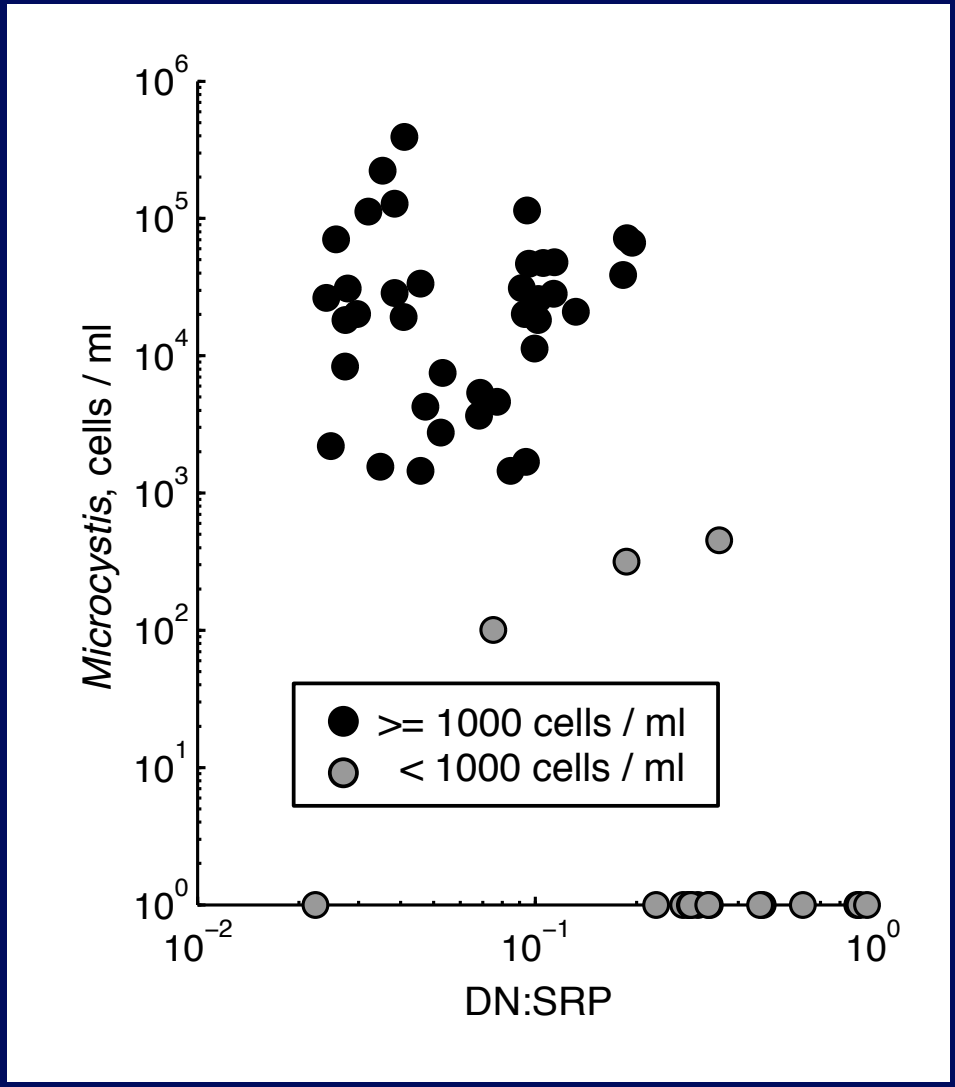
8. **Green Algae**, ln(cells/ml)
9. **Diatoms**, ln(cells/ml)

10. **Depth to Mn^{II}**, mm

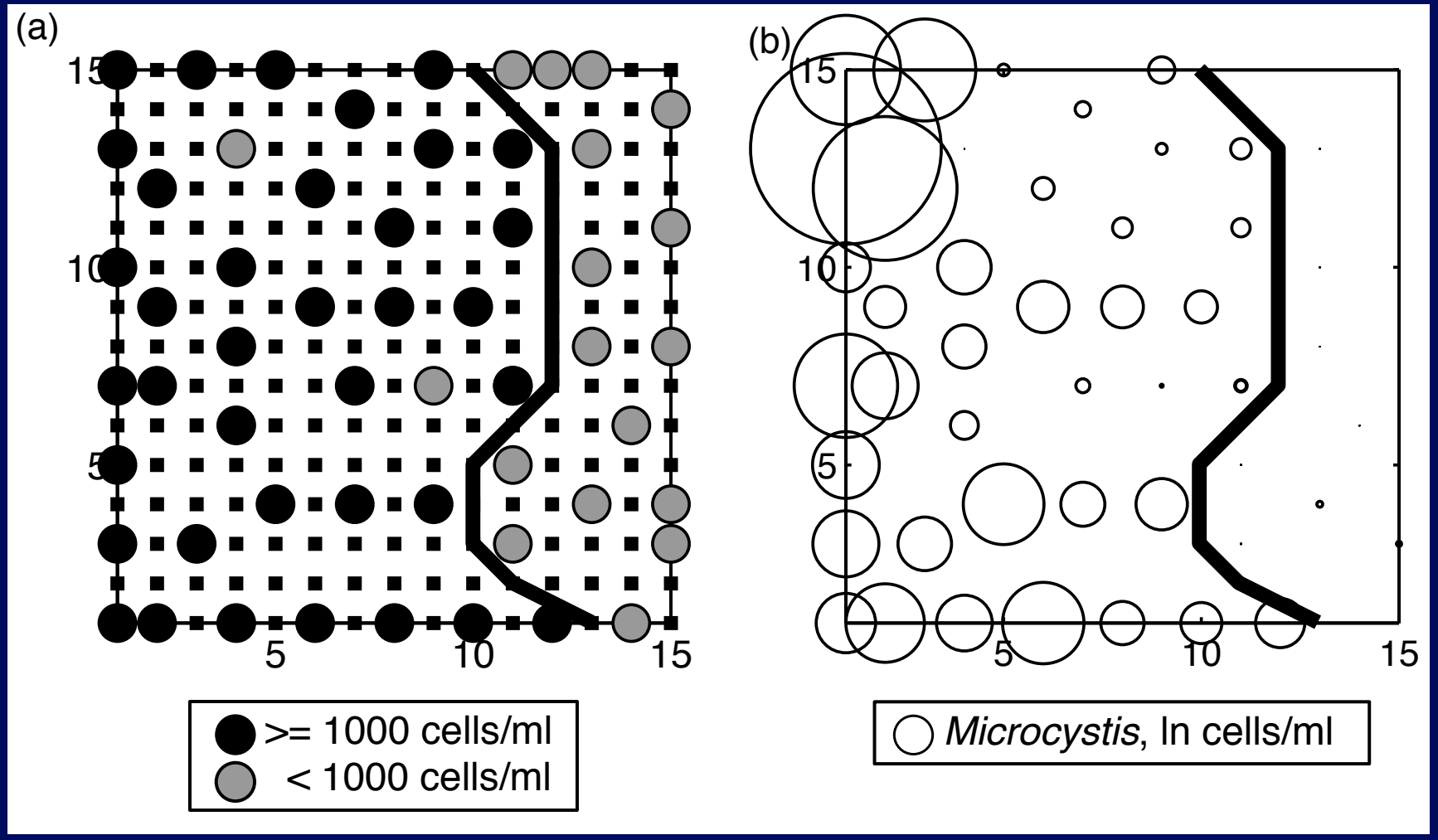
SOM Output – un-weighted by expert opinion



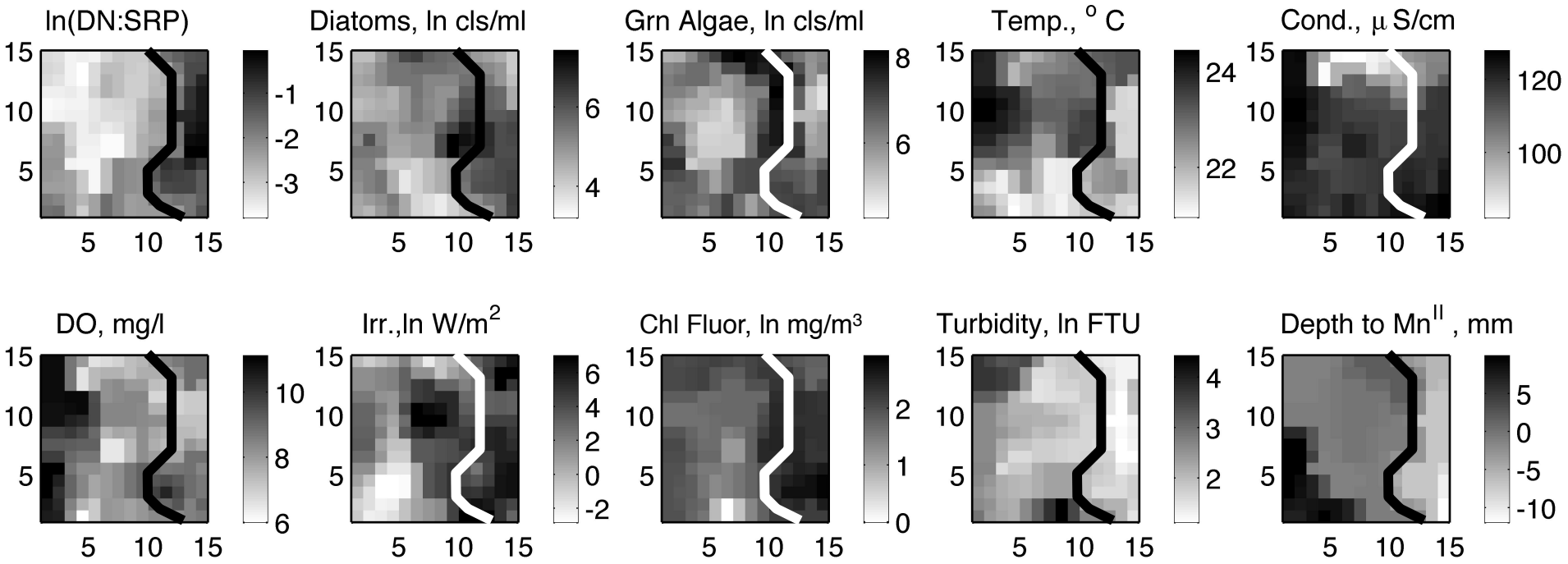
Bloom/No Bloom ?



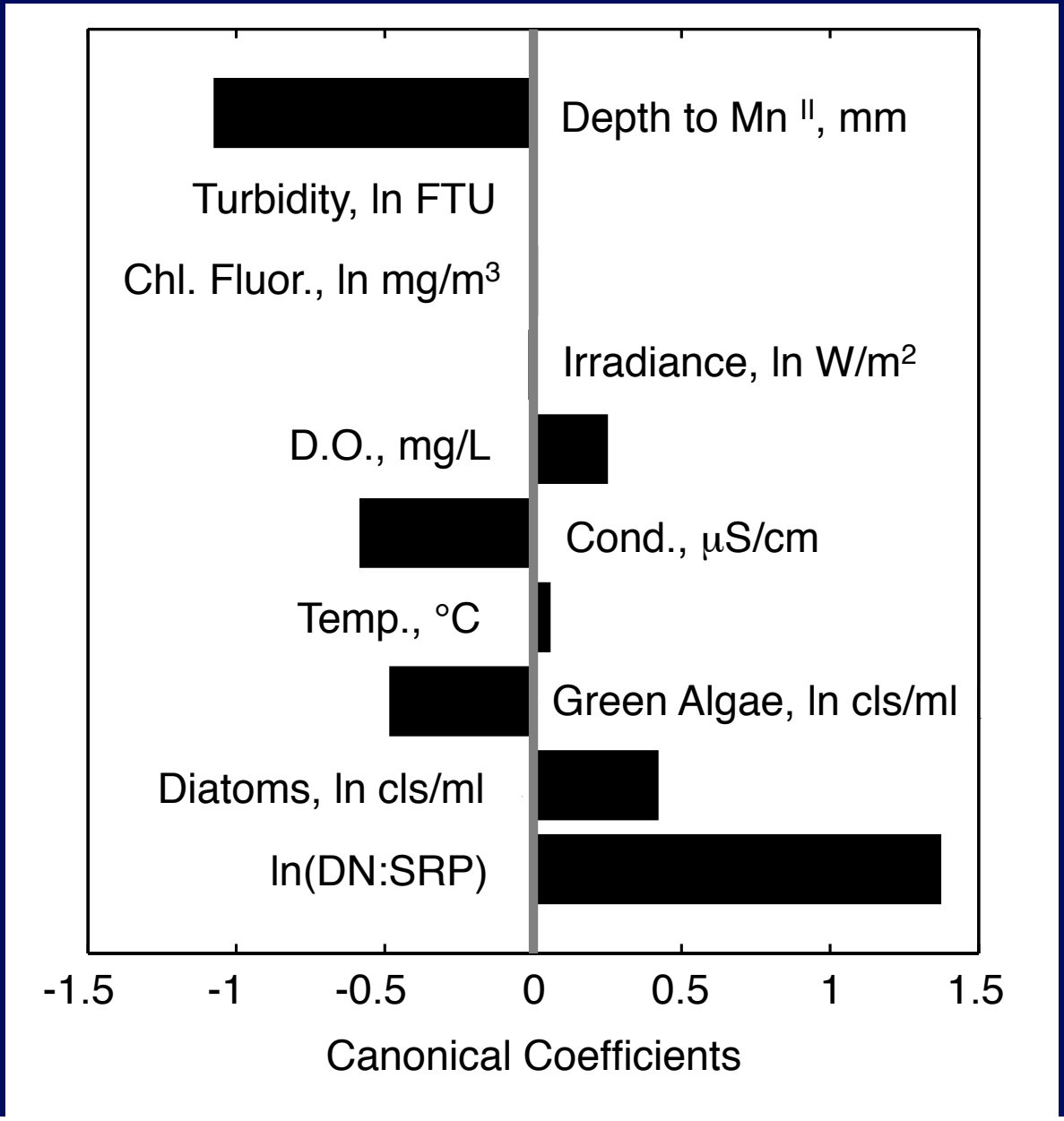
SOM input weighted by expert opinion



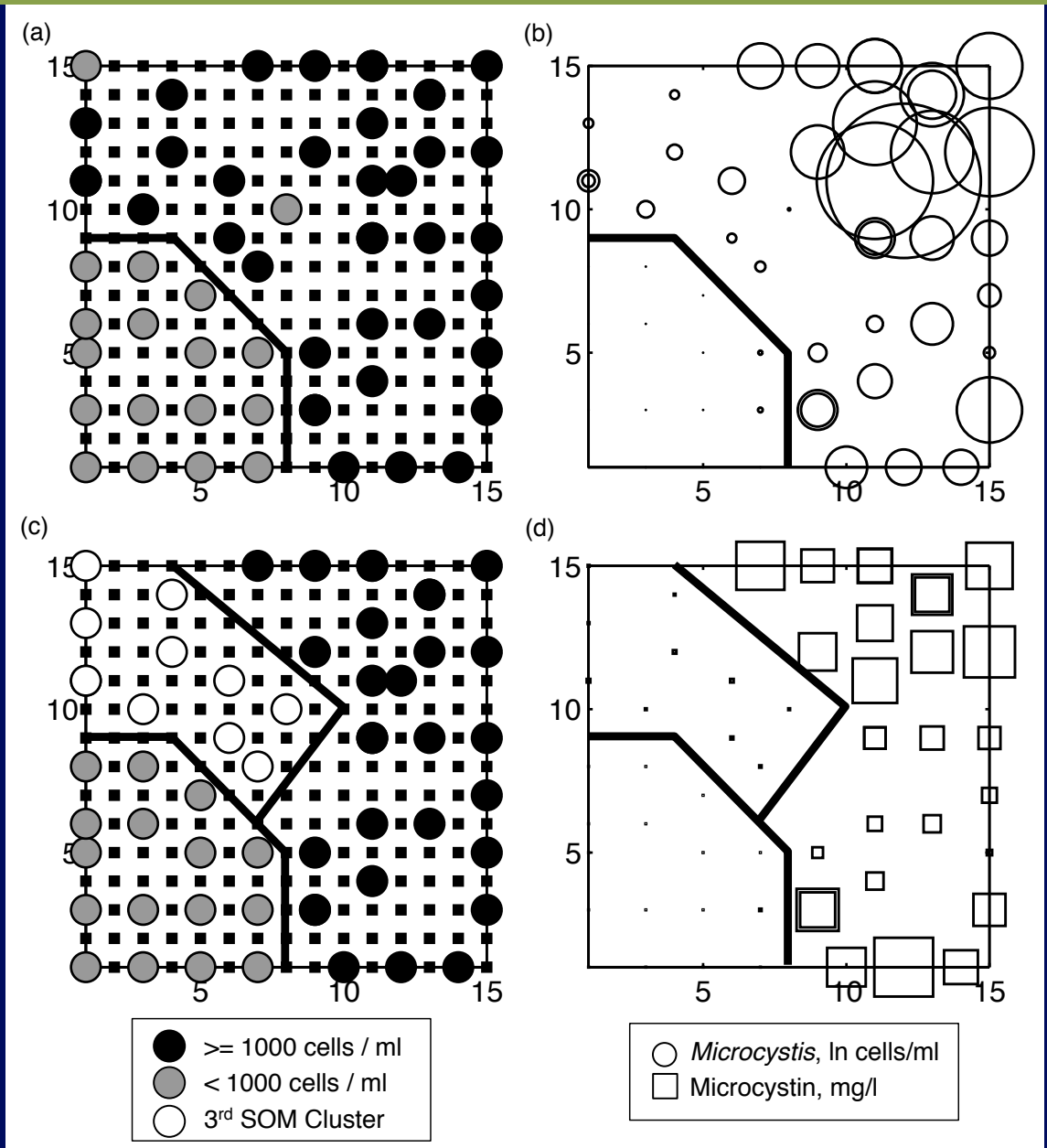
Component planes



Weights from discriminant analysis



SOM input weighted by discriminant analysis



SOM Input – Long-Term Monitoring Dataset

Input to the SOM:

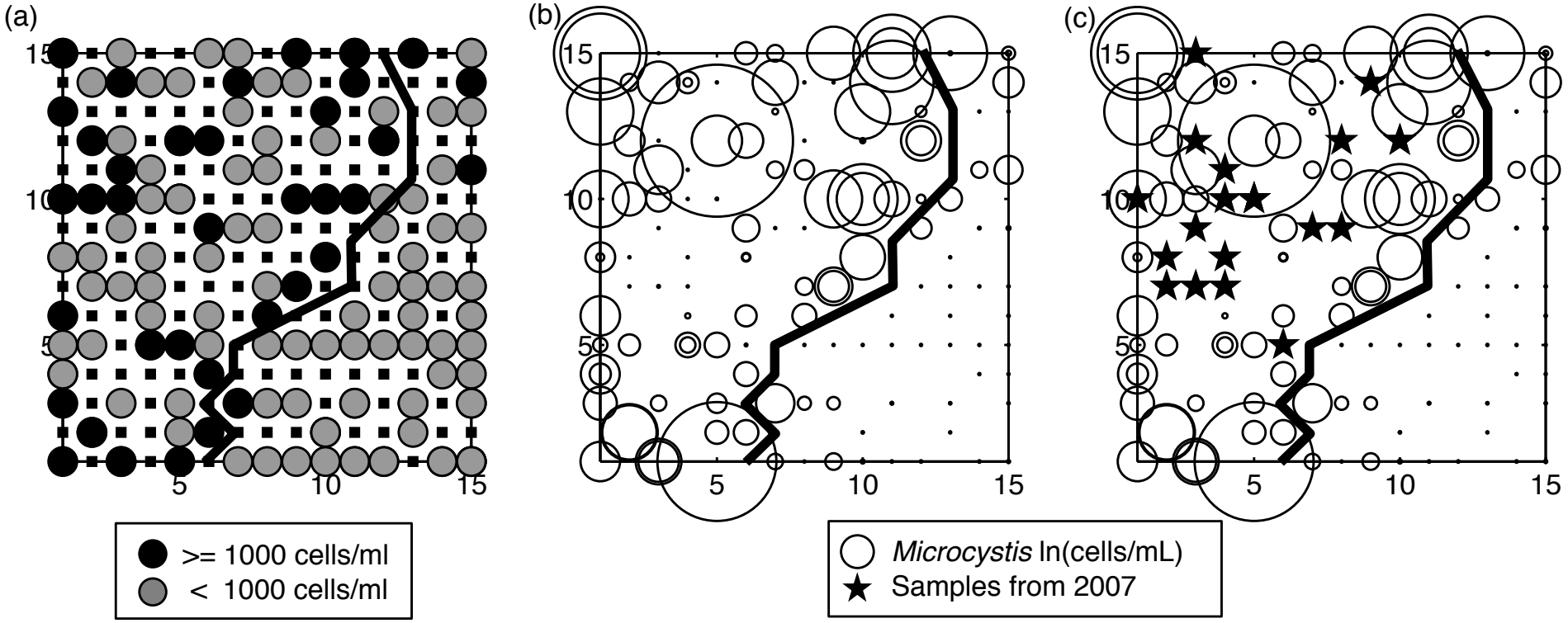
1. ln(**DN:SRP**)

2. ~~**D.O.**, mg/l~~
3. ~~**Conductivity**, μS/cm~~
4. **Temperature**, deg C
5. ~~**Turbidity**, ln(FTU)~~
6. ~~**Chlorophyll Fluorescence**, ln(mg/m^3)~~
7. ~~**Irradiance**, ln(W/m^2)~~

8. **Green Algae**, ln(cells/ml)
9. **Diatoms**, ln(cells/ml)

10. ~~**Depth to Mn^{II}**, mm~~ ≈ Day of Year

Long-term monitoring dataset



Conclusions

The SOM was able to distinguish between samples with high/low *Microcystis* based on other variables.

Sediment anoxia and DN:SRP seem to be the most closely associated with abundant *Microcystis*

2 pieces to the bloom:

- In-lake conditions to support and sustain the bloom
- A trigger – some combination of water/sediment/nutrient input, preditors, other phytoplankton … etc

- Collaboration lead to iterative analysis, better informed researchers and identification of data gaps.

- The SOM results identified the need for higher spatial and temporal sampling frequency to identify the environmental drivers of the bloom.