text
 Data mining
 Model
 Results and analysis
 Conclusion

 000000000
 000000
 000
 000
 000

Mining co-variation patterns from ecological data: a process to aid the construction and validation of computer models

Florent $ARTHAUD^{1,2}$ and Serge $FENET^{1,3}$ and Adriana $PRADO^{1,3}$ and Magali VANGKEOSAY^{1,3}

1 Université de Lyon, CNRS, Université Lyon 1, France

2 LEHNA Lab. UMR5023

3 LIRIS Lab. UMR5205

CompSust'2012, Copenhagen

Context	Data mining	Model	Results and analysis	Conclusion









5 Conclusion

Context	Data mining	Model	Results and analysis	Conclusion
0000000000				
Biodiversity and wetlands	;			

- Fresh waters represent 0.8% of planetary area
 - ... and 2.4% of known species (highest relative richness)
- Most impacted by biodiversity loss



The Dembes sends as	tured:	00000	000	000
0000000000	000000	00000	000	000
Context	Data mining	Model	Results and analysis	Conclusion

The Dombes

- Clayey shelf North of Lyon, France
- 1000 km² with 120 km² of ponds managed by man since Middle Ages
 - Thousands of ponds
 - Period of drought: pond are emptied for cereal culture (every 4-5 years)
 - Period of water: pond are filled for fish breeding
- Great stock of biodiversity



Context ○○●○○○○○○○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Alternative equilibrium	s			
Two stable	e states			





Context	Data mining 000000	Model 00000	Results and analysis	Conclusion
Alternative equilibriums				

Two stable states





Clear state

- Clear water
- Abundant and diversified sub-aquatic vegetation
- No eutrophication
- Few phytoplankton

Context	Data mining	Model	Results and analysis	Conclusion
000000000				
Alternative equilibriums				

Two stable states





Clear state

- Clear water
- Abundant and diversified sub-aquatic vegetation
- No eutrophication
- Few phytoplankton

Turbid state

- Turbid water
- Loss of biodiversity
- Few sub-aquatic vegetation
- Strong eutrophication
- A lot of phytoplankton

Context	Data mining	Model	Results and analysis	Conclusion
0000000000				
Alternative equilibriums				

Widespread process



Context ○○○○●○○○○○○	Data mining 000000	Model 00000	Results and analysis 000	Conclusion 000
Alternative equilibriums				
Widespread	process			



Context ○○○○○●○○○○○	Data mining 000000	Model 00000	Results and analysis	Conclusion
Alternative equilibrium	IS			

Hysteresis phenomenon

How to anticipate the state switch?

We seek to keep ponds clear, but the changing of state is not progressive ; it is violent and hard to anticipate: there are **alternative equilibriums**



Context	Data mining 000000	Model 00000	Results and analysis 000	Conclusion
Questions and data				
<u></u>	6			

Distinctive features

Main characteristics

- Complex and complicated ecosystem
- Agricultural exploitation
- Piscicultural exploitation
- Spatiotemporal-temporal dispersion (pond connectivity, seed dispersions, etc.)
- Injection of pollutants (linked to human activities)

Which questions?

- Climate change impact on biodiversity?
- Local human activities impact on biodiversity?
- How to ensure continuity of ecosystems "services"?
 - Which local private management practices?
 - Which global public practices?

Context	Data mining 000000	Model 00000	Results and analysis 000	Conclusion
Questions and data				
<u></u>	6			

Distinctive features

Main characteristics

- Complex and complicated ecosystem
- Agricultural exploitation
- Piscicultural exploitation
- Spatiotemporal-temporal dispersion (pond connectivity, seed dispersions, etc.)
- Injection of pollutants (linked to human activities)

Which questions?

- Climate change impact on biodiversity?
- Local human activities impact on biodiversity?
- How to ensure continuity of ecosystems "services"?
 - Which local private management practices?
 - Which global public practices?

Context ○○○○○○●○○○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Distinctive	features			

Current goals

Compute optimal management practices to reduce algal blooms

- Conserve/restore biodiversity
- Societal constraints

Context	Data mining	Model	Results and analysis	Conclusion
000000000000				
Questions and data				

Distinctive features

Current goals

Compute optimal management practices to reduce algal blooms

- Conserve/restore biodiversity
- Societal constraints
- Choices of fish species and fish densities
- Supplementary feeding
- Position in pond network



Context	Data mining	Model	Results and analysis	Conclusion
00000000000				
Questions and data				

Distinctive features

Current goals

Compute optimal management practices to reduce algal blooms

- Conserve/restore biodiversity
- Societal constraints
- Choices of fish species and fish densities
- Supplementary feeding
- Position in pond network
- Type of cultures
- Rhythm of drought
- Fertilization, liming, permaculture...



Context	Data mining	Model	Results and analysis	Conclusion
000000000000	000000	00000		000
Questions and data				
Available	data			

Data difficult to obtain

- Anthropic constraints (reluctant landowners)
- Few measurement campaigns (financial cost)
- Noisy data
- Long process

Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

• Collect water samples

Vegetation communities



Context ○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
A				

Available data

Phytoplankton communities

- Collect water samples
- Count specimen under microscope

Vegetation communities



Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

• Define the boundaries of quadras



Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

- Define the boundaries of quadras
- Collect plant samples

Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

- Define the boundaries of quadras
- Collect plant samples
- Identify plant characteristics



Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

Propagule communities

• Define propagule context



Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

- Define propagule context
- Collect propagules



Context ○○○○○○○○○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

- Define propagule context
- Collect propagules
- Plant propagules and...



Context ○○○○○○○○○●○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

- Define propagule context
- Collect propagules
- Plant propagules and...
- ... let grow in favorable environment



Context ○○○○○○○○○	Data mining 000000	Model 00000	Results and analysis	Conclusion 000
Questions and data				
Available	data			

Vegetation communities

- Define propagule context
- Collect propagules
- Plant propagules and...
- ... let grow in favorable environment
- Identify plant characteristics



Context ○○○○○○○○○●	Data mining 000000	Model 00000	Results and analysis 000	Conclusion
Questions and data				
Raw data				

• 154 parameters for 90 ponds

• Near-monthly basis from 2009 to 2011, and more is coming

- Architecture of the pond and management policies
 - Surface
 - Geographical position and links with neighbouring ponds
 - Fishing and agricultural data
- Established lake vegetation
 - 55 samples species
 - Floating VS submerged
- Several chemical levels
 - Nitrogen
 - Nitrate compounds
 - Carbon compounds
 - Ionic activity

• 94 genera of bacteria and algae belonging to 7 families

Context ○○○○○○○○○●	Data mining 000000	Model 00000	Results and analysis 000	Conclusion
Questions and data				
Raw data				

- 154 parameters for 90 ponds
- Near-monthly basis from 2009 to 2011, and more is coming
 - Architecture of the pond and management policies
 - Surface
 - · Geographical position and links with neighbouring ponds
 - Fishing and agricultural data
 - Established lake vegetation
 - 55 samples species
 - Floating VS submerged
 - Several chemical levels
 - Nitrogen
 - Nitrate compounds
 - Carbon compounds
 - Ionic activity

• 94 genera of bacteria and algae belonging to 7 families

Context	Data mining 000000	Model 00000	Results and analysis 000	Conclusion
Questions and data				
Raw data				

- 154 parameters for 90 ponds
- Near-monthly basis from 2009 to 2011, and more is coming
 - Architecture of the pond and management policies
 - Surface
 - Geographical position and links with neighbouring ponds
 - Fishing and agricultural data
 - Established lake vegetation
 - 55 samples species
 - Floating VS submerged
 - Several chemical levels
 - Nitrogen
 - Nitrate compounds
 - Carbon compounds
 - Ionic activity
 - 94 genera of bacteria and algae belonging to 7 families

Context	Data mining 000000	Model 00000	Results and analysis 000	Conclusion
Questions and data				
Raw data				

- 154 parameters for 90 ponds
- Near-monthly basis from 2009 to 2011, and more is coming
 - Architecture of the pond and management policies
 - Surface
 - · Geographical position and links with neighbouring ponds
 - Fishing and agricultural data
 - Established lake vegetation
 - 55 samples species
 - Floating VS submerged
 - Several chemical levels
 - Nitrogen
 - Nitrate compounds
 - Carbon compounds
 - Ionic activity

• 94 genera of bacteria and algae belonging to 7 families

Context	Data mining 000000	Model 00000	Results and analysis 000	Conclusion
Questions and data				
Raw data				

- 154 parameters for 90 ponds
- Near-monthly basis from 2009 to 2011, and more is coming
 - Architecture of the pond and management policies
 - Surface
 - Geographical position and links with neighbouring ponds
 - Fishing and agricultural data
 - Established lake vegetation
 - 55 samples species
 - Floating VS submerged
 - Several chemical levels
 - Nitrogen
 - Nitrate compounds
 - Carbon compounds
 - Ionic activity
 - 94 genera of bacteria and algae belonging to 7 families

Context 00000000000	Data mining ●00000	Model 00000	Results and analysis	Conclusion 000						
ooking for co-variation patterns in data										
Co-variati	on patterns									
· · · · · · · · · · · · · · · · · · ·	1811 x x # x # x + x + x + x + x + x + x + x		* * * * * * * * * * * * * * * *							

			-18 -1 -1		1 4 1 4								-					-					-	1.4		~ ~	- 48	-	 1		
	-				Parami and	in motor	feedback from	de Dadage	strates in	na inci yan inci	n bein	a Bastan	inche Gentre	in the second	Party Party .	a salar ing		· · · · · · · ·	start and	÷		and the second s									
And in case of	100	age as	Contraction of the local distribution of the	******			-	-		10.00		10 8			Hald store	-	1-84	ALCONA.	4543	-	14 .11	181	114						 		
and the second	100	E: 22													Belle Lamo		100	0.000	-		18 R	140									
The second value of	-	C: CC								B-00					mote and			Con Contractor	4543												
and the second distance	1000	21 22			. :										DOM LED	CA HEADA		A LOOK	00 3			1.007									
States in the local division of	200	21 22								H-0 1	646 8	0 X	2 44		2014 202	48 +1.25.8		an anna	063 3		45 50	184	443		2 124						
sufficient designed		S: 22													2010 88	9 15458		Can Juppern	(95) 2		10 10	3804	968		100		10				
The second secon	100	2: 22		: :::::						10 I		: ;	: :		2010 AL	-	100	A LONG	00 3	ACC 140		1411			:					: :	
The second of	2010	2: 22													1 10 10 10 10 10 10 10 10 10 10 10 10 10			In Assess	1042			2.000					10 C				
the second second	1.000	21 22								-			0 24		and and	-		an anna a	4344		68 64	2404	6.00	12 A			141				
The second secon	100	21 22													and the	1000		a conta	33 3			150			: ::						
and the second s	.00	2: 22		1 22212						22 1	H 1				Bold 170		- 25	1000	221 1		11 12	100	277	12 D	1 11						1.1
The second se		2: 22																1000	1000				200		: ::						- E
and the second	Jani									95	342				40-14 +14	1005		in Josef			68 1	1.808	044		2 66.3		1		_		_
and the second	100	21 22		: :::::						-		:			2-12 -12	100		an Junio				1000		14 3							
and the second	200	2: 22								-						1000	-	tan parama				6.057	100							: :	
the second	1.004									-	-					and the second	- 84	and statements				1000			-		8				
- C - C - C - C - C - C - C - C - C - C	220	FI 55								-	22				272 22	222		and a state of			21 12	:::::	12		1 22		2				
and the second		21 22		: :::::						-	2	:			200 100	100		an could			3 13	155	- 22		1 21						- A
and the local division of the local division		2: 22				: :	1 1			22	21				20 21	222	- 84	IN LUMM			1 12	1007	227		: 12						1.12
Statement Statement		S								-						****		In conten				1304	0.07				-				6.6
the second se	1.000	C: C2								-					-	-		Tan Parkets				1.000	0.04				-				
1000	200	21 22		: ::::::						-	22	:			2-12 -12	1000		IN JOLM			6) 100	1100			1 14						
and the local division of		2: 22					1 1			95	×2				2-12 -10	4005	- 64	in links			14 14	3.348	0.04		3 204						
and in case		SI 22								-	84				10-12 -00	4040	- 644	148 24506			67 90	1300	100		1 111						
THE REAL	100	G 1 2 2						- 11		-		÷ .			11 100			tis curies	147305	1212		100							_		_
and the second	+364	2: 22		1 22212			1 1			25	50 B				10 100	2 2000		134 J.4768	147340		54 DE	1100	212	12 1	1 12						1.1
and in case of	2954	SI 22							- 1	÷.	100				10 10		- 64	131 24768	10,000		47 90	here?	100								
THE R OWNER	1.164	SI 22								-								IN DOMES	10700			1174									
and the second	1000	2: 22								-								the success	10700			LHC									
and the second second	214	21 22							- 1	-	111				641 69	al babe		in house	ed Table		43 648	1000	443	a a			<u> </u>				
The Artist		F: 55							- 5	÷.	444				- +4 +e			the subset	16586		66 86	3296	0.00		4 64						
THE REAL		21 22								-	100				10 10	A 1910		in lots	107000			1400			1 10		<u>.</u>				
State of Concession, Name of State of S	100	2: 22								=	100					LA 1400		In LUNCE	107.000			3303					-				1
The start	1244	2: 22								=								the states	10700			2004			: *						
and the second		21 22								-						-		in house	and Loope			1000									
and desired		PI 22								-	21 1							in lots	1.1.1		2 2		122								
And Address	144.54	C: 22		: :::::						-		-	•			1000	-	THE PARTY OF	10100	100		1004							_	· · · ·	_
All Crusters		21 22														(Calmin	- 63	- 24/44	al a		44 9.3	110			1 12						
All Crambos	200	F: 55														10404	1.23	a 10(44	10100												
Contraction of the local division of the loc		21 22														1790	- 123	and second	10100												
Contraction of Contraction	1000	2: 22		: ::::::		: :										17 miles	- 23	ALL DATES	ALC: NO			1000			:					: :	
THE LOCKER																1000		and statistic	and the								-				
the states		21 22														1000	- 23	10100	100			100	12		1 21						
TR Conten		21 22														1000	- 23	and and and	104.00		- 11 L	1.00			: 22		8				
Carlos Carlos		21 22		: :::::												100	1.23	an sarah	and the second			2305	142		. 31		14				
Calific States	100	2: 22				: :										1000	- 53	10735	and the		22 22	100	227								
The Party of the P	100															11000	1.00	an erran	and the			1114									
The local division in	1.1.1	CI 22														11000	- 23	and survey	and the			1.010			1		2 C				
Outer	100.04			: ::::::	. :											27000	-23	and Andreas	104.00		10 NO	LHD	0.04	10 X	4 61				_		_
THE OWNER.	100		-							-			-		tanta surr	an chicker	-	NR CORDS	10.000			1.0.0	1.04								
the state of the	-									22			: 2			1000		and contract			- 21 - 22	100	100		1.1						
Mini Outstand											12	: :	: 3			10000		AN LONGE	10.000			1.010					2				
and Containing	100			: ::::::						22 1	12	: :	: 3		101 12	1000		NA LINGS	1.110	55		2105	227	2 2	: 53		3				1.1
Canada Canada										52 3	12	: :			10.11 10.	10101		KR EDWER	10000		21 22	1.04	22	10 1	112		10			1 1	
Canada Canada										-					10-12 12	the state	- 60	10.0403	10.000			100	100	-							6 R
Containing Containing	2414			: :::::						22 1		: :	: 2		Maria Ma	No thicks		No. COMPLE	10.000		11 10	1000									
Concession of the local division of the loca	100			1 22212						20-00	0-00	: :	1 2		10-11 64	in Links	- 20	ALL DOWNER	113500	1241		1114					141				
Contained Contained	2.6									8-0	848				8-9 64	a istro		ALC: LONGE	43690		55 50	1306	44								
And Distances	- 3									840	0-00				10-10 24	10 10530	- 84	NR LONGS	183576		5.0 0.0	3758	948	121 X			101				
Canada Constants										240		: :	: 2		10-12 -04	100.00	1.00	NA LONG	10.000		12 10	2007	010		4 130						
AND CONTRACTOR	1000									2040	240				tana and	a show		ALL PROPERTY.	10.2000	1041	17 10	1,758	847						-		-
Contraction of Case	2004	5								18-60	10			191	20-02 240	10 4400	1.00	NAL PROPERTY.	107354	141	- 10	1.856	- 11	ND 8	1.40						

Use data mining techniques to identify important relationships in the data that could guide modeling and data gathering process

- Database of at most 154 attributes per measure
- No discretization of data

Context 0000000000	Data mining ●00000	Model 00000	Results and analysis 000	Conclusion 000							
Looking for co-variation patterns in data											
Co-variation	Co-variation patterns										

Use data mining techniques to identify important relationships in the data that could guide modeling and data gathering process

- Database of at most 154 attributes per measure
- No discretization of data
- Ability to combine numerical and categorical attributes
- Find all frequent patterns

Context 0000000000	Data mining 0●0000	Model 00000	Results and analysis	Conclusion 000						
Looking for co-variation patterns in data										
Co-variatio	n patterns									

Identify sets of numerical attributes that behave similarly (with **itemset mining** [Prado et al. EGC'2012]):

- Records are sorted w.r.t every attributes ~> index ranking
- For all sets of attributes we compute the ratio of the number of pairs of records (among all possible pairs) for which all attributes in the set have the ranking index in the first record significantly different than in the second one ~> correlation ratio
- High ratio implies that attributes in the set behave similarly

 → potentially interesting covariation patterns
- Extract all patterns with a ratio higher than a user defined minimum support threshold

Context 0000000000	Data mining 0●0000	Model 00000	Results and analysis	Conclusion						
Looking for co-variation patterns in data										
Co-variatio	on patterns									

Identify sets of numerical attributes that behave similarly (with **itemset mining** [Prado et al. EGC'2012]):

- **(**) Records are sorted w.r.t every attributes \rightsquigarrow index ranking
- Por all sets of attributes we compute the ratio of the number of pairs of records (among all possible pairs) for which all attributes in the set have the ranking index in the first record significantly different than in the second one ~> correlation ratio
- High ratio implies that attributes in the set behave similarly

 potentially interesting covariation patterns
- Extract all patterns with a ratio higher than a user defined **minimum support threshold**
| Context
0000000000 | Data mining
0●0000 | Model
00000 | Results and analysis | Conclusion |
|--------------------------|-----------------------|----------------|----------------------|------------|
| Looking for co-variation | n patterns in data | | | |
| Co-variatio | on patterns | | | |

Identify sets of numerical attributes that behave similarly (with **itemset mining** [Prado et al. EGC'2012]):

- **(**) Records are sorted w.r.t every attributes \rightsquigarrow index ranking
- ② For all sets of attributes we compute the ratio of the number of pairs of records (among all possible pairs) for which all attributes in the set have the ranking index in the first record significantly different than in the second one → correlation ratio
- High ratio implies that attributes in the set behave similarly
 potentially interesting covariation patterns
- Extract all patterns with a ratio higher than a user defined minimum support threshold

Context 0000000000	Data mining 0●0000	Model 00000	Results and analysis	Conclusion 000
Looking for co-variation	patterns in data			
Co-variatio	n patterns			

Identify sets of numerical attributes that behave similarly (with **itemset mining** [Prado et al. EGC'2012]):

- **(**) Records are sorted w.r.t every attributes \rightsquigarrow index ranking
- ② For all sets of attributes we compute the ratio of the number of pairs of records (among all possible pairs) for which all attributes in the set have the ranking index in the first record significantly different than in the second one → correlation ratio
- High ratio implies that attributes in the set behave similarly ~> potentially interesting covariation patterns
- Extract all patterns with a ratio higher than a user defined minimum support threshold

Context 0000000000	Data mining 0●0000	Model 00000	Results and analysis	Conclusion 000
Looking for co-variation	patterns in data			
Co-variatio	n patterns			

Identify sets of numerical attributes that behave similarly (with **itemset mining** [Prado et al. EGC'2012]):

- **(**) Records are sorted w.r.t every attributes \rightsquigarrow index ranking
- ② For all sets of attributes we compute the ratio of the number of pairs of records (among all possible pairs) for which all attributes in the set have the ranking index in the first record significantly different than in the second one → correlation ratio
- High ratio implies that attributes in the set behave similarly
 potentially interesting covariation patterns
- Extract all patterns with a ratio higher than a user defined minimum support threshold

Context 0000000000	Data mining ○○●000	Model 00000	Results and analysis	Conclusion
Results				
Results				

```
Examples of covariation patterns
```

```
[CHLOROPHYTES+, CYANOBACTERIES-] 0.58
P_Ptot+,ch1_a+,EUGLENOPHYTES-,CHLOROPHYTES-,CYANOBACTERIES+#0.52536
...
[P-Ptot(mg/L)[c]+, CYANOBACTERIES+] 0.52
[P-Ptot(mg/L)[c]+, CHLOROPHYTES-] 0.52
[N-Ntot(mg/L)+, P-Ptot(mg/L)[c]+, ch1.a+] 0.55731
[N-Ntot(mg/L)+, ch1.a+, DINOPHYCEES+, CHLOROPHYTES-] 0.50988
[Month+, N-Ntot(mg/L)+, P-P04(mg/L)-, ch1.a+] 0.52964
...
```

Context 0000000000	Data mining ○○●000	Model 00000	Results and analysis 000	Conclusion
Results				
Results				

Examples of covariation patterns

```
[CHLOROPHYTES+, CYANOBACTERIES-] 0.58
P_Ptot+,ch1_a+,EUGLENOPHYTES-,CHLOROPHYTES-,CYANOBACTERIES+#0.52536
...
[P-Ptot(mg/L)[c]+, CYANOBACTERIES+] 0.52
[P-Ptot(mg/L)[c]+, CHLOROPHYTES-] 0.52
[N-Ntot(mg/L)+, P-Ptot(mg/L)[c]+, ch1.a+] 0.55731
[N-Ntot(mg/L)+, ch1.a+, DINOPHYCEES+, CHLOROPHYTES-] 0.50988
[Month+, N-Ntot(mg/L)+, P-P04(mg/L)-, ch1.a+] 0.52964
```

. . .

Warning

These are **not** rules . . . yet.

Context 0000000000	Data mining ○○●000	Model 00000	Results and analysis 000	Conclusion
Results				
Results				

Examples of covariation patterns

```
[CHLOROPHYTES+, CYANOBACTERIES-] 0.58
P_Ptot+,ch1_a+,EUGLENOPHYTES-,CHLOROPHYTES-,CYANOBACTERIES+#0.52536
...
[P-Ptot(mg/L)[c]+, CYANOBACTERIES+] 0.52
[P-Ptot(mg/L)[c]+, CHLOROPHYTES-] 0.52
[N-Ntot(mg/L)+, P-Ptot(mg/L)[c]+, ch1.a+] 0.55731
[N-Ntot(mg/L)+, ch1.a+, DINOPHYCEES+, CHLOROPHYTES-] 0.50988
[Month+, N-Ntot(mg/L)+, P-P04(mg/L)-, ch1.a+] 0.52964
...
```

We can consider all the ponds together (global behaviour) ... or specific ponds (local behaviour)

... or specific attributes (ecologically pertinent questions)

Context 0000000000	Data mining ○○○●○○	Model 00000	Results and analysis	Conclusion 000
Results				
Going furtl	her			

How to build a global view of covariation networks?

Context 00000000000	Data mining ○○○○●○	Model 00000	Results and analysis	Conclusion 000
Results				
Going furth	her			



Context	Data mining	Model	Results and analysis	Conclusion
	000000			
Beaulte				

Going further



- Identify pertinent attributes (e.g. all the chemicals that are positively correlated to an increase of Chlorophyll-A)
- Identify ecological relationships (e.g. during the periods when Cyanobacteria increase, how do the Euglenophytes behave?)
- Locate feedback loops
- Help building a computer model to explore a precise behaviour of the system

Context 0000000000	Data mining ○○○○○●	Model 00000	Results and analysis 000	Conclusion 000
Results				
Examples				

- Nitrogen and Phosphorus are positively linked in all ponds
- Nitrogen and Nitrates are negatively linked in most of the ponds
- The "Épansardières" pond is the most representative
- The "Aubergères" pond is the most singular
- Floating and submerged plants behave differently and are very representative of global pond state
- Green and blue algae behave differently and are very representative of pond states

Context 0000000000	Data mining ○○○○○●	Model 00000	Results and analysis 000	Conclusion 000
Results				
Examples				

- Nitrogen and Phosphorus are positively linked in all ponds
- Nitrogen and Nitrates are negatively linked in most of the ponds
- The "Épansardières" pond is the most representative
- The "Aubergères" pond is the most singular
- Floating and submerged plants behave differently and are very representative of pond states
- Green and blue algae behave differently and are very representative of pond states

Context 00000000000	Data mining 000000	Model ●0000	Results and analysis 000	Conclusic
Principles				
Vegetati	on			
Flo	ating:			
6 6	 Leaves spread atop the surface Very competitive for light Less competitive for nutrients 			
Sul	omerged:			
a a	Less competitivVery competiti	ve for light ve for nutrients		
_				
Phytople	ankton			

- Chlorophytes (green algae):
 - Faster reproduction
 - Less competitive for light
- Cyanobacteria (blue algae):
 - More competitive for light
 - Strong effect on water turbidity



Context		Data mining 000000	Mo:	del Doo	Results and analysis	Conclusion 000
Equatio	ns					
	Variation of population	=	x	x	-	
	concentration					
	$\frac{dS}{dt}$	=	х	x	-	
	dF dt	=	×	×	-	
	dG dt	=	x	x	-	
	<u>dB</u>	=	x	×	-	

- Basis nutrient : N in mg.L⁻
- Nutrients available for plants : $n_v = \frac{N}{1+a_f F + a_s S}$
- Nutrients available for phytoplankton : $n_p = n_v G B$
- Turbidity : $E = e_g G + e_b B$ in m^-
- Depth : D in m

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Equations				

Variation of population concentration	=	exponential growth rate	×	х	-	
dS dt	=	r _s S	×	×	-	
dF dt	=	r _f F	×	x	-	
dG dt	=	r _g G	x	x	-	
dB dt	=	<mark>r</mark> ьВ	x	x	-	

- Basis nutrient : N in $mg.L^{-1}$
- Nutrients available for plants : $n_v = \frac{N}{1+q_f F + q_s S}$
- Nutrients available for phytoplankton : $n_p = n_v G B$
- Turbidity : $E = e_g G + e_b B$ in m^-
- Depth : D in m

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Equations				

Variation of population concentration	=	exponential growth rate	×	competition for nutrients	×	-
$\frac{dS}{dt}$	=	r₅S	x	$\frac{n_V}{n_V + h_S}$	x	-
$\frac{dF}{dt}$	=	r _f F	×	$\frac{n_V}{n_V + h_f}$	x	-
dG dt	=	r _g G	x	$\frac{n_p}{n_p+h_g}$	x	-
<u>dB</u> dt	=	r _b B	×	$\frac{n_p}{n_p+h_h}$	×	-

- Basis nutrient : N in $mg.L^{-1}$
- Nutrients available for plants : $n_v = \frac{N}{1+q_f F + q_s S}$
- Nutrients available for phytoplankton : $n_p = n_v G B$
- Turbidity : $E = e_g G + e_b B$ in m^{-1}
- Depth : D in m

Context		Data 1 0000	mining 00		Model ○○●○○		Results and an 000	alysis	Conclusion 000
Equatio	ns								
	Variation of population concentration	=	exponential growth rate	x	competition for nutrients	x	competition for light	-	
	dS dt	=	r₅S	×	$\frac{n_V}{n_V+h_S}$	×	$\frac{1}{1+a_{S}S+bF+DE}$	-	
	dF dt	=	r _f F	x	$\frac{n_V}{n_V+h_f}$	x	$\frac{1}{1+a_f F}$	-	
	dG dt	=	r _g G	×	$\frac{n_P}{n_P+h_g}$	x	$rac{h_{sg}}{h_{sg} + DE + cF}$	-	
	<u>dB</u>	=	r _b B	x	$\frac{n_p}{p_p}$	x	h _{sb}	-	

- Basis nutrient : N in $mg.L^{-1}$
- Nutrients available for plants : $n_v = \frac{N}{1+q_f F + q_s S}$
- Nutrients available for phytoplankton : $n_p = n_v G B$
- Turbidity : $E = e_g G + e_b B$ in m^{-1}
- Depth : D in m

Context		Data 0000	mining 00		Model ○○●○○		Results and ar	alysis		Conclusion
Equatio	ns									
	Variation of population	=	exponential growth rate	×	competition for nutrients	x	competition for light	-	mortality rate	
	concentration									-
	dS dt	=	r₅S	x	$\frac{n_V}{n_V+h_S}$	x	$\frac{1}{1+a_SS+bF+DE}$	-	I ₅ S	
	dF dt	=	r _f F	x	$\frac{n_V}{n_V+h_f}$	x	$\frac{1}{1+a_fF}$	-	l _f F	
	dG dt	=	r _g G	×	$\frac{n_p}{n_p+h_g}$	x	$rac{h_{sg}}{h_{sg}+DE+cF}$	-	$(I_g + f)G$	
	dB dt	=	r _b B	×	$\frac{n_p}{n_p+h_b}$	x	$\frac{h_{sb}}{h_{sb}+DE+cF}$	-	$(l_b + f)B$	

- Basis nutrient : N in $mg.L^{-1}$
- Nutrients available for plants : $n_v = \frac{N}{1+q_f F + q_s S}$
- Nutrients available for phytoplankton : $n_p = n_v G B$
- Turbidity : $E = e_g G + e_b B$ in m^{-1}
- Depth : D in m

Context		Data 0000	mining 000		Model ○○●○○		Results and ar	alysis		Conclusion
Equatio	ns									
	Variation of population	=	exponential growth rate	×	competition for nutrients	×	competition for light	-	mortality rate	
	concentration									-
	dS dt	=	r₅S	x	$\frac{n_V}{n_V+h_S}$	x	$\frac{1}{1+a_SS+bF+DE}$	-	I _s S	
	dF dt	=	r _f F	x	$\frac{n_V}{n_V+h_f}$	x	$\frac{1}{1+a_fF}$	-	I _f F	
	$\frac{dG}{dt}$	=	r _g G	×	$\frac{n_p}{n_p+h_g}$	x	$rac{h_{sg}}{h_{sg}+DE+cF}$	-	$(I_g + f)G$	
	dB dt	=	r _b B	×	$\frac{n_p}{n_p+h_b}$	x	$\frac{h_{sb}}{h_{sb}+DE+cF}$	-	$(l_b + f)B$	

- Basis nutrient : N in $mg.L^{-1}$
- Nutrients available for plants : $n_v = \frac{N}{1+q_f F + q_s S}$
- Nutrients available for phytoplankton : $n_p = n_v G B$
- Turbidity : $E = e_g G + e_b B$ in m^{-1}
- Depth : D in m

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Model simulations				

- Progressive change in nutrient disponibility ~>> brutal shift of equilibrium
- Progressive increasing of available nutrients (N) (e.g. from 0 to 10 mg.L⁻¹ with 0.01 steps)
- Progressive decreasing of available nutrients (e.g. from 10 to 0 mg.L⁻¹ with 0.01 steps)
- Evolution on 200 days between every variation step (equilibrium)

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Model simulations				

- Progressive change in nutrient disponibility ~>> brutal shift of equilibrium
- Progressive increasing of available nutrients (N) (e.g. from 0 to 10 $mg.L^{-1}$ with 0.01 steps)
- Progressive decreasing of available nutrients (e.g. from 10 to 0 mg.L⁻¹ with 0.01 steps)
- Evolution on 200 days between every variation step (equilibrium)

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Model simulations				

- Progressive change in nutrient disponibility ~>> brutal shift of equilibrium
- Progressive increasing of available nutrients (N) (e.g. from 0 to 10 $mg.L^{-1}$ with 0.01 steps)
- Progressive decreasing of available nutrients (e.g. from 10 to $0 mg.L^{-1}$ with 0.01 steps)
- Evolution on 200 days between every variation step (equilibrium)

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Model simulations				

- Progressive change in nutrient disponibility ~>> brutal shift of equilibrium
- Progressive increasing of available nutrients (N) (e.g. from 0 to 10 mg.L⁻¹ with 0.01 steps)
- Progressive decreasing of available nutrients (e.g. from 10 to $0 mg.L^{-1}$ with 0.01 steps)
- Evolution on 200 days between every variation step (equilibrium)

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Implementation				

- Few available nutrients
 - Competition for nutrients
 - Advantage to submerged plants and chlorophytes
 - Clear state
- Many nutrients available
 - Competition for light
 - Advantage to floating plants and cyanobacterias
 - Turbid state

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Implementation				

- Few available nutrients
 - Competition for nutrients
 - Advantage to submerged plants and chlorophytes
 - Clear state
- Many nutrients available
 - Competition for light
 - Advantage to floating plants and cyanobacterias
 - Turbid state

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Implementation				

- Few available nutrients
 - Competition for nutrients
 - Advantage to submerged plants and chlorophytes
 - Clear state
- Many nutrients available
 - Competition for light
 - Advantage to floating plants and cyanobacterias
 - Turbid state

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Implementation				

- Few available nutrients
 - Competition for nutrients
 - Advantage to submerged plants and chlorophytes
 - Clear state
- Many nutrients available
 - Competition for light
 - Advantage to floating plants and cyanobacterias
 - Turbid state

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Implementation				

- Few available nutrients
 - Competition for nutrients
 - Advantage to submerged plants and chlorophytes
 - Clear state
- Many nutrients available
 - Competition for light
 - Advantage to floating plants and cyanobacterias
 - Turbid state

Context	Data mining	Model	Results and analysis	Conclusion
		00000		
Implementation				

- Few available nutrients
 - Competition for nutrients
 - Advantage to submerged plants and chlorophytes
 - Clear state
- Many nutrients available
 - Competition for light
 - Advantage to floating plants and cyanobacterias
 - Turbid state



Phytoplankton dynamics

• Values of N before point B: chlorophytes dominance





• Between A and B: two stable states depending on earlier system state



Vegetation dynamics

• Values of N before point B: submerged plants dominance



Vegetation dynamics

• Values of N after point A: floating plants dominance



Vegetation dynamics

 Between A and B: two stable states depending on earlier system state



Increasing of available nutrients N.


Decreasing of available nutrients N.



Before point D: only one equilibrium where submerged plants and chlorophytes dominate



After point B: only one equilibrium where floating plants and cyanobacteria dominate



Between C and A: these two equilibriums coexist



Between D and C and between A and B: third equilibrium where floating plants and chlorophytes dominate

Context	Data mining	Model	Results and analysis	Conclusion
				000
Assessment				

Alternative equilibriums

- Clear state with submerged plants and chlorophytes for low values of N
- Turbid state with floating plants and cyanobacteria for high values of N
- Third stable state with floating plants and chlorophytes
- Resistance of the submerged plants to the high depth (surprise!)
- Important effect of depth and spatial diffusion

Context	Data mining	Model	Results and analysis	Conclusion
				000
Assessment				

- Alternative equilibriums
- $\bullet\,$ Clear state with submerged plants and chlorophytes for low values of N
- Turbid state with floating plants and cyanobacteria for high values of N
- Third stable state with floating plants and chlorophytes
- Resistance of the submerged plants to the high depth (surprise!)
- Important effect of depth and spatial diffusion

Context	Data mining	Model	Results and analysis	Conclusion
				000
Assessment				

- Alternative equilibriums
- Clear state with submerged plants and chlorophytes for low values of N
- Turbid state with floating plants and cyanobacteria for high values of N
- Third stable state with floating plants and chlorophytes
- Resistance of the submerged plants to the high depth (surprise!)
- Important effect of depth and spatial diffusion

Context	Data mining	Model	Results and analysis	Conclusion
				000
Assessment				

- Alternative equilibriums
- Clear state with submerged plants and chlorophytes for low values of N
- Turbid state with floating plants and cyanobacteria for high values of N
- Third stable state with floating plants and chlorophytes
- Resistance of the submerged plants to the high depth (surprise!)
- Important effect of depth and spatial diffusion

Context	Data mining	Model	Results and analysis	Conclusion
				000
Assessment				

- Alternative equilibriums
- $\bullet\,$ Clear state with submerged plants and chlorophytes for low values of N
- Turbid state with floating plants and cyanobacteria for high values of N
- Third stable state with floating plants and chlorophytes
- Resistance of the submerged plants to the high depth (surprise!)
- Important effect of depth and spatial diffusion

Context	Data mining	Model	Results and analysis	Conclusion
				000
Assessment				

- Alternative equilibriums
- $\bullet\,$ Clear state with submerged plants and chlorophytes for low values of N
- Turbid state with floating plants and cyanobacteria for high values of N
- Third stable state with floating plants and chlorophytes
- Resistance of the submerged plants to the high depth (surprise!)
- Important effect of depth and spatial diffusion

Context	Data mining	Model	Results and analysis	Conclusion
				000
Prospects				

- Parameterize the model wrt. pond data (iRace, dedicated measurements occurring right now...)
- Incorporate more parameters into the model
- Study the impact of bottom topology
- Consider the effect of regular drainings
- Generate optimal policies considering agriculture and fish farming constraints
- How to influence policy makers and/or farmers? ()

Context	Data mining	Model	Results and analysis	Conclusion
				000
Prospects				

- Parameterize the model wrt. pond data (iRace, dedicated measurements occurring right now...)
- Incorporate more parameters into the model
- Study the impact of bottom topology
- Consider the effect of regular drainings
- Generate optimal policies considering agriculture and fish farming constraints
- How to influence policy makers and/or farmers? (

Context	Data mining	Model	Results and analysis	Conclusion
				000
Prospects				

- Parameterize the model wrt. pond data (iRace, dedicated measurements occurring right now...)
- Incorporate more parameters into the model
- Study the impact of bottom topology
- Consider the effect of regular drainings
- Generate optimal policies considering agriculture and fish farming constraints
- How to influence policy makers and/or farmers? (!)

Context	Data mining	Model	Results and analysis	Conclusion
				000
Prospects				

- Parameterize the model wrt. pond data (iRace, dedicated measurements occurring right now...)
- Incorporate more parameters into the model
- Study the impact of bottom topology
- Consider the effect of regular drainings
- Generate optimal policies considering agriculture and fish farming constraints
- How to influence policy makers and/or farmers? (!)

Context	Data mining	Model	Results and analysis	Conclusion
				000
Prospects				

- Parameterize the model wrt. pond data (iRace, dedicated measurements occurring right now...)
- Incorporate more parameters into the model
- Study the impact of bottom topology
- Consider the effect of regular drainings
- Generate optimal policies considering agriculture and fish farming constraints
- How to influence policy makers and/or farmers? (!)

Context	Data mining	Model	Results and analysis	Conclusion
				000
Prospects				

- Parameterize the model wrt. pond data (iRace, dedicated measurements occurring right now...)
- Incorporate more parameters into the model
- Study the impact of bottom topology
- Consider the effect of regular drainings
- Generate optimal policies considering agriculture and fish farming constraints
- How to influence policy makers and/or farmers? (!)

Context	Data mining	Model	Results and analysis	Conclusion
				000
Prospects				



Thank you for your attention!