Geographic Information Systems & Statistics

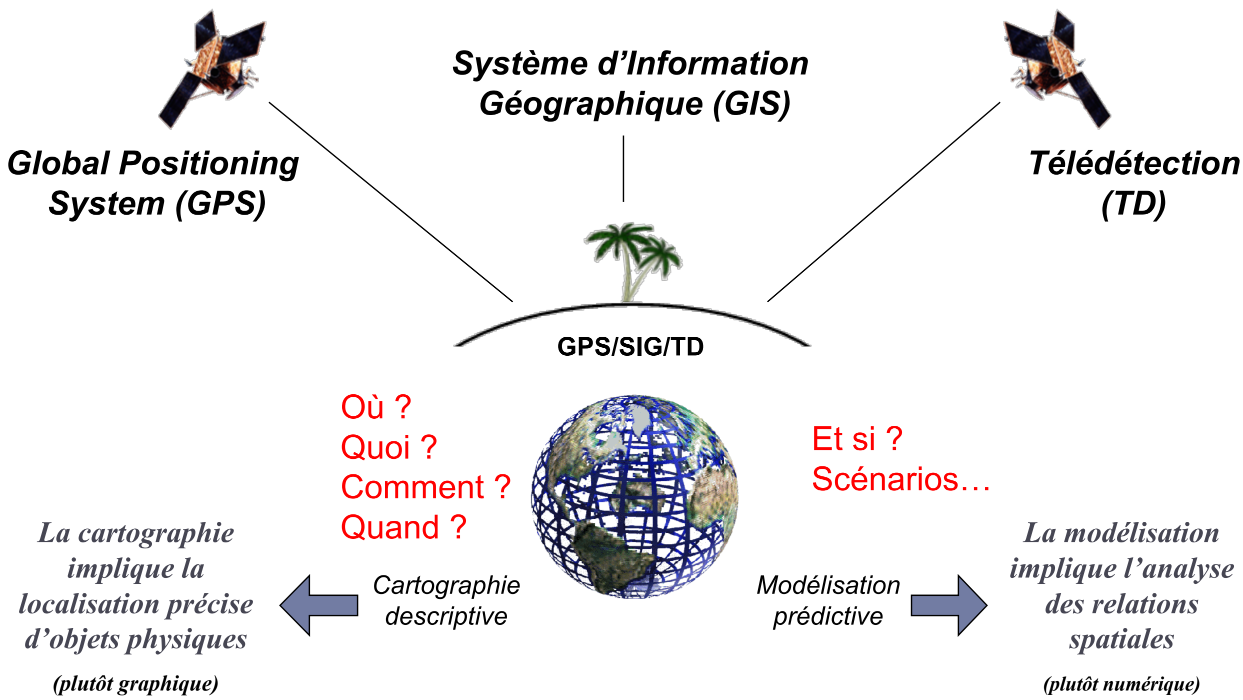
Anthony Lehmann

Nov. 21 2017

Déjeuner R

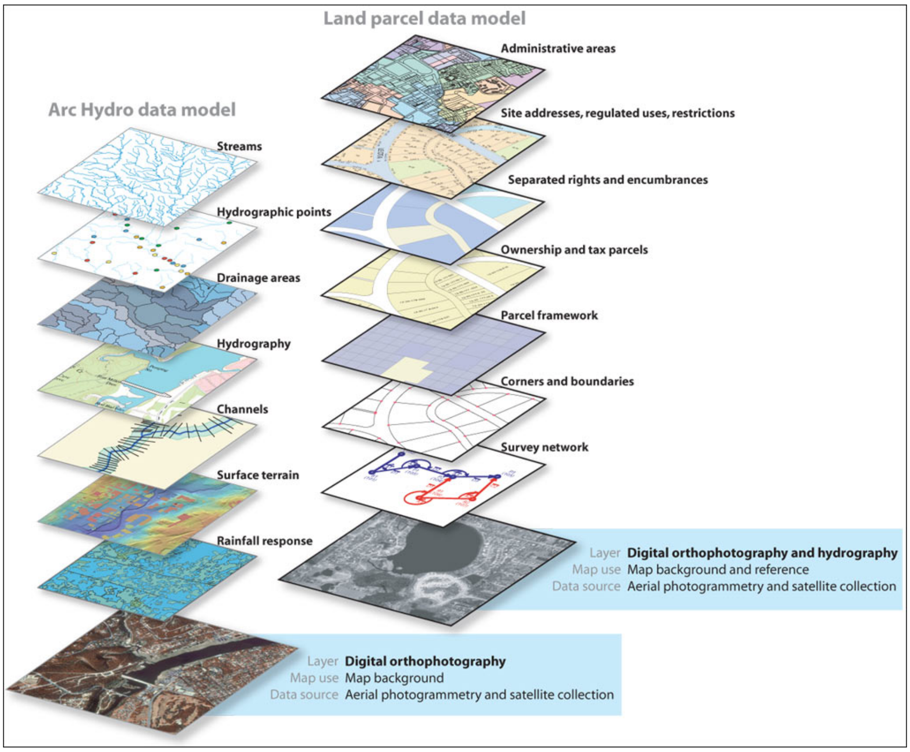
UNIVERSITÉ DE GENÈVE

Géotechnologies



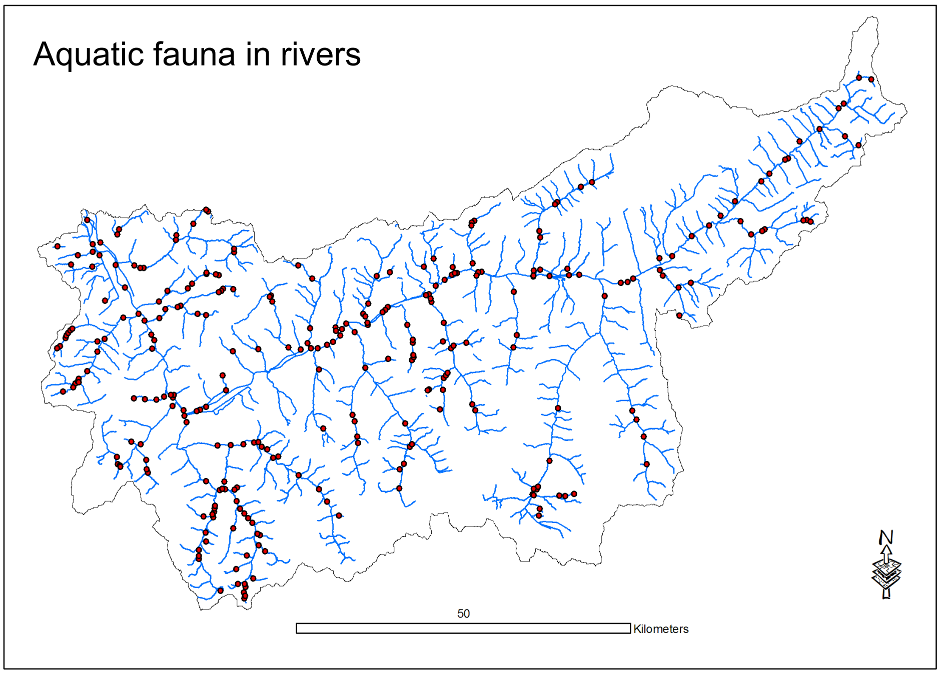
UNIVERSITÉ DE GENÈVE

SIG: Une modélisation du monde réel



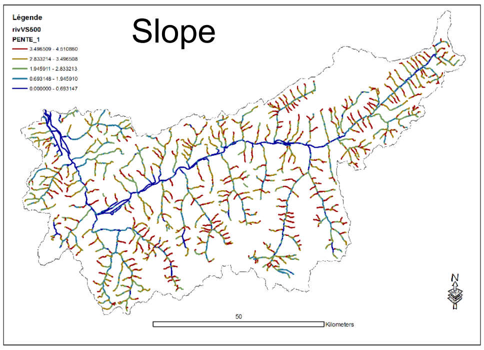
UNIVERSITÉ DE GENÈVE

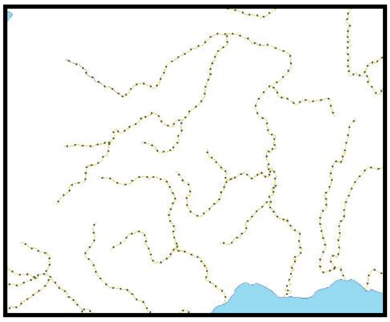
Introduction: Swiss rivers in Valais region

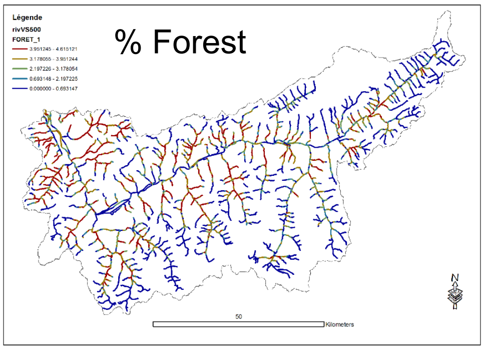


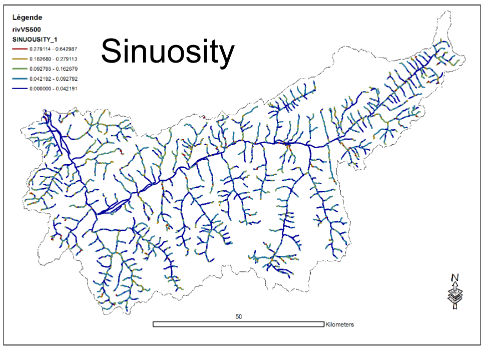
UNIVERSITÉ DE GENÈVE

Introduction : 500m segments of rivers

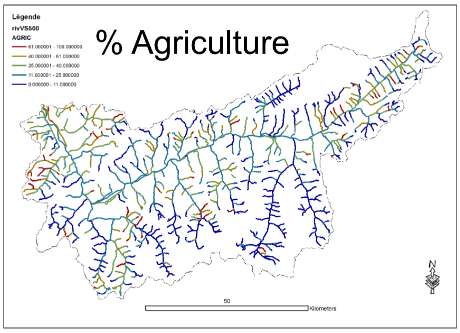


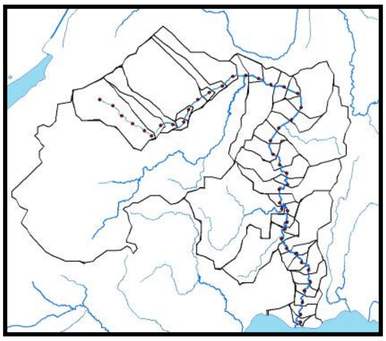


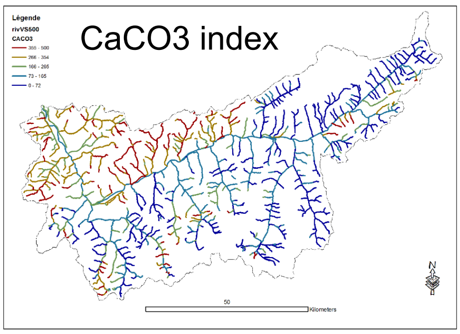


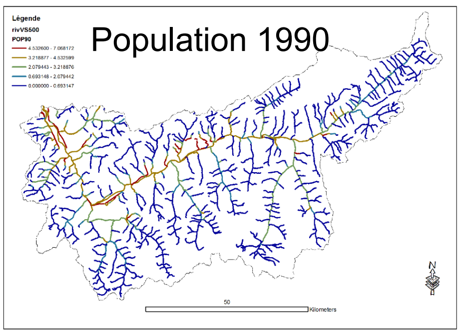


Introduction: watersheds every 1000m

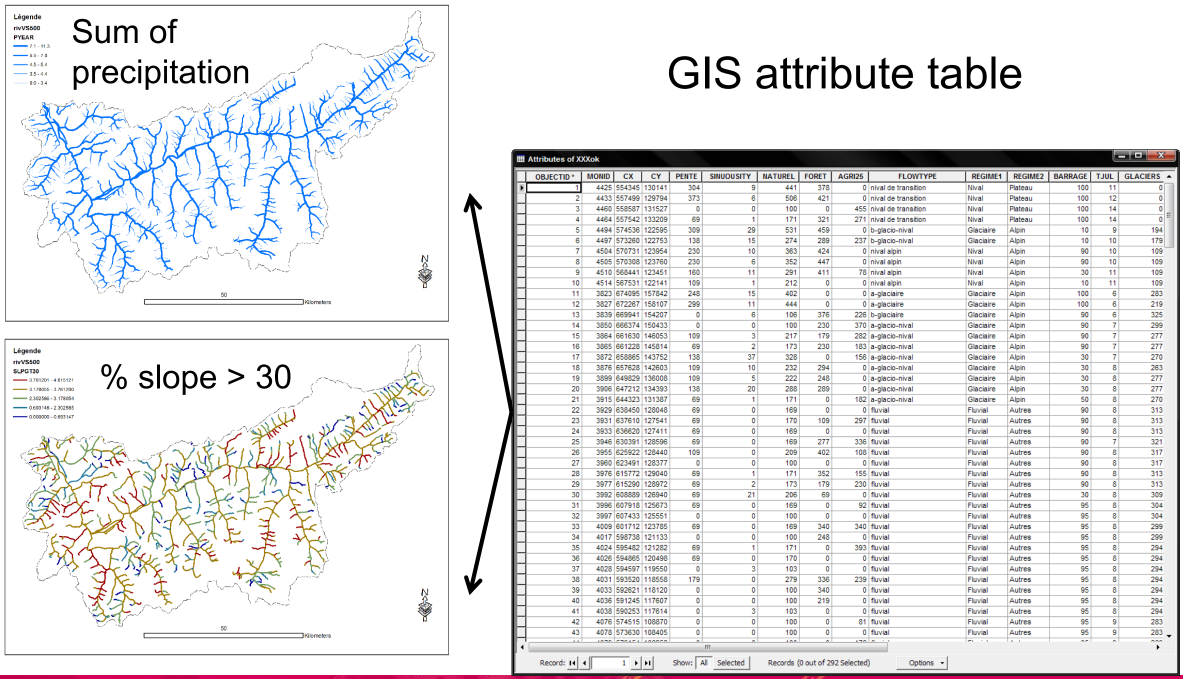






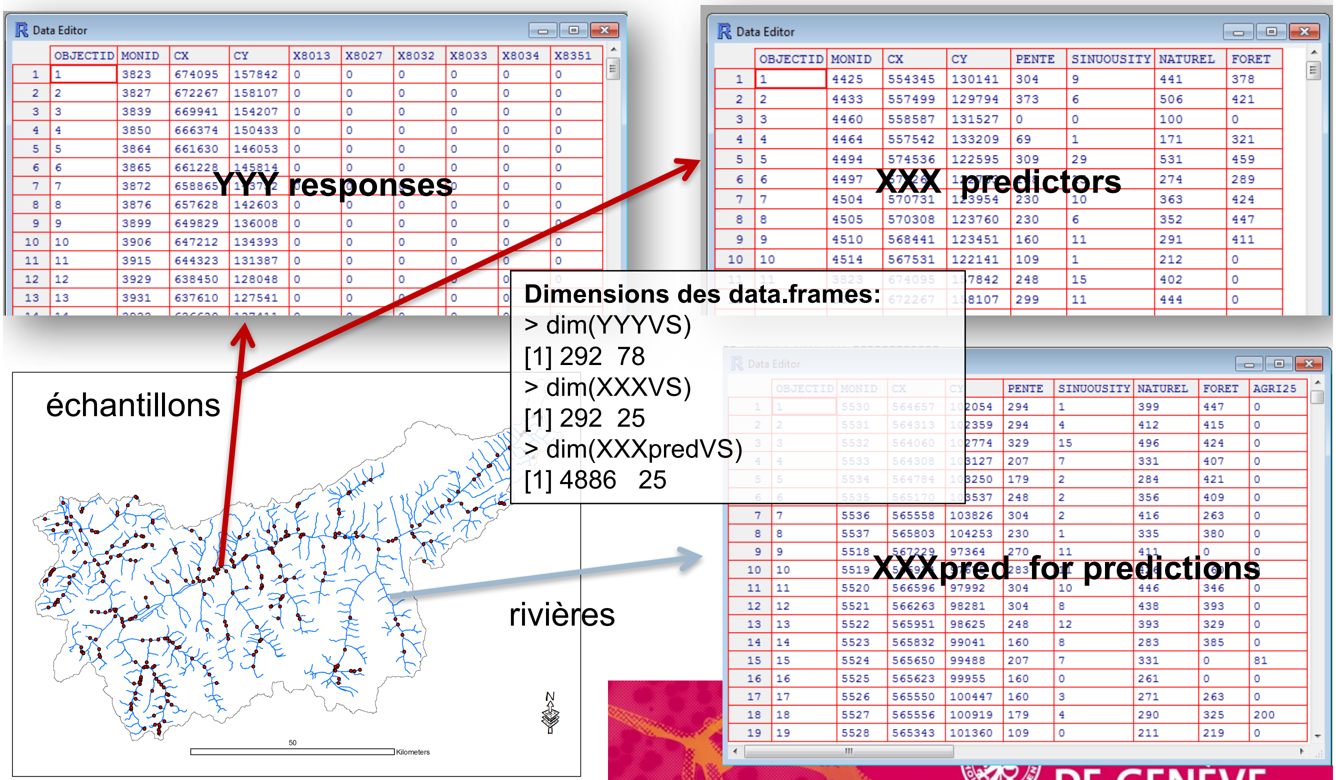


Introduction



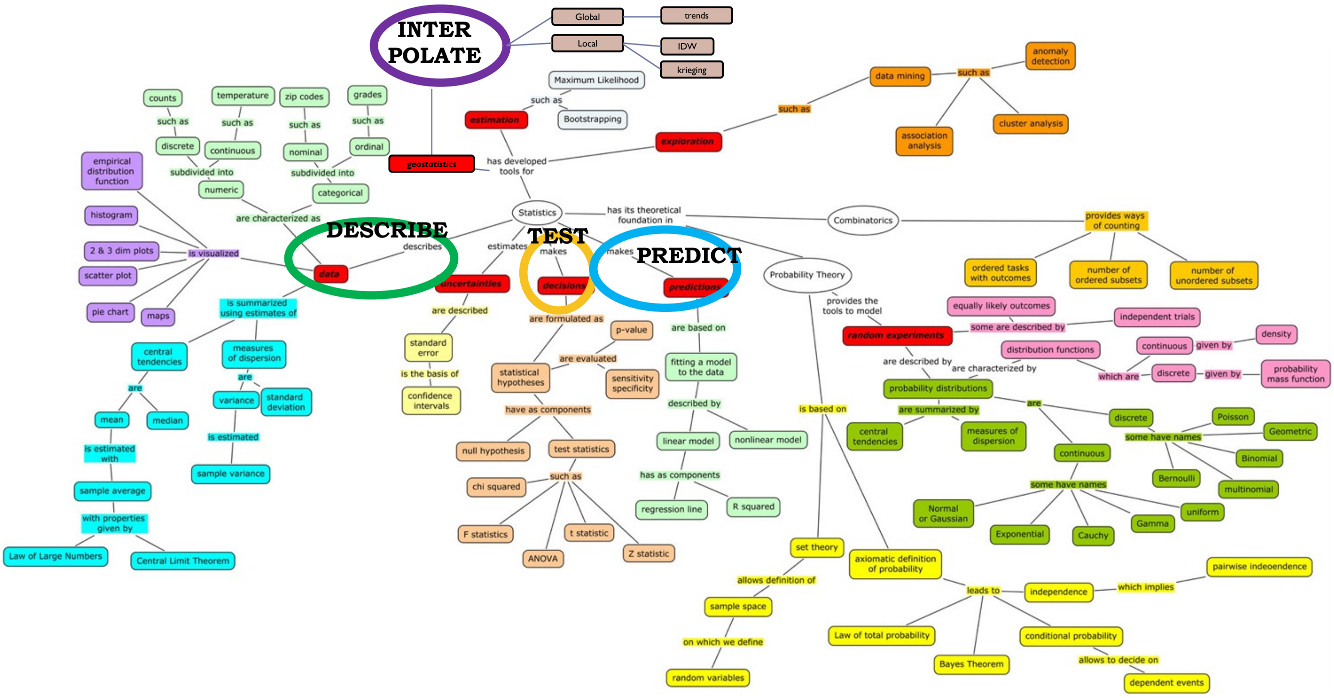
UNIVERSITÉ DE GENÈVE

Data preparation



DE GENÈVE

What can we do with statistics?



Test…

Data Editor

YYY responses

	OBJECTID	MONID	CX	CY	X8013	X8027	X8032	X8033	X8034	X8351
1	1	3823	674095	157842	0	0	0	0	0	0
2	2	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	0
3	3	3839	669941	154207	0	0	0	0	0	0
4	4	3850	666374	150433	0	0	0	0	0	0
5	5	3864	661630	146053	0	0	0	0	0	0
6	6	3865	661228	145814	0	0	0	0	0	0
7	7	3872	658865	143752	0	0	0	0	0	0
8	8	3876	657628	142603	0	0	0	0	0	0
9	9	3899	649829	136008	0	0	0	0	0	0
10	10	3906	647212	134393	0	0	0	0	0	0
11	11	3915	644323	131387	0	0	0	0	0	0
12	12	3929	638450	128048	0	0	0	0	0	0
13	13	3931	637610	127541	0	0	0	0	0	0

Data Editor

XXX predictors

	OBJECTID	MONID	CX	CY	PENTE	SINUOUSITY	NATUREL	FORET
1	[illegible]	[illegible]	[illegible]	[illegible]	304	9	441	378
2	2	4433	557499	129794	373	6	506	421
3	3	4460	558587	131527	0	0	100	0
4	4	4464	557542	133209	69	1	171	321
5	5	4494	574536	122595	309	29	531	459
6	6	4497	573260	122753	138	15	274	289
7	7	4504	570731	123954	230	10	363	424
8	8	4505	570308	123760	230	6	352	447
9	9	4510	568441	123451	160	11	291	411
10	10	4514	567531	122141	109	1	212	0
11	11	3823	674095	157842	248	15	402	0
12	12	3827	672267	158107	299	11	444	0

GOF

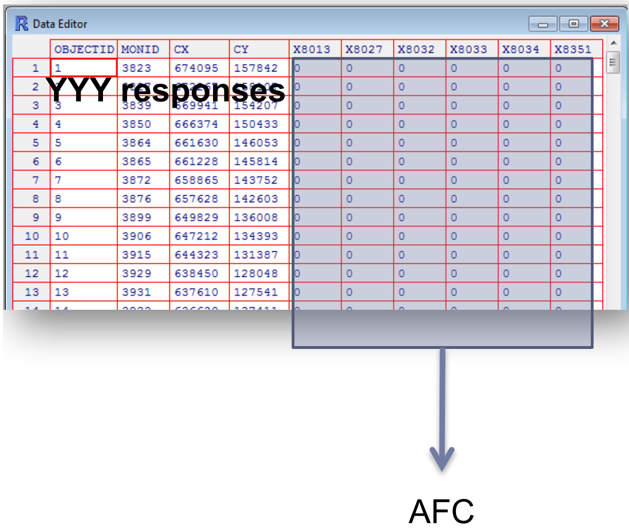
Ho: normality

Parametrical — Non parametrical

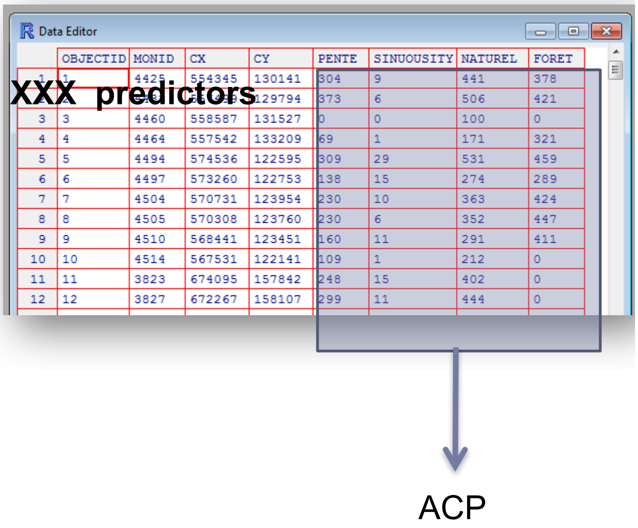
Diff. of means btw 2 groups: Test t — Ho: m1=m2 — Wilcoxon

Diff. of means btw more than 2 groups : Anova — Ho: m1=m2=m3… — Kruskal

Describe…

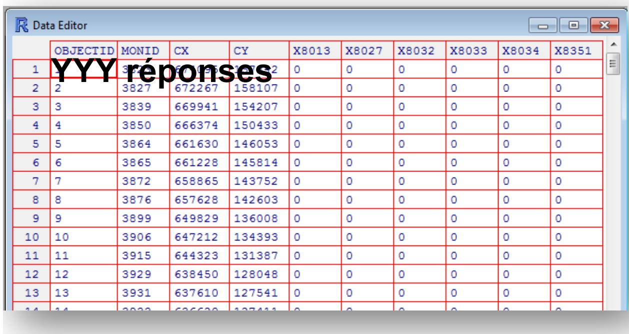


	OBJECTID	MONID	CX	CY	X8013	X8027	X8032	X8033	X8034	X8351
1	1	3823	674095	157842	0	0	0	0	0	0
2	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0
3	3	3839	669941	154207	0	0	0	0	0	0
4	4	3850	666374	150433	0	0	0	0	0	0
5	5	3864	661630	146053	0	0	0	0	0	0
6	6	3865	661228	145814	0	0	0	0	0	0
7	7	3872	658865	143752	0	0	0	0	0	0
8	8	3876	657628	142603	0	0	0	0	0	0
9	9	3899	649829	136008	0	0	0	0	0	0
10	10	3906	647212	134393	0	0	0	0	0	0
11	11	3915	644323	131387	0	0	0	0	0	0
12	12	3929	638450	128048	0	0	0	0	0	0
13	13	3931	637610	127541	0	0	0	0	0	0

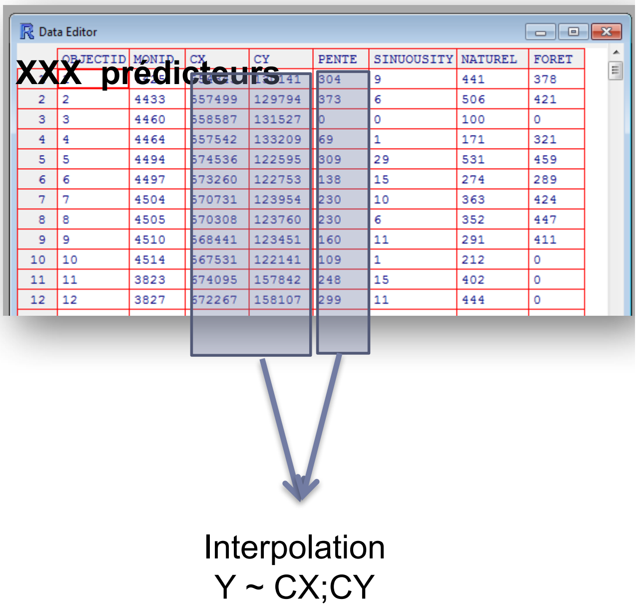


	OBJECTID	MONID	CX	CY	PENTE	SINUOUSITY	NATUREL	FORET
1	1	4425	554345	130141	304	9	441	378
2	2	[illegible]	[illegible]	129794	373	6	506	421
3	3	4460	558587	131527	0	0	100	0
4	4	4464	557542	133209	69	1	171	321
5	5	4494	574536	122595	309	29	531	459
6	6	4497	573260	122753	138	15	274	289
7	7	4504	570731	123954	230	10	363	424
8	8	4505	570308	123760	230	6	352	447
9	9	4510	568441	123451	160	11	291	411
10	10	4514	567531	122141	109	1	212	0
11	11	3823	674095	157842	248	15	402	0
12	12	3827	672267	158107	299	11	444	0

Interpolate…

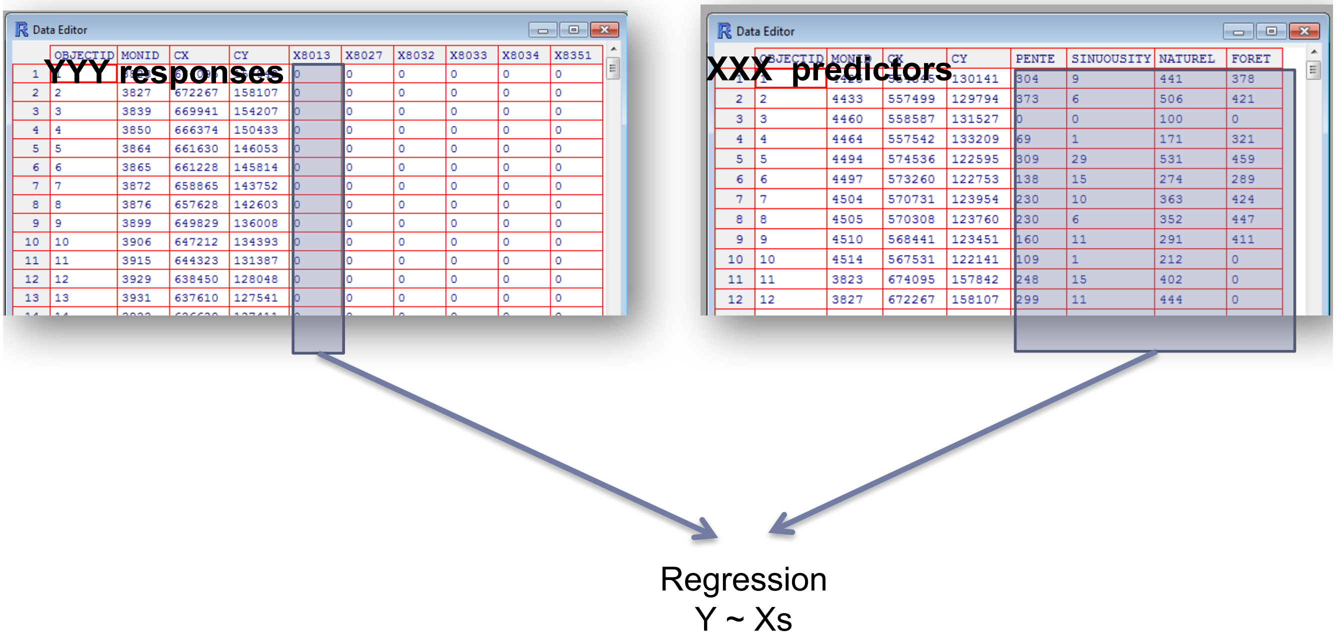


	OBJECTID	MONID	CX	CY	X8013	X8027	X8032	X8033	X8034	X8351
1	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0
2	2	3827	672267	158107	0	0	0	0	0	0
3	3	3839	669941	154207	0	0	0	0	0	0
4	4	3850	666374	150433	0	0	0	0	0	0
5	5	3864	661630	146053	0	0	0	0	0	0
6	6	3865	661228	145814	0	0	0	0	0	0
7	7	3872	658865	143752	0	0	0	0	0	0
8	8	3876	657628	142603	0	0	0	0	0	0
9	9	3899	649829	136008	0	0	0	0	0	0
10	10	3906	647212	134393	0	0	0	0	0	0
11	11	3915	644323	131387	0	0	0	0	0	0
12	12	3929	638450	128048	0	0	0	0	0	0
13	13	3931	637610	127541	0	0	0	0	0	0



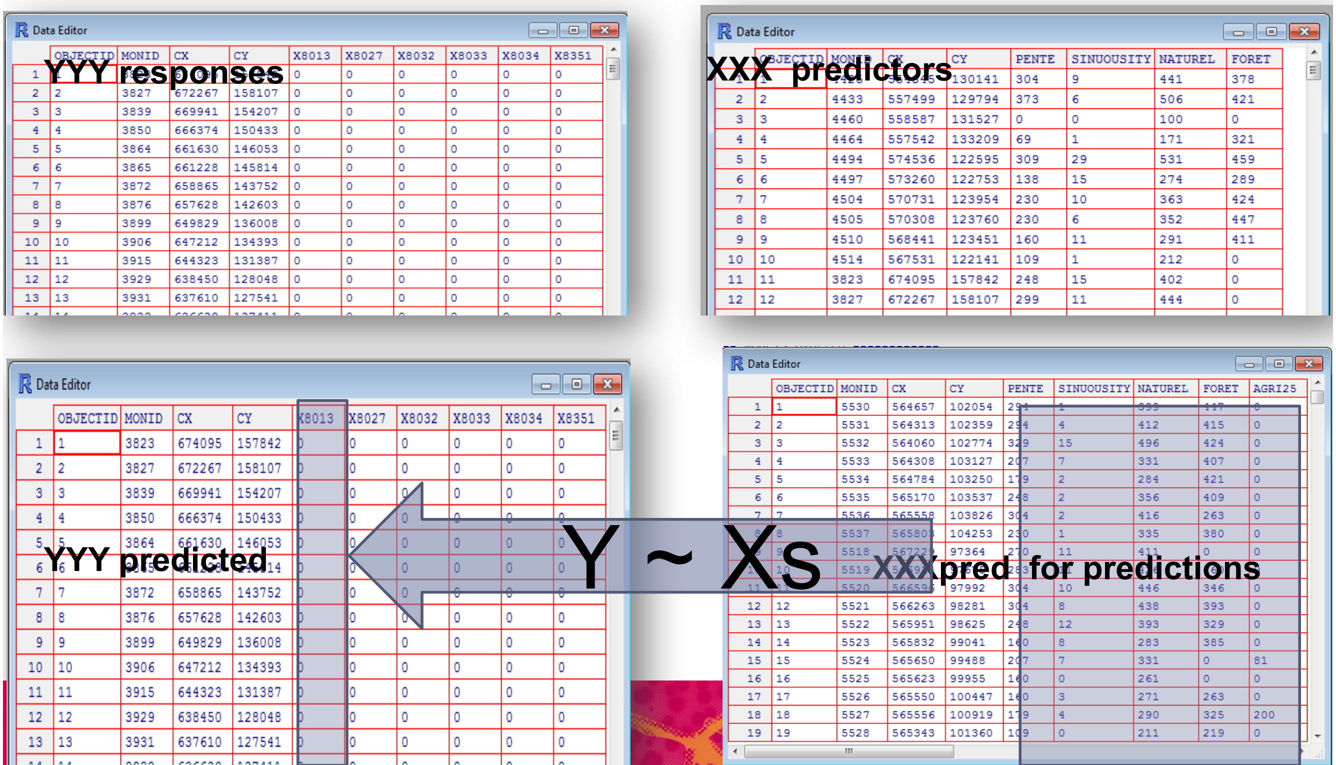
	OBJECTID	MONID	CX	CY	PENTE	SINUOUSITY	NATUREL	FORET
1	[illegible]	[illegible]	[illegible]	[illegible]	304	9	441	378
2	2	4433	557499	129794	373	6	506	421
3	3	4460	558587	131527	0	0	100	0
4	4	4464	557542	133209	69	1	171	321
5	5	4494	574536	122595	309	29	531	459
6	6	4497	573260	122753	138	15	274	289
7	7	4504	570731	123954	230	10	363	424
8	8	4505	570308	123760	230	6	352	447
9	9	4510	568441	123451	160	11	291	411
10	10	4514	567531	122141	109	1	212	0
11	11	3823	674095	157842	248	15	402	0
12	12	3827	672267	158107	299	11	444	0

Predict…



UNIVERSITÉ DE GENÈVE

Predict…



DE GENEVE

Predict: Predictive statistics

a) **Simple Linear Regression (LR)**
b) **Multiple Linear Regression (MLR)**
c) **Generalized Linear Models (GLM)**
d) **Generalized Additive Models(GAM)**
e) **…**

Regression

What is the purpose of regression models?

1 / To describe the relationship between a so-called dependent variable (or explained variable) and one or more dependent variables (or explanatory variables)

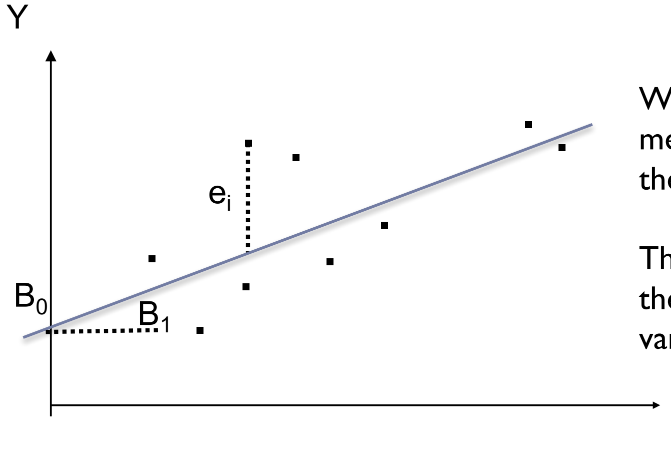
2 / To test the contribution of each variable to the explanation of the dependent variable

3 / To predict the value of the dependent variable as a function of the independent variables

UNIVERSITÉ DE GENÈVE

Simple Linear Regression (LR)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1,\ldots,n \;\; observations$$



We estimate β_0 and β_1 by the least squares method which consists in minimizing the sum of the squares of the residuals.

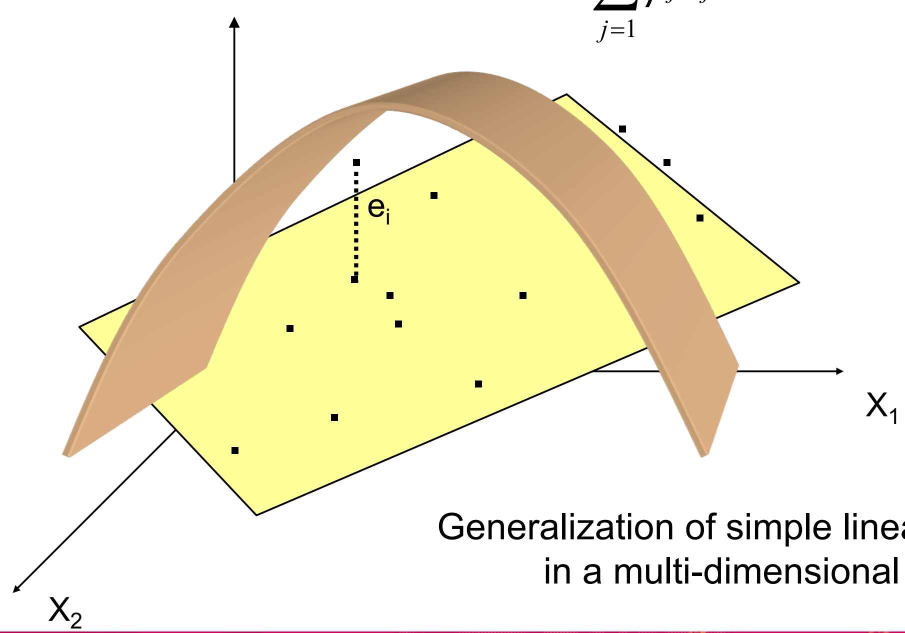
This method relies on a normal distribution of the errors and therefore of the explained variable Y

$$\sum_{i=1}^{n}(Yi - \overline{Y})^2 = \sum_{i=1}^{n}(\hat{Y}i - \overline{Y})^2 + \sum_{i=1}^{n}(Yi - \hat{Y}i)^2$$

$$SCtotal = SCrégression + SCrésidus$$

Multiple linear regression (MLR)

$$Y_i = \alpha + \sum_{j=1}^{p} \beta_j X_{j\,i} + \varepsilon_i \qquad i = 1,\ldots,n \,/\, j = 1,\ldots,p$$



Generalization of simple linear regression in a multi-dimensional space

General Linear Models (GLM)

$$g(\mu) \quad = \quad \alpha + \sum_{j=1}^{p} \beta_j x_{ji} + \varepsilon_i$$

Generalisation of regression models for other distribution:

- multiple regression becomes a special case of GLM
- use of a link function g() to transform the mean of Y in a plane defined by the explanatory variables
- The GLM thus make it possible to model data whose distribution is not Normal (binomial, Poisson, ...)
- Maximum of Likelihood for estimating the parameters

	Normal	**Binomial**	**Poisson**
Distribution	continuous	discrete	discrete
Link functions	μ	Log{μ/(1-μ)}	Log(μ)
Utilisation	Similar to LR	presence/absence	counts

General Additive Models (GAM)

$$g(\mu) \quad = \quad \alpha + \sum_{j=1}^{p} f_j(x_{ji}) + \varepsilon_i$$

Non Parametric Extension of GLM:

The form of the response curves of the explanatory variables depends directly on the data and not on a priori parametric model (X, X2, X3 ...).

- Use of smoothing function f() to establish the response curves
- Backfitting algorithm to find the shape of the response curves

UNIVERSITÉ DE GENÈVE

Synthesis

The regression aims at modeling the linear relationship between an explained variable Y and one or more explanatory variables $X_1, X_2, \ldots, X_n$

	LMR	GLM	GAM
Distribution of Y	Normal	Normal, binomial, Poisson	Normal, binomial, Poisson
Response curves	Parametrical ($X, X^2, X^3 \ldots$)	Parametrical ($X, X^2, X^3 \ldots$)	Non parametrical, any shape
Origines	<1900	1970	1986

$$Y_i = \alpha + \sum_{j=1}^{p} \beta_j X_{ji} + \varepsilon_i \qquad i = 1, \ldots, n$$

$$g(\mu) = \alpha + \sum_{j=1}^{p} \beta_j x_{ji} + \varepsilon_i$$

$$g(\mu) = \alpha + \sum_{j=1}^{p} f_j(x_{ji}) + \varepsilon_i$$

Conclusions

What is a good model ?

Accuracy: Although this is the main objective, it should not be forgotten that we always succeed in getting closer to the variable explained if we add enough explanatory variables

Simplicity: according to a principle of parsimony we always try to use the least possible explanatory variables in order to extract the main structures of the dataset

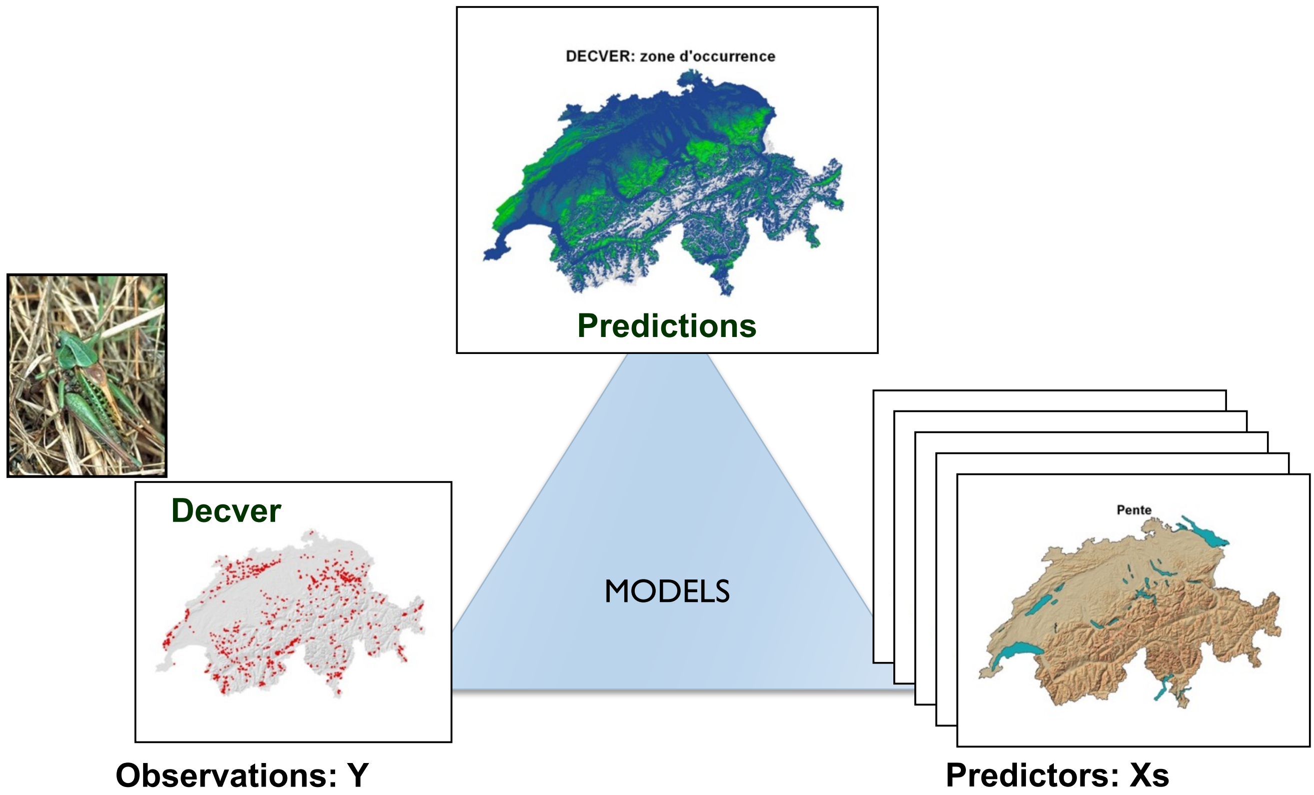
Generalization: A good model is also a model that can be used with new data from another spatial region or another period.

>> Modeling is an art and we must admit that all models are false and that one seeks the least false!

APPLICATIONS

- as a tool for understanding the processes studied
- as a diagnostic tool by comparison between observation and prediction
- as a tool for predicting the future state of a system, in order to provide scenarios for managers

Exemple d'application en Ecologie: Species Distribution Modeling



UNIVERSITÉ DE GENÈVE

Comparing methods

Novel methods improve prediction of species' distributions from occurrence data

ECOGRAPHY 29: 129–151, 2006

Jane Elith*, Catherine H. Graham*, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann, Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E. Schapire, Jorge Soberón, Stephen Williams, Mary S. Wisz and Niklaus E. Zimmermann

Method	Class of model, and explanation	Data[1]	Software	Std errors?[2]	Contact person
BIOCLIM	envelope model	p	DIVA-GIS	no	CG, RH
BRT	boosted decision trees	pa	R, gbm package	no	JE
BRUTO	regression, a fast implementation of a gam	pa	R and Splus, mda package	yes	JE
DK-GARP	rule sets from genetic algorithms; desktop version	pa	DesktopGarp	no	ATP
DOMAIN	multivariate distance	p	DIVA-GIS	no	CG, RH
GAM	regression: generalised additive model	pa	S-Plus, GRASP add-on	yes	AG,AL,JE
GDM	generalised dissimilarity modelling; uses community data	pacomm	Specialized program not general released; uses Arcview and Splus	no	SF
GDM-SS	generalised dissimilarity modelling; implementation for single species	pa	as for GDM	no	SF
GLM	regression; generalised linear model	pa	S-Plus, GRASP add-on	yes	AG,AL,JE
LIVES	multivariate distance	p	Specialized program not general released	no	JLi
MARS	regression; multivariate adaptive regression splines	pa	R, mda package plus new code to handle binomial responses	yes	JE, FH
MARS-COMM	as for MARS, but implemented with community data	pacomm	as for MARS	yes	JE
MARS-INT	as or MARS; interactions allowed	pa	as for MARS	yes	JE
MAXENT	maximum entropy	pa	Maxent	no	SP
MAXENT-T	maximum entropy with threshold features	pa	Maxent	no	SP
OM-GARP	rule sets derived with genetic algorithms; open modeller version	pa	new version of GARP not yet available	no	ATP

Elith et al. 2006

UNIVERSITÉ DE GENÈVE

Comparing methods

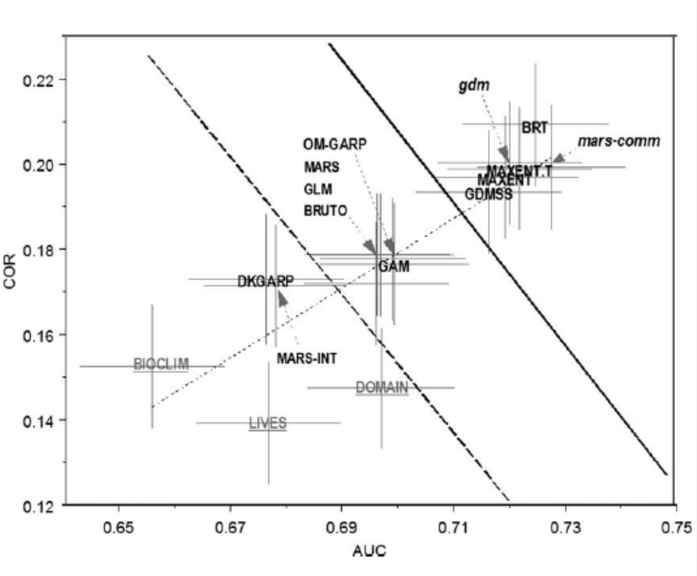
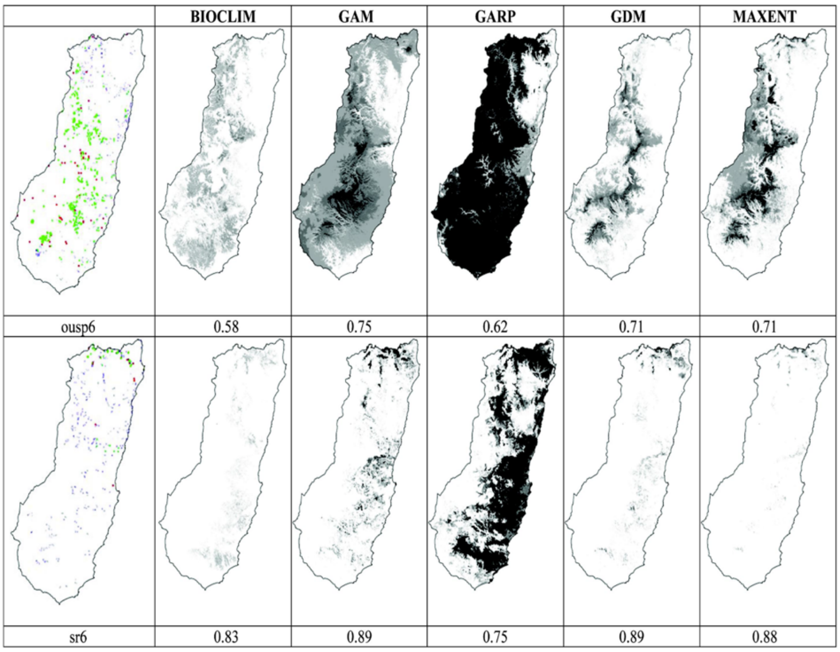


Fig. 3. Mean AUC vs mean correlation (COR) for modelling methods, summarised across all species. The grey bars are standard errors estimated in the GLMM (see Appendix), reflecting variation for an average species in an average region. The labels are broad classifications of the methods: grey underlined = only use presence data, black capitals = use presence and background samples, black lower case italics = community methods.

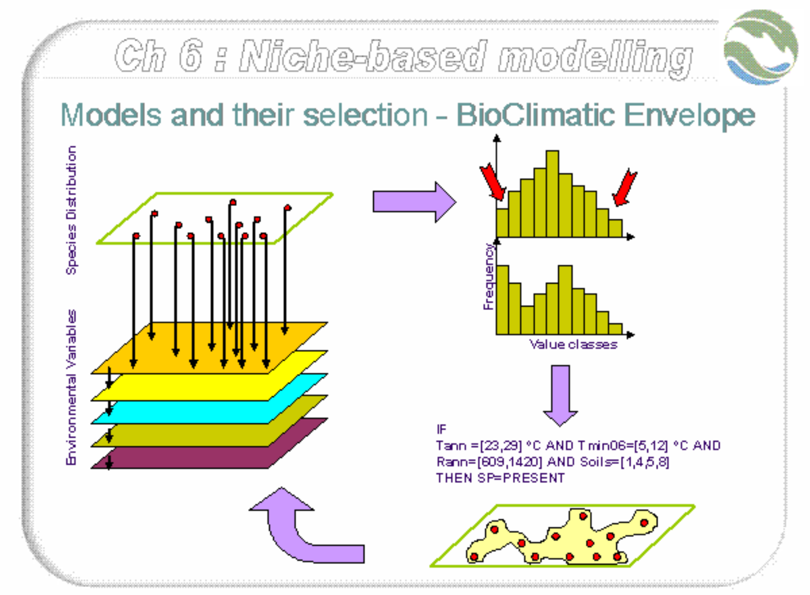
Elith et al. 2006

Comparing methods



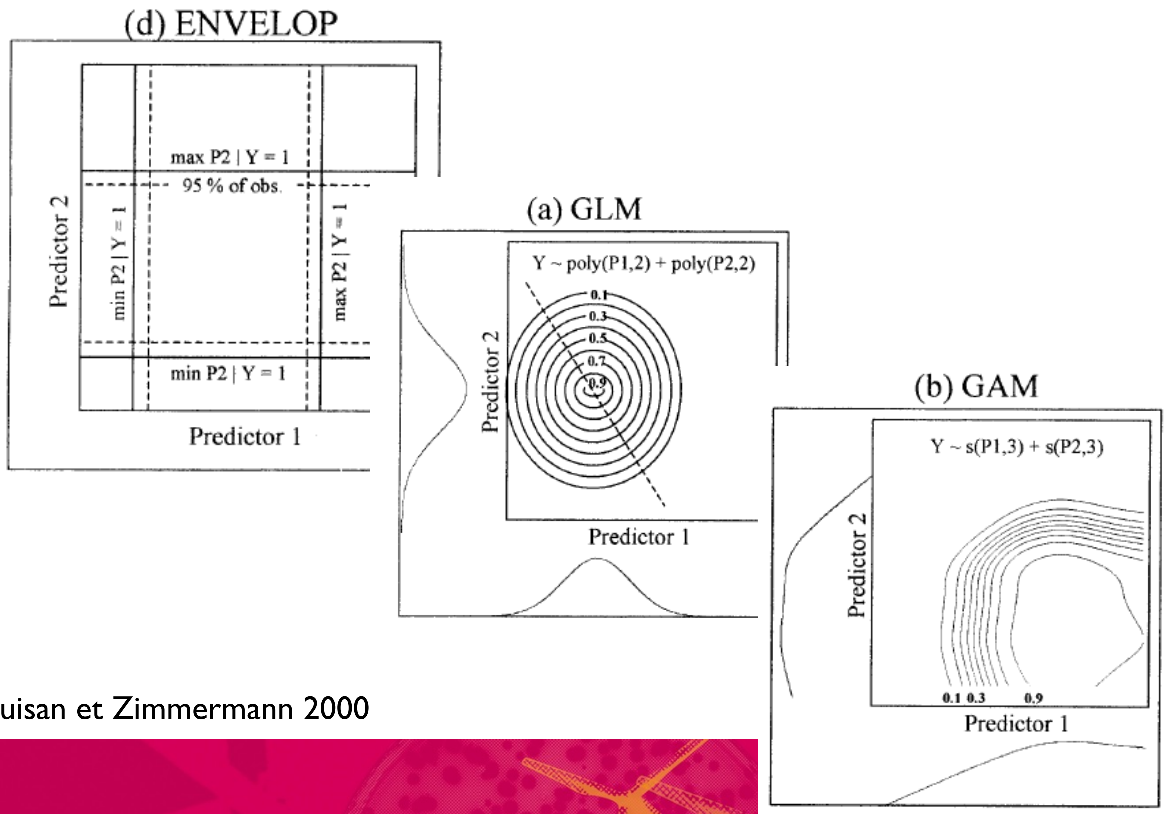
Elith et al. 2006

BioClim envelop



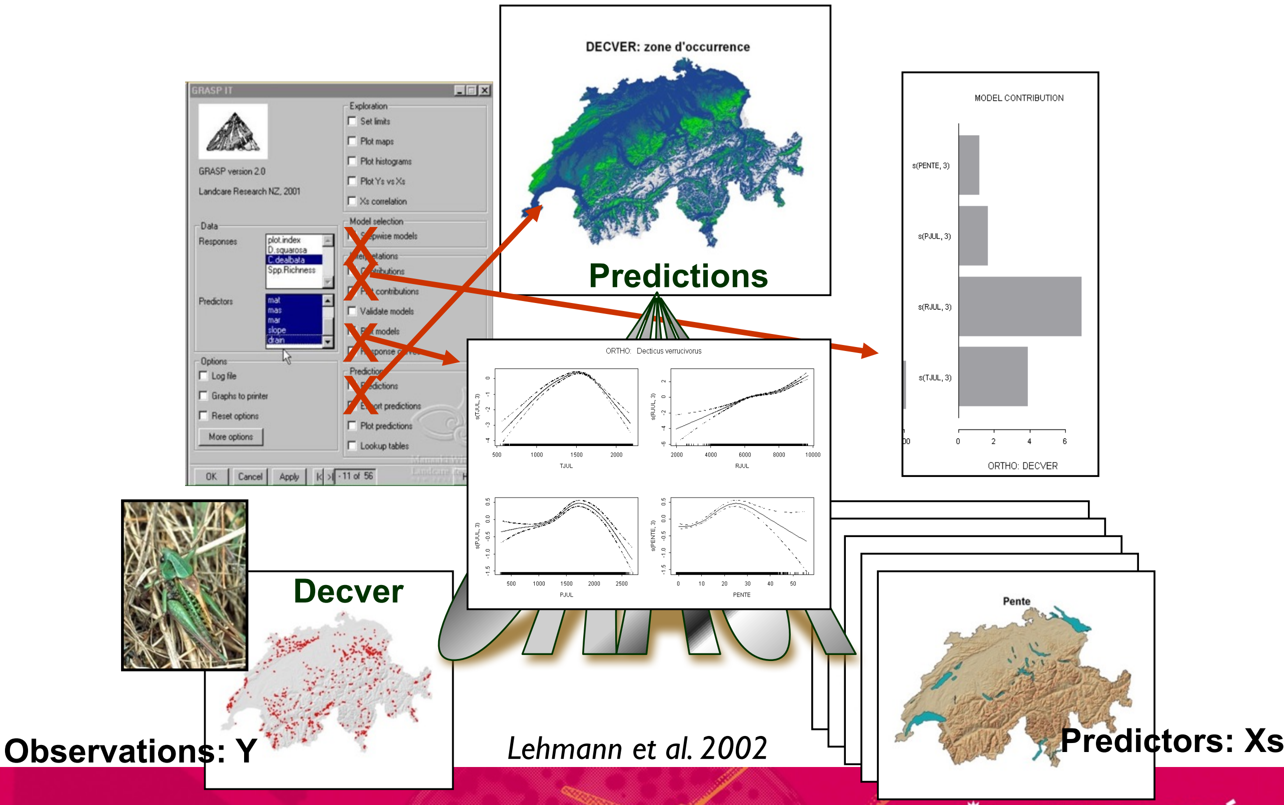
http://planet.botany.uwc.ac.za/nisl/Climate_change/page_198.htm

Generalized Linear Models



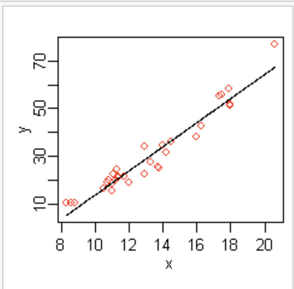
Guisan et Zimmermann 2000

Generalized Additive Models

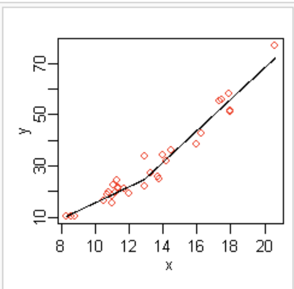


Lehmann et al. 2002

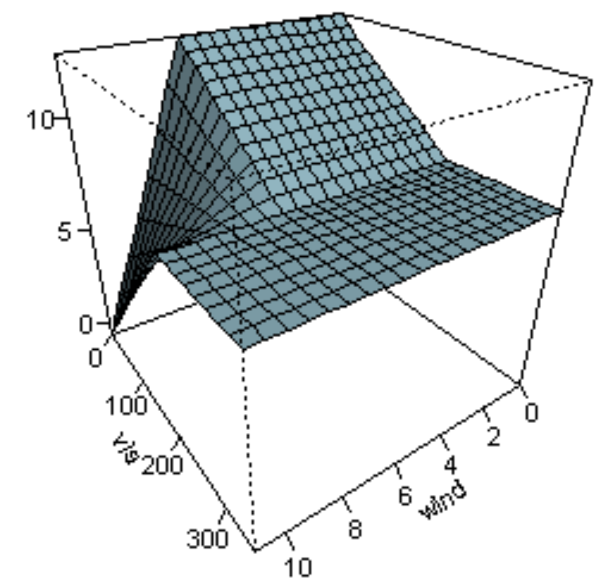
Multivariate Adaptive Regression Splines



A linear model

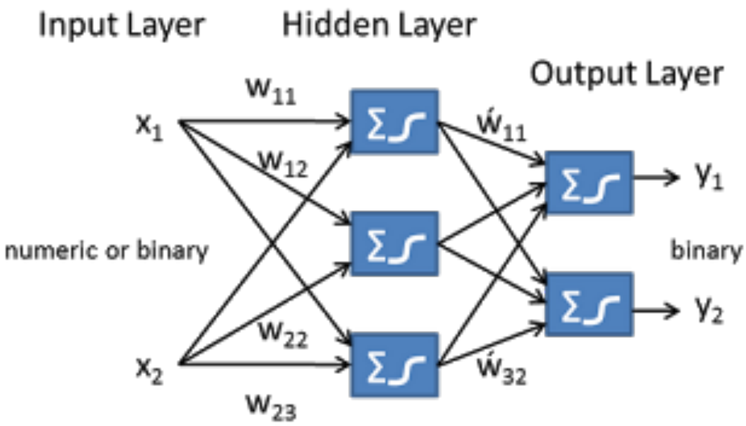


A simple MARS model of the same data



http://en.wikipedia.org/wiki/Multivariate_adaptive_regression_splines

Neural Networks



http://en.wikipedia.org/wiki/Artificial_neural_network

Generalized Dissimilarity Modelling

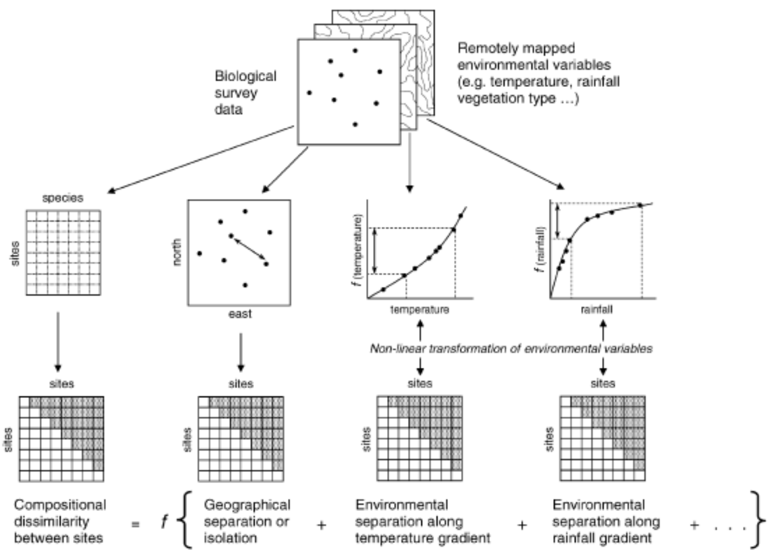


Figure 8. A diagrammatic illustration of the GDM approach to modelling compositional dissimilarity (based on Ferrier 2002).

Ferrier et al. 2002

UNIVERSITÉ DE GENÈVE

Generalized Dissimilarity Modelling

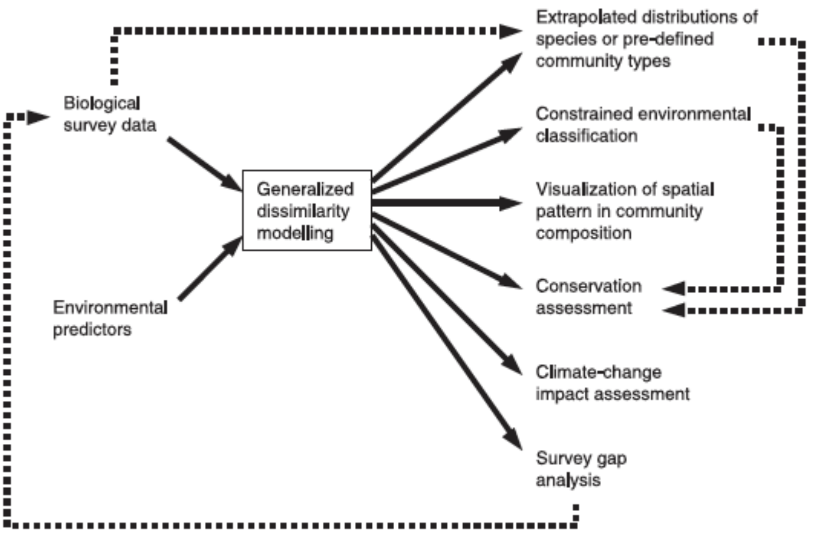
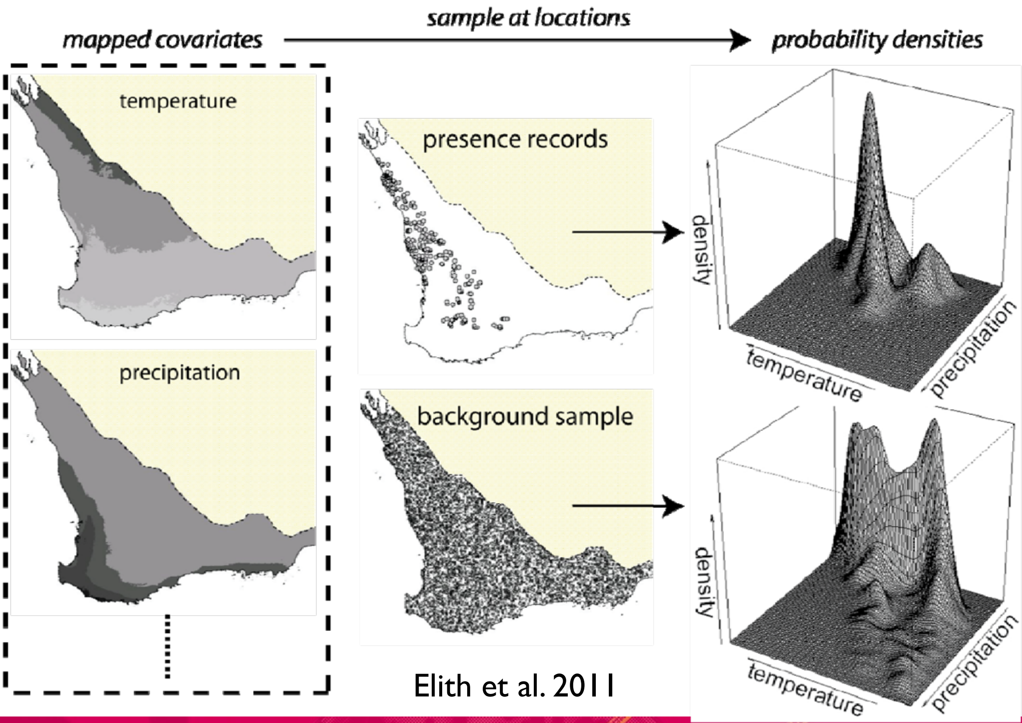


Figure 5 Applications of generalized dissimilarity modelling.

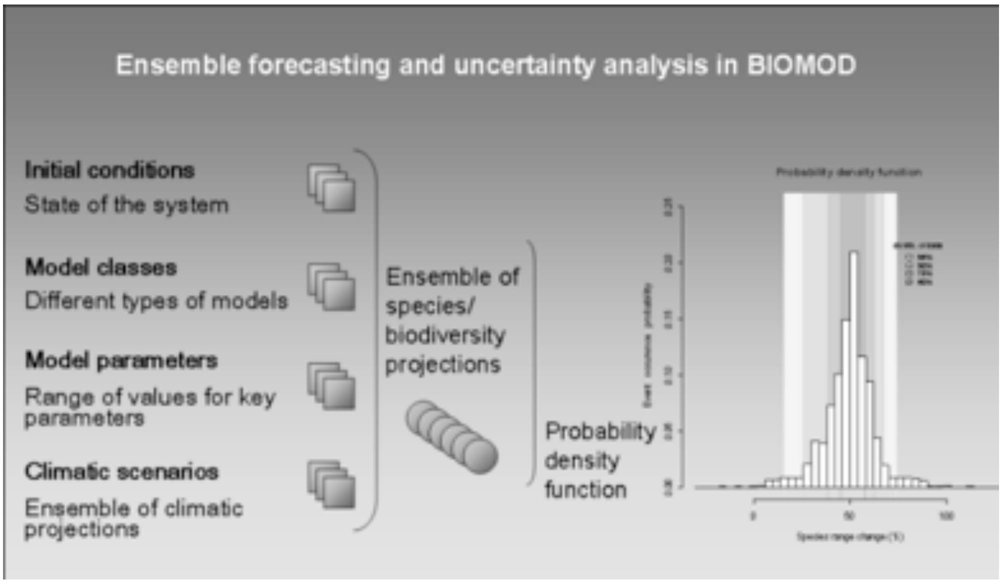
Ferrier et al. 2007

Maximum Enthropy



Elith et al. 2011

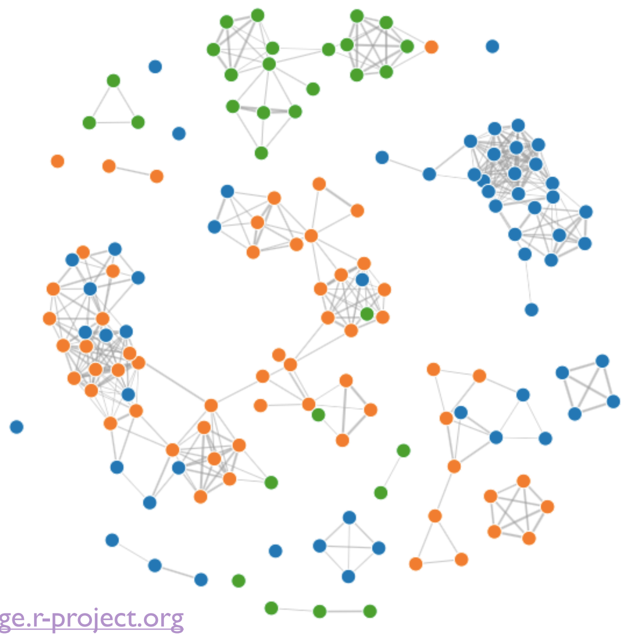
Ensemble Forecasting: Biomod



http://cran.r-project.org/web/packages/biomod2/index.html

Ensemble Forecasting: Caret

- Bagging Models
- Bayesian Model
- Boosting Models
- Cost Sensitive Learning Models
- Discriminant Analysis Models
- Ensemble Model
- Feature Extraction Models
- Feature Selection Wrapper Models
- Gaussian Process Models
- Generalized Additive Model
- Generalized Linear Model
- Implicit Feature Selection Models
- Kernel Method
- L1 Regularization Models
- L2 Regularization Models
- Linear Classifier Models
- Linear Regression Models
- Logic Regression Models
- Logistic Regression Models
- Mixture Model
- Model Tree
- Multivariate Adaptive Regression Splines Models
- Neural Network Models
- Oblique Tree Models
- Partial Least Squares Models
- Polynomial Model
- Prototype Models
- Quantile Regression Models
- Radial Basis Function Models
- Random Forest Models
- Regularization Models
- Relevance Vector Machines
- Ridge Regression Models
- Robust Methods
- Robust Model
- ROC Curves Models
- Rule-Based Model
- Self-Organising Maps
- Support Vector Machines
- Tree-Based Model



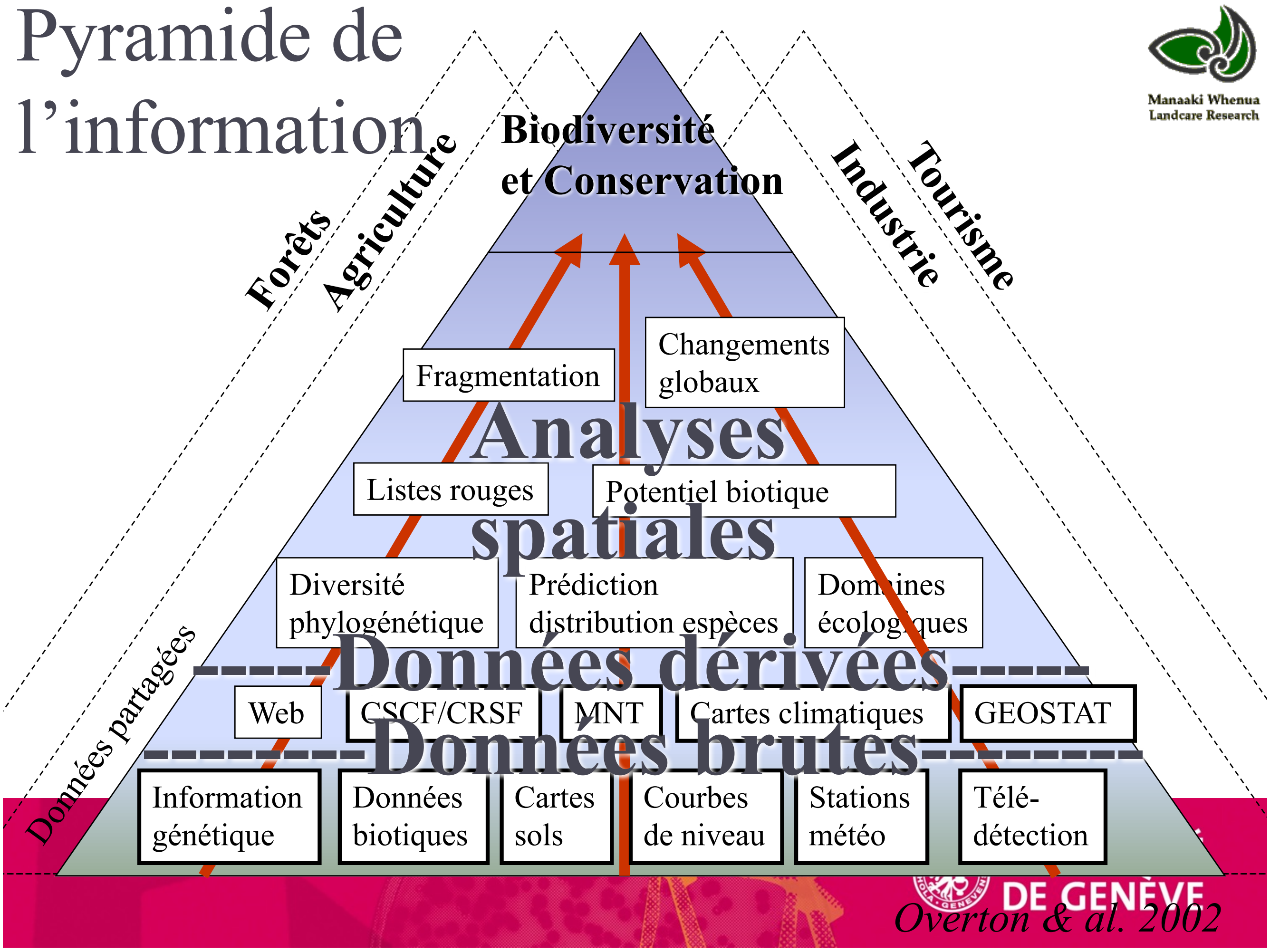
http://caret.r-forge.r-project.org
+ http://www.youtube.com/watch?v=7Jbb2ItbTC4

UNIVERSITÉ DE GENÈVE

Generalized Regression Analyses
and Spatial Predictions
GRASP

UNIVERSITÉ
DE GENÈVE

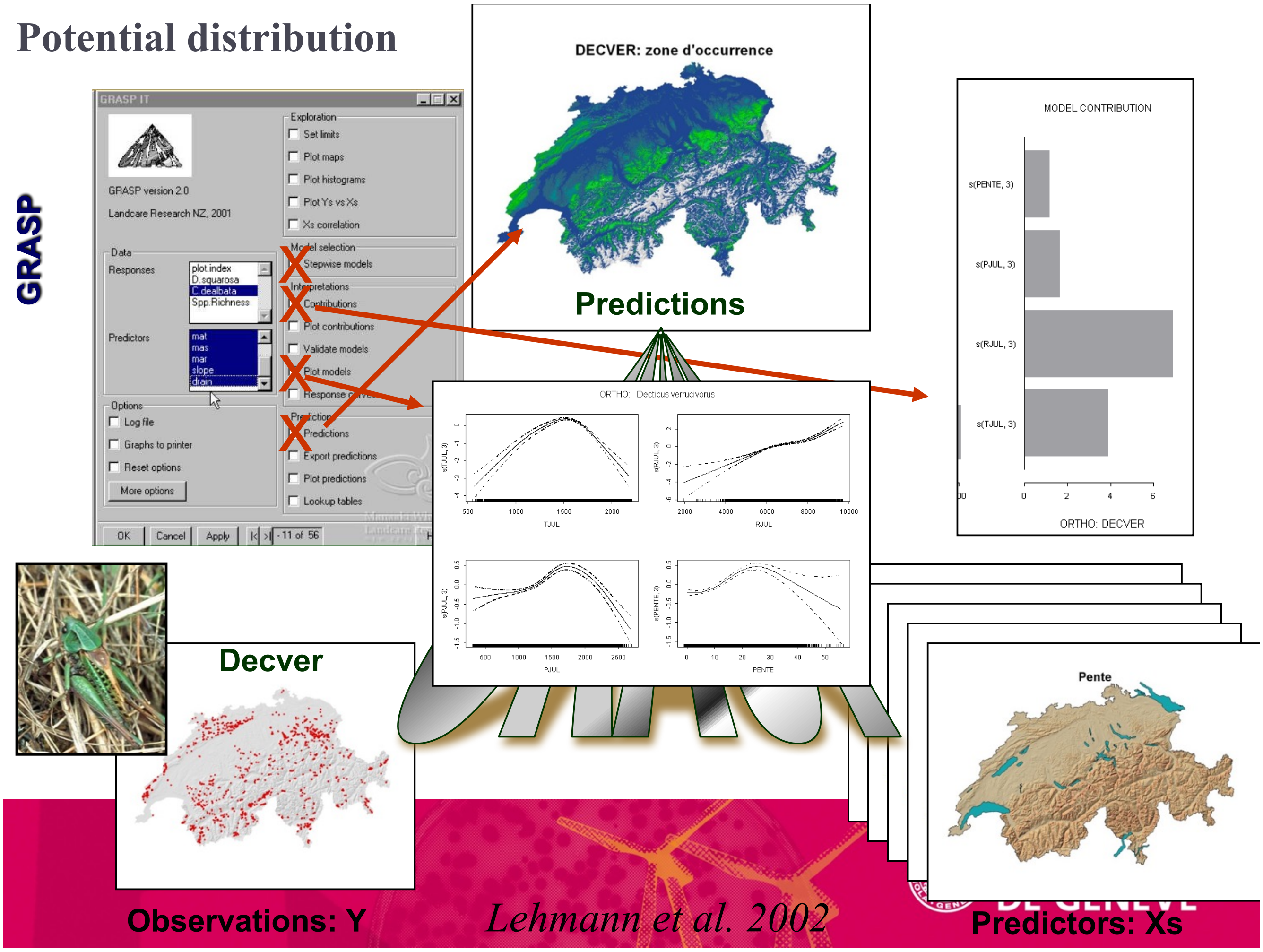
Pyramide de l'information



Manaaki Whenua
Landcare Research

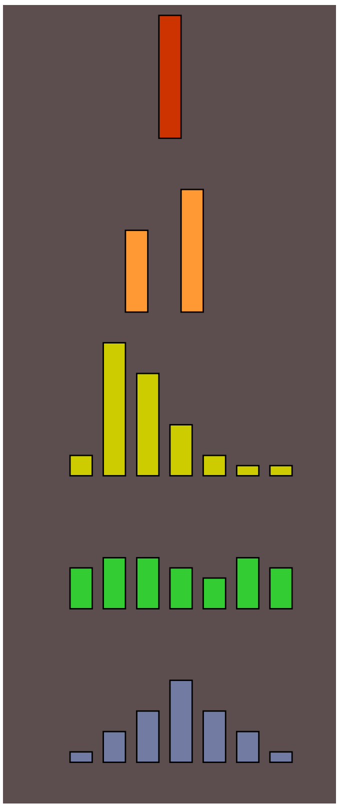
Overton & al. 2002

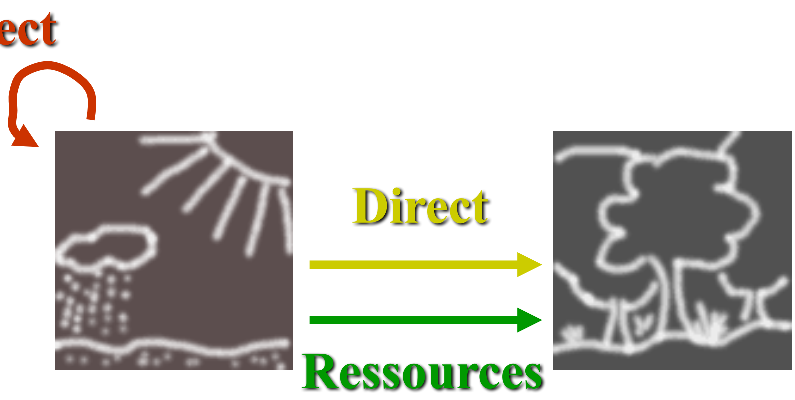
Potential distribution



Données	Méthodes	Résultats
Présences seul.:	PCA Biomapper, faux 0 > binomiale	Probabilité
Présence absence:	binomiale	Probabilité
Abondance:	Poisson	Abondance
Richesse:	Poisson	Richesse
Couverture :	binomiale	Pourcentage
Biomasse :	normale	biomasse

Distribution





	Indirects	Directs	Ressources
Continus	Altitude	Température	Lumière
Discrets	Classe pédologique	pH	Nutriments

Regression & GIS:

	Least Square Regression	Generalized Linear Model	Generalized Additive Model
Y and error distribution:	Normal	Normal, binomial, multinomial, Poisson	Normal, binomial, multinomial, Poisson
Response curves:	Parametric (X, X^2, X^3...)	Parametric (X, X^2, X^3...)	Non parametric, any shape
Combination with GIS:	Substitution of Xs by maps in GIS	Substitution of Xs by maps in GIS	Estimation of Y in statistical package and export to GIS

$$Y_i = \alpha + \sum_{j=1}^{p} \beta_j X_{j\,i} + \varepsilon_i$$

$$g(\mu) = \alpha + \sum_{j=1}^{p} \beta_j x_{ji} + \varepsilon_i$$

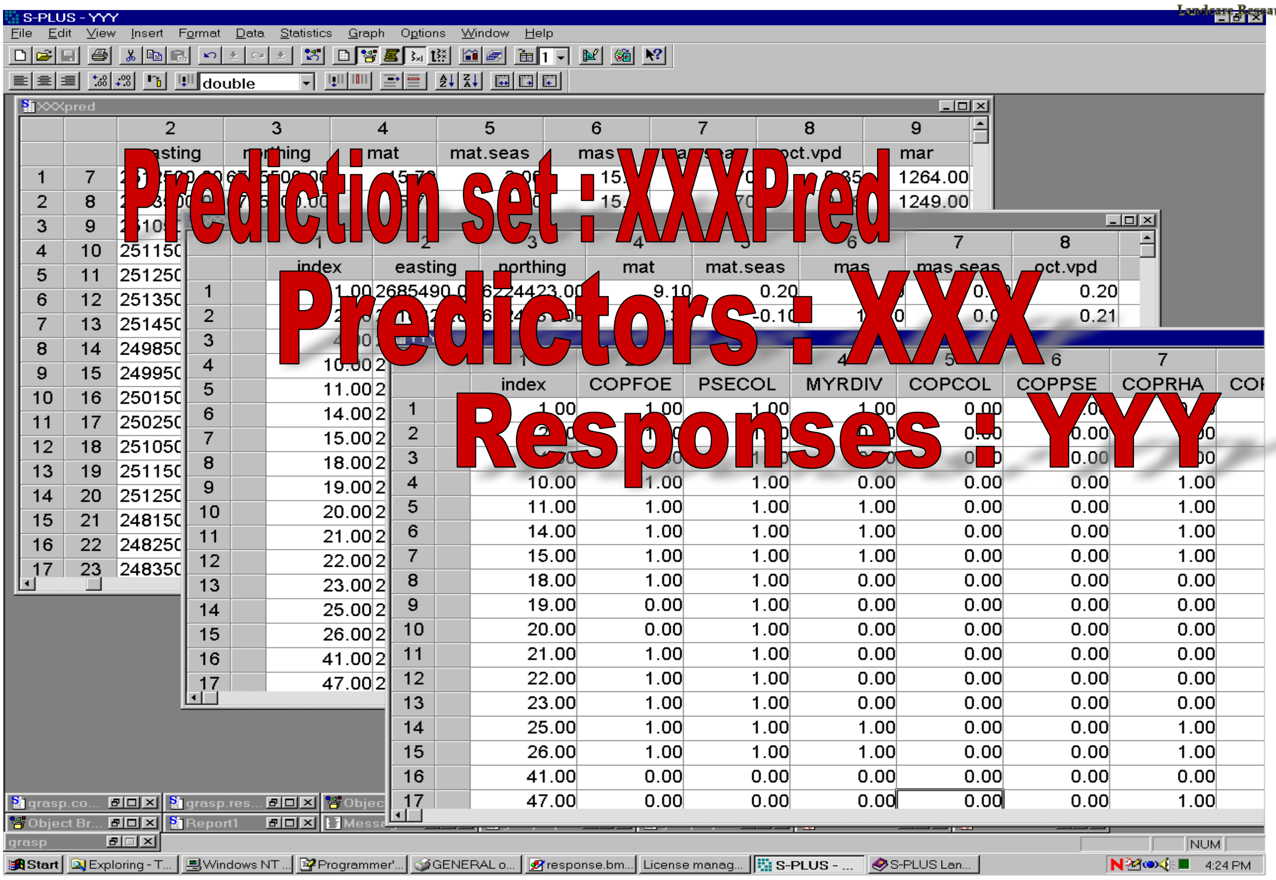
$$g(\mu) = \alpha + \sum_{j=1}^{p} f_j(x_{ji}) + \varepsilon_i$$

Running GRASP from a script in Splus and R

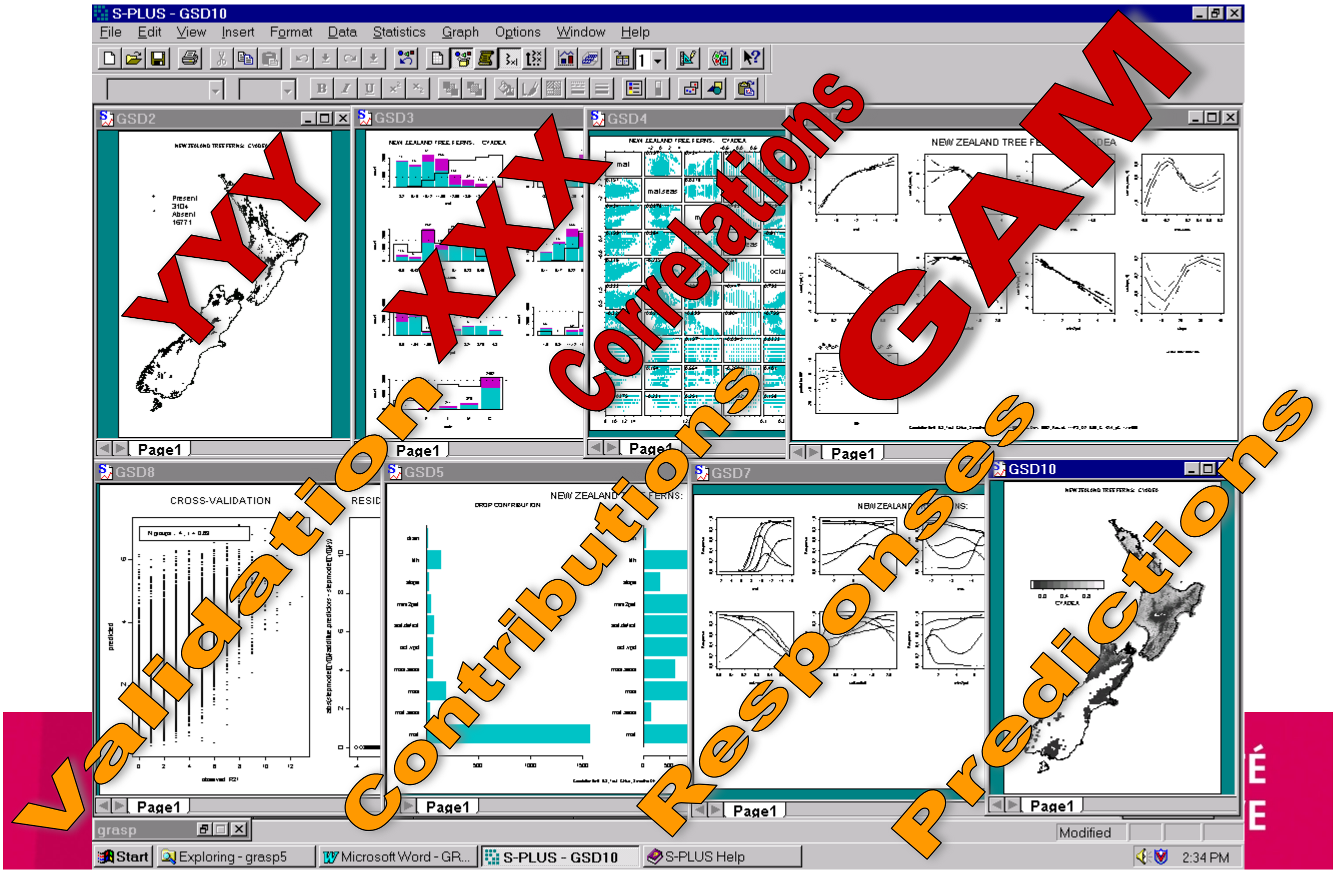
```
grasp(selected.responses = c(2,3)
selected.predictors = c(4,5,6,7,8,9),
gr.fam = "binomial",
# DATA OBSERVATION
plot.maps = T, plot.distry = T, plot.histograms = T, plot.respvspred = T,
plot.correlation = T, corlim = 80, plot.xpred = T, use.correlation = T, reset.cor = F,
# MODEL SELECTION
test = "CROSS", stepwise.models = T, contributions = T, plot.contributions = T,
plot.models = T, validate.models = T,
# PREDICTIONS
predictions = T, export.predictions = T, plot.predictions = T)
}
```

UNIVERSITÉ DE GENÈVE

Manaaki Whenua Landcare Research

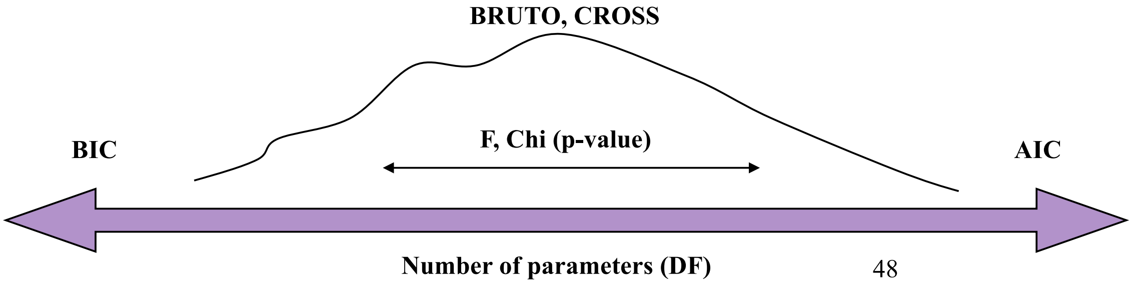


GRASP Outputs

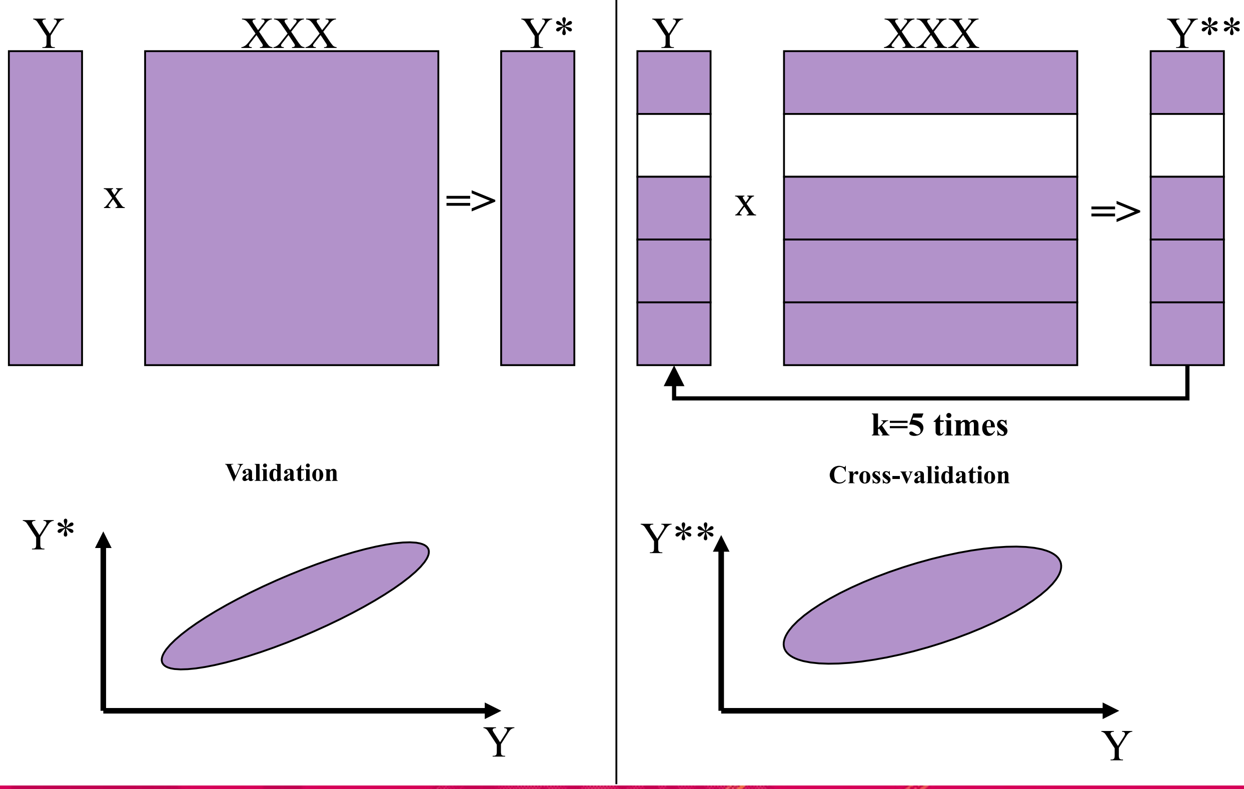


Stepwise selection methods

- Parcymony: avoid under- or over-predicting
- Stepwise: Remove or add one predictor and test for differences
- Directions: backward, forward, both
- Criteria and Tests:
  - o Chi: compare models with a Chi test and a p-value
  - o F: F test for quasi models
  - o AIC: - 2logL(M) + 2***K**, with K parameters, Likelihood, Model
  - o BIC: - 2logL(M) + log (**N**) **K**, with N sample size
  - o BRUTO: Backfitting algorithm selecting parameters and DF
  - o CROSS: selection based on cross-validation at each step



Validation vs. Cross Validation

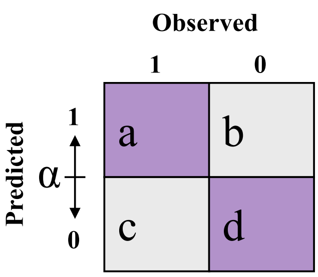


- Spearman correlation r for Gaussian and Poisson distributions
- AUC ROC statistics for binomial distribution

UNIVERSITÉ DE GENÈVE

Binomial model validation

Confusion matrix:

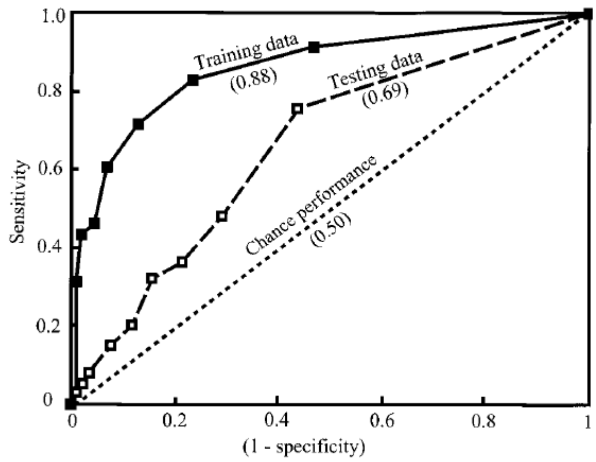


Prevalence:	(a + c)/N
Overall diagnostic power:	(b + d)/N
Correct classification rate:	(a + d)/N
Sensitivity:	a/(a + c)
Specificity:	d/(b + d)
False positive rate:	b/(b + d)
False negative rate:	c/(a + c)
Positive predictive power (PPP):	a/(a + b)
Negative predictive power (NPP):	d/(c + d)
Misclassification rate:	(b + c)/N
Odds-ratio:	(ad)/(cb)

Kappa: [(a+d)-(((a+c)(a+b)+(b+d)(c+d))/N)]/[N-(((a+c)(a+b)+(b+d)(c+d))/N)]

(with poor K , 0.4; good 0.4 , K , 0.75 and excellent K.0.75)

AUC ROC:

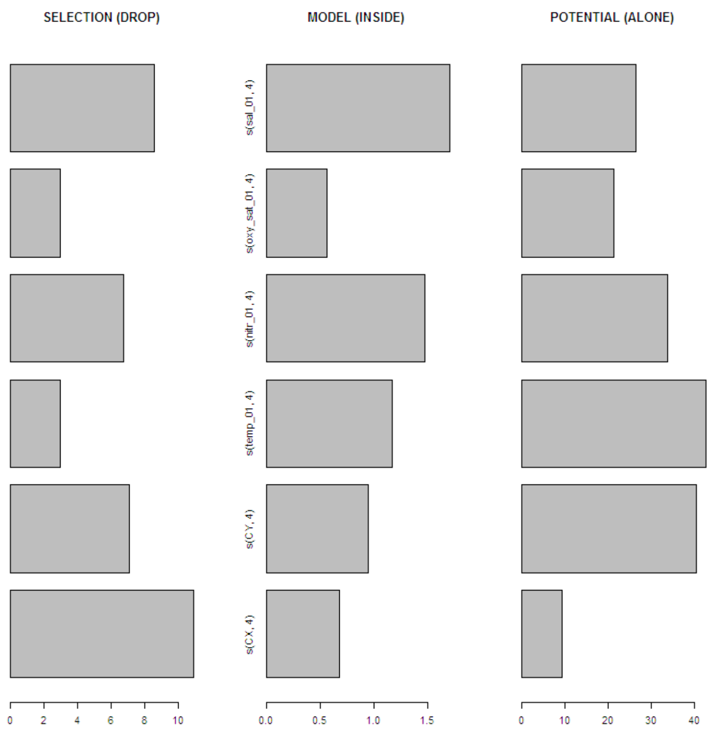


Best Kappa calculates K at different treshold values of α
and selects the best treshold value
as a compromise between PPP and NPP

Fielding 1997

UNIVERSITÉ DE GENÈVE

Model Contributions



deviance LP scale deviance

DROP: drop of explained deviance when removing a variable from a model

MODEL: Sensitivity of the model to changes of values along the range of each predictor

ALONE: Explained deviance by a single variable model

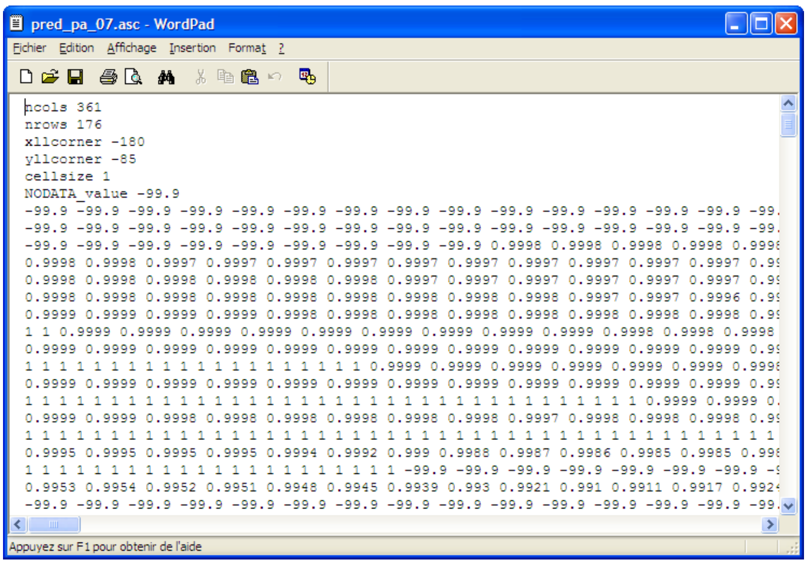
Lehmann et al. 2002

Exporting GAM predictions to a GIS

• If n obervation < 200000, then make prediction in S or Splus with predict.gam function, and export as ESRI ascii file

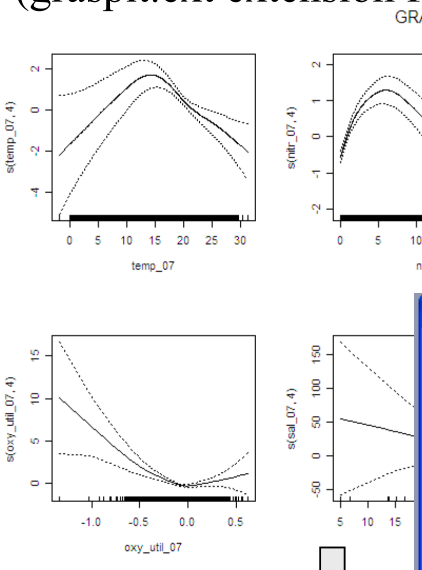
```
> gr.predmat[1:100,5]
  [1] 0.1179 0.0940 0.1713 0.0715
  [5] 0.1021 0.1174 0.1553 0.0590
  [9] 0.0983 0.5032 0.1520 0.7362
 [13] 0.0900 0.0982 0.3434 0.1341
 [17] 0.0695 0.2760 0.0871 0.7726
 [21] 0.0911 0.0733 0.0420 0.0866
 [25] 0.1216 0.8594 0.0408 0.0776
 [29] 0.0938 0.1182 0.0522 0.0545
 [33] 0.5306 0.0472 0.1736 0.2210
 [37] 0.2474 0.9998 0.3117 0.4586
 [41] 0.9999 0.9514 0.6029 0.4998
 [45] 1.0000 0.3477 0.9713 0.5479
 [49] 0.6805 0.2154 0.2205 0.2192
 [53] 0.2441 0.1619 0.2106 0.2220
 [57] 0.2328 0.2255 0.2310 0.2547
 [61] 0.2142 0.2115 0.3600 0.7507
 [65] 0.2298 0.1884 0.2020 0.2079
 [69] 0.2112 0.2228 0.8458 0.2053
 [73] 0.2336 0.2240 0.2553 0.2177
 [77] 0.2208 1.0000 0.2133 0.2395
 [81] 0.2113 0.2772 0.0940 0.2201
 [85] 0.2060 0.2066 0.2322 0.2063
 [89] 0.2389 0.2145 0.2669 0.2034
 [93] 0.2316 0.3172 0.4533 0.1986
 [97] 0.2143 0.2151 0.2231 0.2155
```

→



Exporting GAM predictions to a GIS

- Else, build a Lookup Table (LUT) to describe each repsonse curve, export it as a text file, and read it with a function in GIS (graspit.ext extension for Arcview 3.x)



pa_07_lut.txt

	A	B	C	D	E	F	G
1	row.names	temp_07	nitr_07	oxy_dis_07	oxy_util_07	sal_07	explantions
2	1	-2.279347	-0.560688	3.3185	24.3639	55.9519	linear additive contribution
3	2	-2.228792	-0.427655	3.21358	24.09496	55.5997	"
4	3	-2.178238	-0.296545	3.10867	23.82601	55.2475	"
5	4	-2.127684	-0.169255	3.00375	23.55707	54.8953	"
196	195	-2.290814	-0.399622	7.97532	5.56975	-2.9124	"
197	196	-2.332806	-0.405924	8.05455	5.63827	-2.967	"
198	197	-2.374797	-0.412226	8.13378	5.70678	-3.0216	"
199	198	-2.416789	-0.418528	8.21301	5.7753	-3.0762	"
200	199	-2.45878	-0.42483	8.29225	5.84381	-3.1308	"
201	200	-2.500772	-0.431131	8.37148	5.91232	-3.1854	"
202	201	-1.77	0	4	-1.33	5.01	min of X in XXX (calibration)
203	202	31.32	24.92	9.37	0.62	39.28	max of X in XXX (calibration)
204	203	200	200	200	200	200	number of rows
205	204	-0.0346812	-0.0346812	-0.0346812	-0.0346812	-0.0346812	model intercept
206	205	-2.1	0	3.83	-2.7	5	min of X in XXXpred or XXXlut (prediction)
207	206	33.23	37.35	10.3	2.44	40.65	max of X in XXXpred or XXXlut (prediction)
208							

Exercice dans R: GAM et prédictions spatiales

```
# Spatial Predictions
install.packages("sp") # sp for spatial
library(sp)

load("datareg/YYY.RData")
load("datareg/XXX.RData")
load("datareg/XXXpred.RData")
names(YYY)
names(XXX)
names(XXXpred)

par(mfrow=c(1,1))
plot(XXXpred$CX, XXXpred$CY)
TEMP<-data.frame("CX"=XXXpred$CX,"CY"=XXXpred$CY,"TEMP"=XXXpred$TEMP)
coordinates(TEMP) = c("CX", "CY")gridded(TEMP) = TRUETEMP = as(TEMP,
"SpatialGridDataFrame") # to full
gridimage(TEMP)
image(TEMP, col = terrain.colors(250))

install.packages("gam") library(gam)gam4 <-
gam(YYY$RICHNESS~s(TEMP)+s(RAD)+s(PREC)+s(NNESS)+s(SLOPE),
family=poisson,select=TRUE, data=XXX)summary(gam4)
par(mfrow=c(3,2))
plot(gam4, scale=0)
```

Exercice dans R: GAM et prédictions spatiales

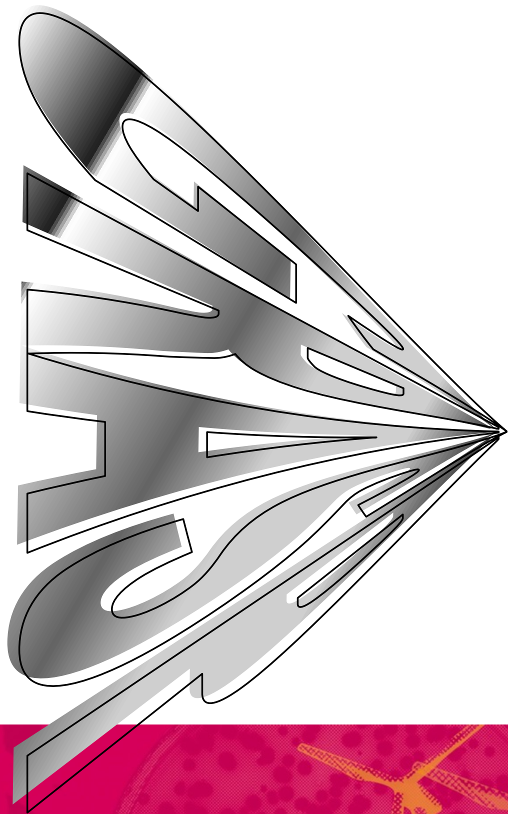
```
gam4pred<-predict.gam(gam4,XXXpred)par(mfrow=c(1,1)

PRED.SPR<-data.frame("CX"=XXXpred$CX,"CY"=XXXpred$CY,"PRED.SPR"=gam4pred)

coordinates(PRED.SPR) = c("CX", "CY")

gridded(PRED.SPR) = TRUEPRED.SPR = as(PRED.SPR, "SpatialGridDataFrame") # to
full grid

image(PRED.SPR, col = terrain.colors(250))

plot(PRED.SPR["PRED.SPR"], zlim = c(0,max(gam4pred)), col=terrain.colors(20),
axes = TRUE)title("Species richness")

points(XXX$CX, XXX$CY, col = 'red', cex=YYY$RICHNESS/10)

box()
```

UNIVERSITÉ DE GENÈVE

Fougères NZ

Manaaki Whenua
Landcare Research

- **Température**
- **Radiation**
- **Humidité air**
- **Pluie / Evaporation**
- **Humidité du sol**
- **Pente**
- **Lithologie**
- **Ecoulement**
- **Distribution espèces**

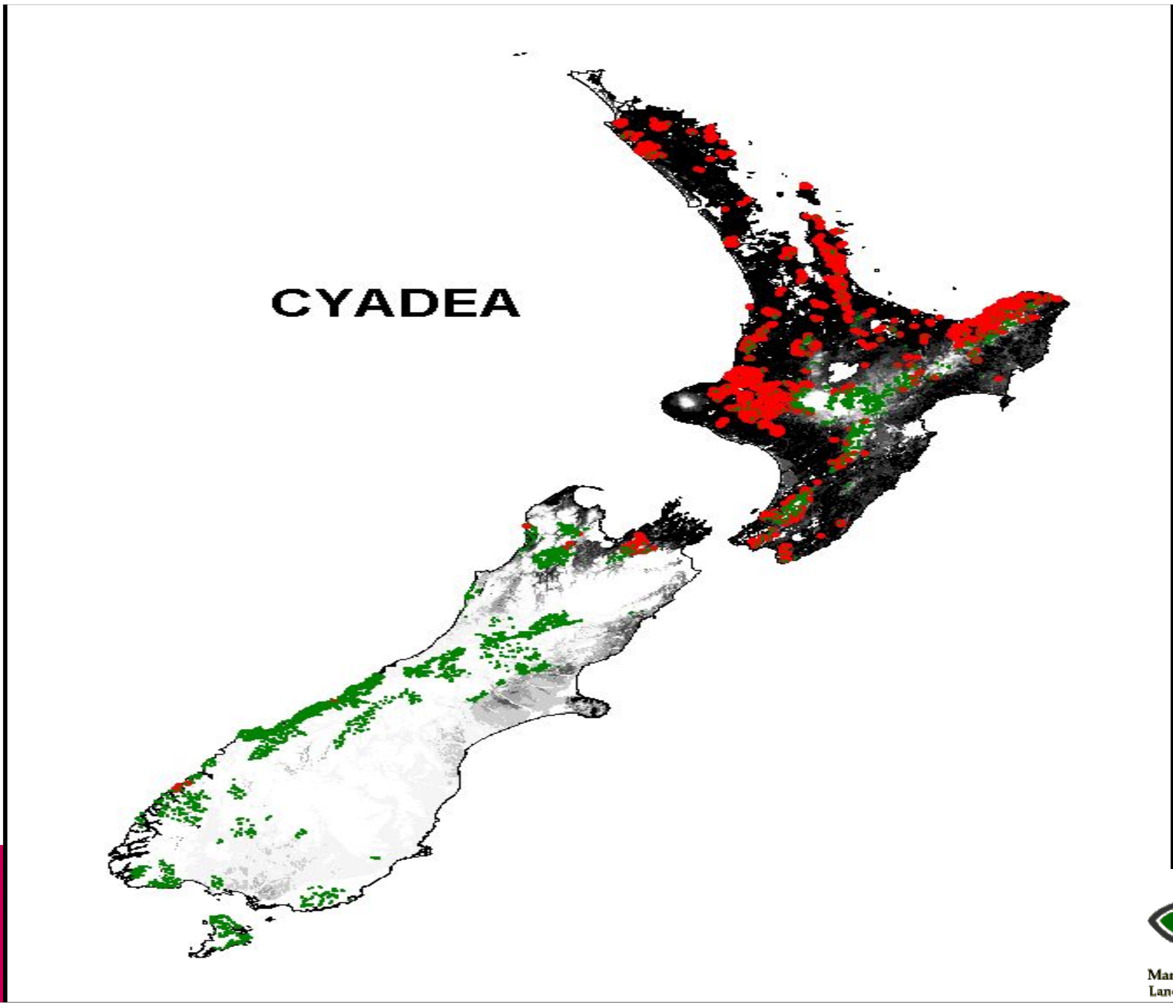


Prédire la distribution des espèces

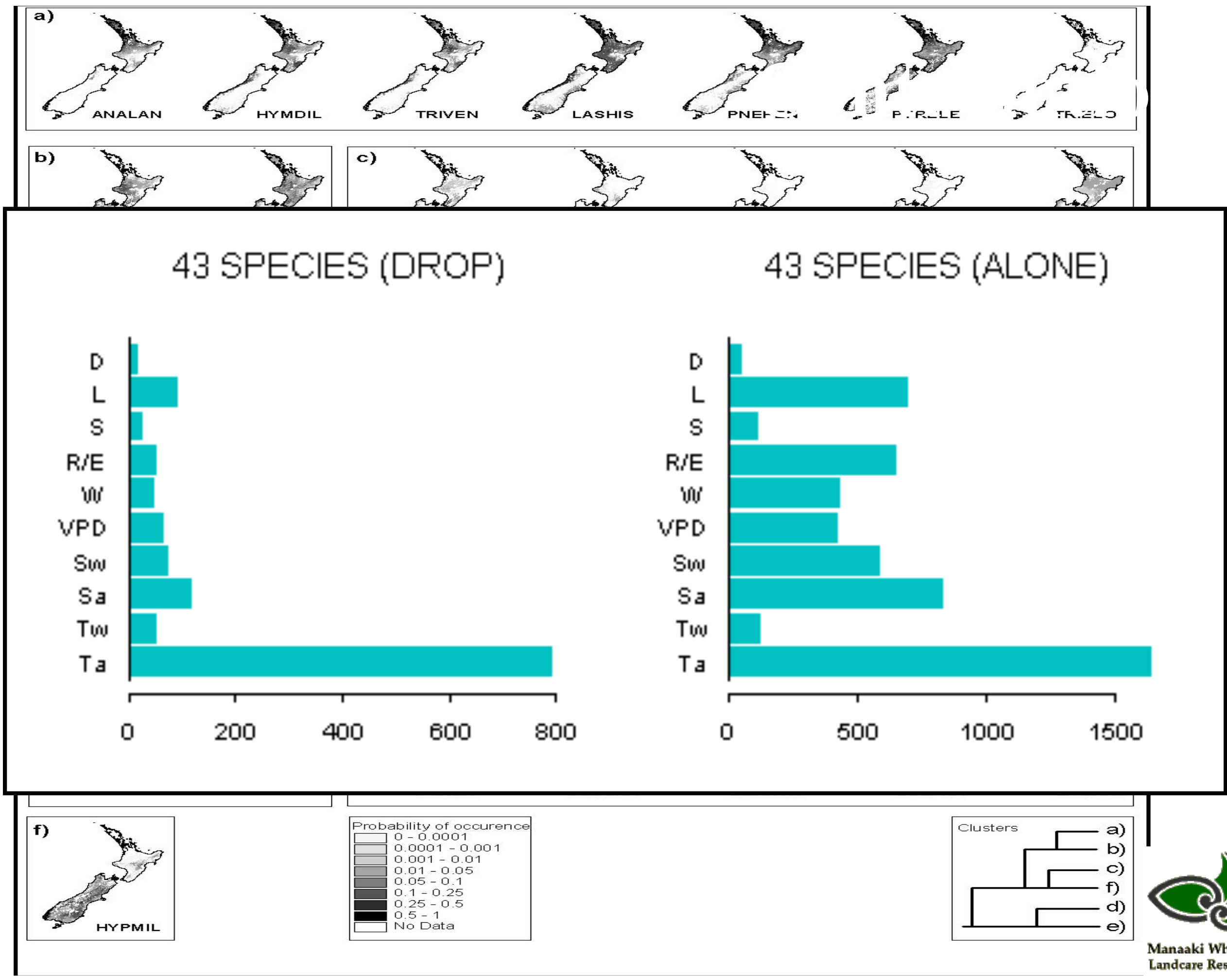
Identifier les « hotspot » de biodiversité

Lehmann et al., 2002.

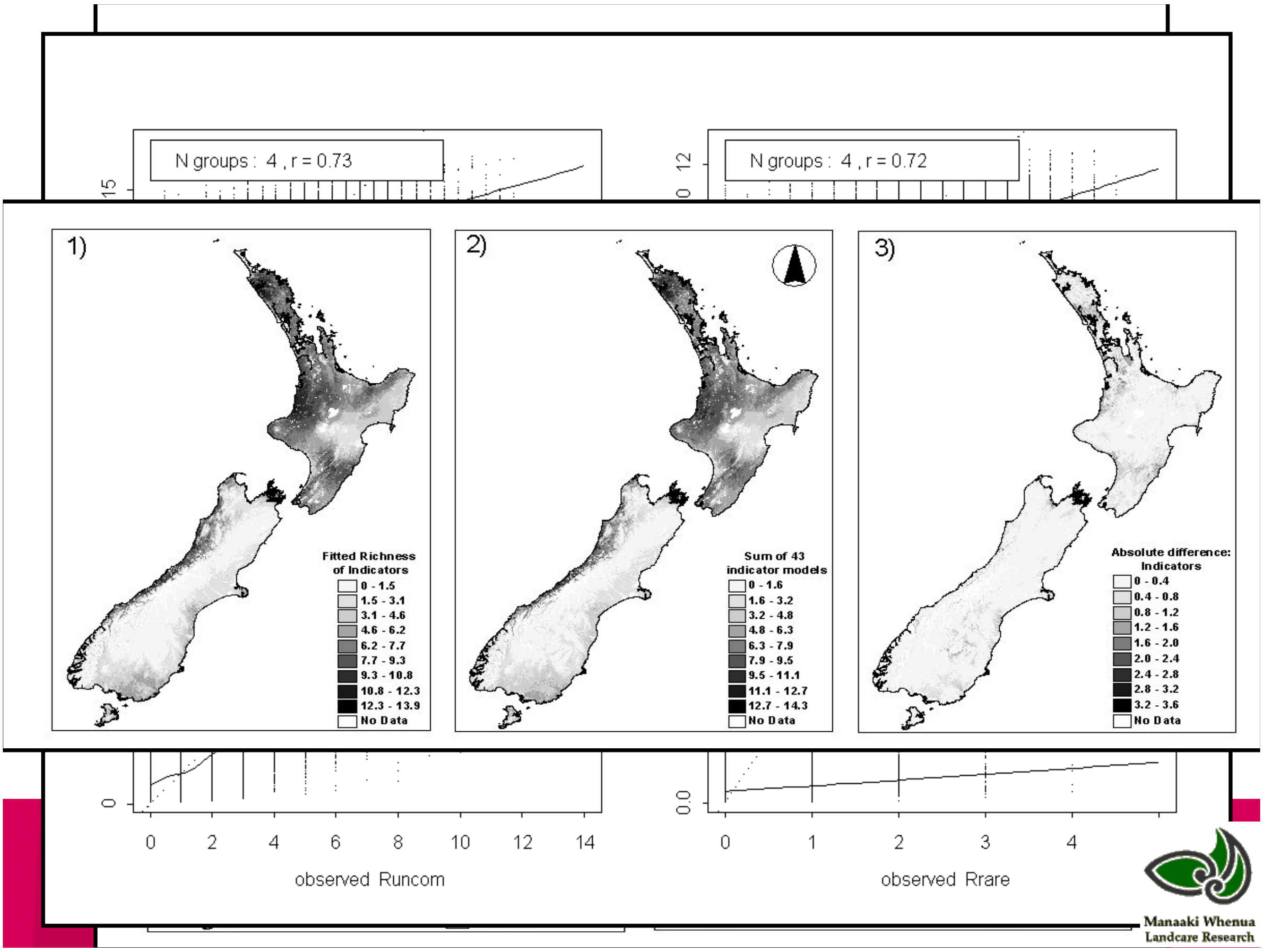
UNIVERSITÉ DE GENÈVE



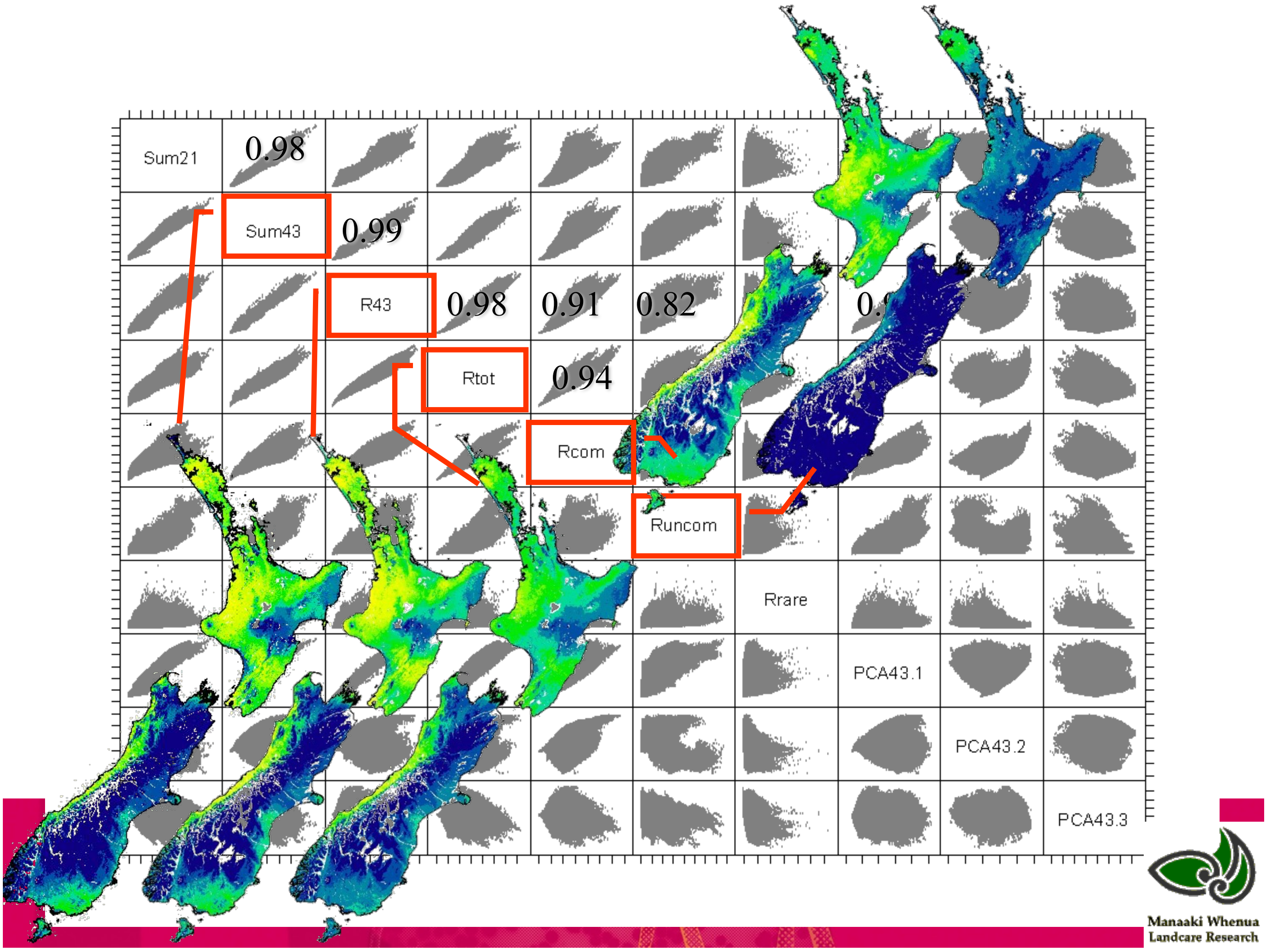
Manaaki Whenua
Landcare Research



Manaaki Whenua
Landcare Research



N groups : 4 , r = 0.73
N groups : 4 , r = 0.72
1)
Fitted Richness of Indicators
0 - 1.5
1.5 - 3.1
3.1 - 4.6
4.6 - 6.2
6.2 - 7.7
7.7 - 9.3
9.3 - 10.8
10.8 - 12.3
12.3 - 13.9
No Data
2)
Sum of 43 indicator models
0 - 1.6
1.6 - 3.2
3.2 - 4.8
4.8 - 6.3
6.3 - 7.9
7.9 - 9.5
9.5 - 11.1
11.1 - 12.7
12.7 - 14.3
No Data
3)
Absolute difference: Indicators
0 - 0.4
0.4 - 0.8
0.8 - 1.2
1.2 - 1.6
1.6 - 2.0
2.0 - 2.4
2.4 - 2.8
2.8 - 3.2
3.2 - 3.6
No Data
observed Runcom
observed Rrare
Manaaki Whenua
Landcare Research



Sum21
0.98
Sum43
0.99
R43
0.98
0.91
0.82
Rtot
0.94
Rcom
Runcom
Rrare
PCA43.1
PCA43.2
PCA43.3
Manaaki Whenua
Landcare Research

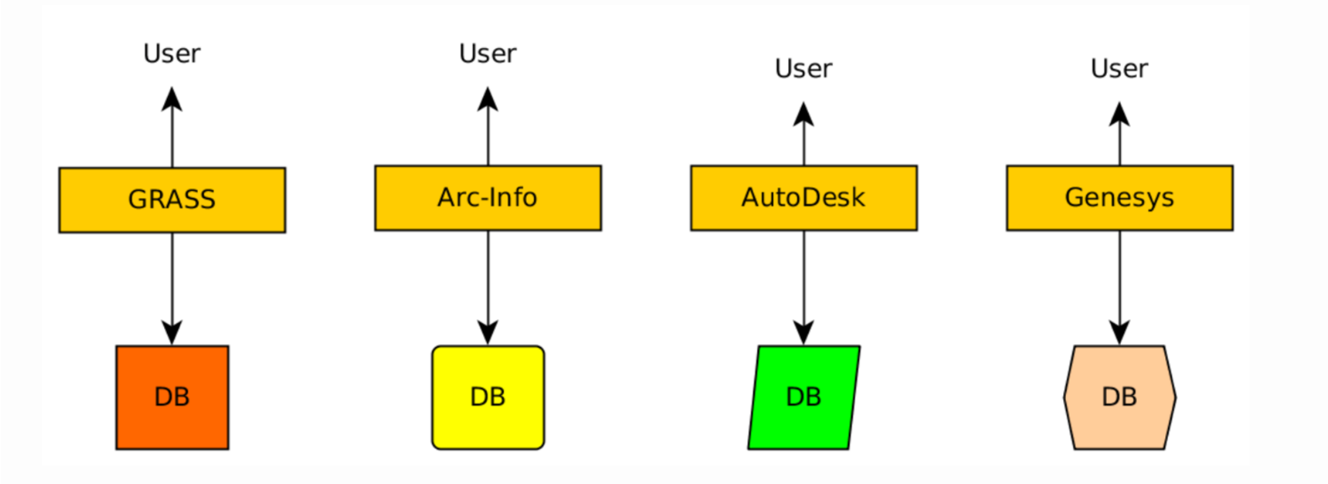
GRASP:

- ↳ permet de modéliser des données provenants de différents types de distribution sans *a priori* sur la forme des courbes de réponses
- ↳ permet de standardiser et d'automatiser partiellement la modélisation
- ↳ apporte: vitesse, flexibilité, généralisation, uniformité
- ↳ (l'outil) utilise les GAMs dans R pour faciliter les prédictions (spatiales)
- ↳ (la méthode) vise à promouvoir les prédictions spatiales dans la gestion de l'environnement

Working with spatial data in R

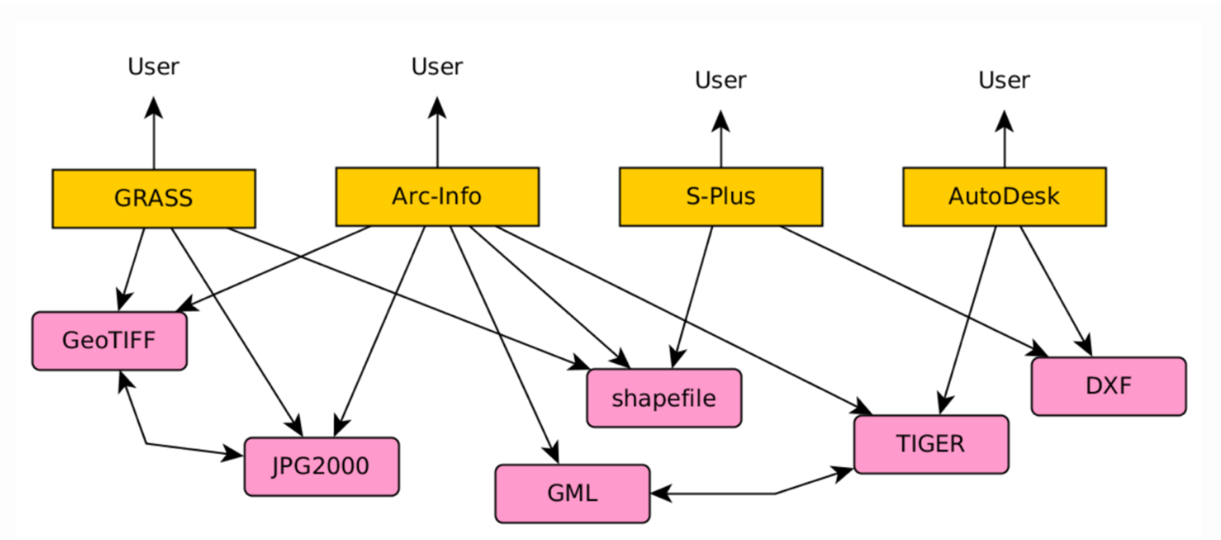
UNIVERSITÉ DE GENÈVE

80's



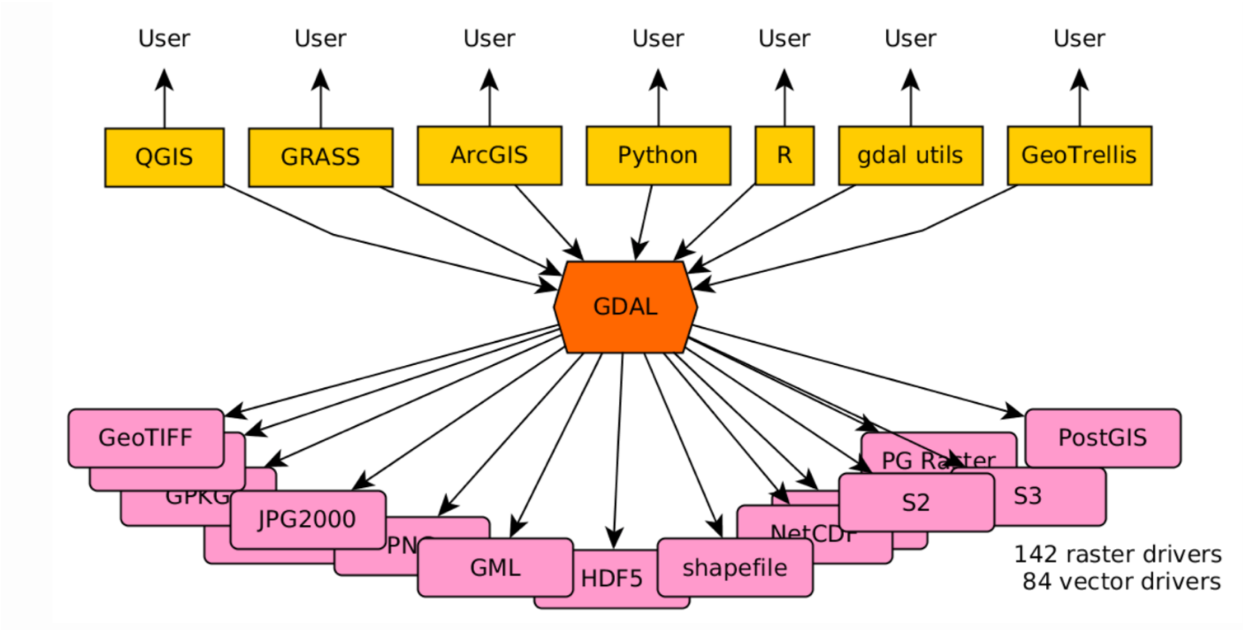
http://r-spatial.org/2016/11/29/openeo.html

90's



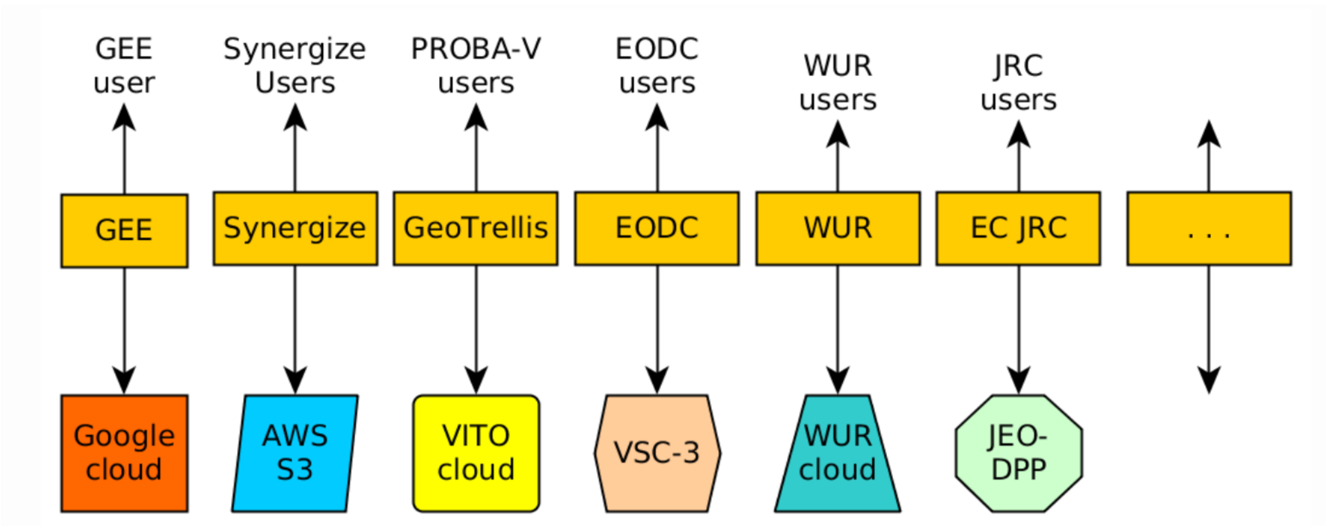
http://r-spatial.org/2016/11/29/openeo.html

00’s



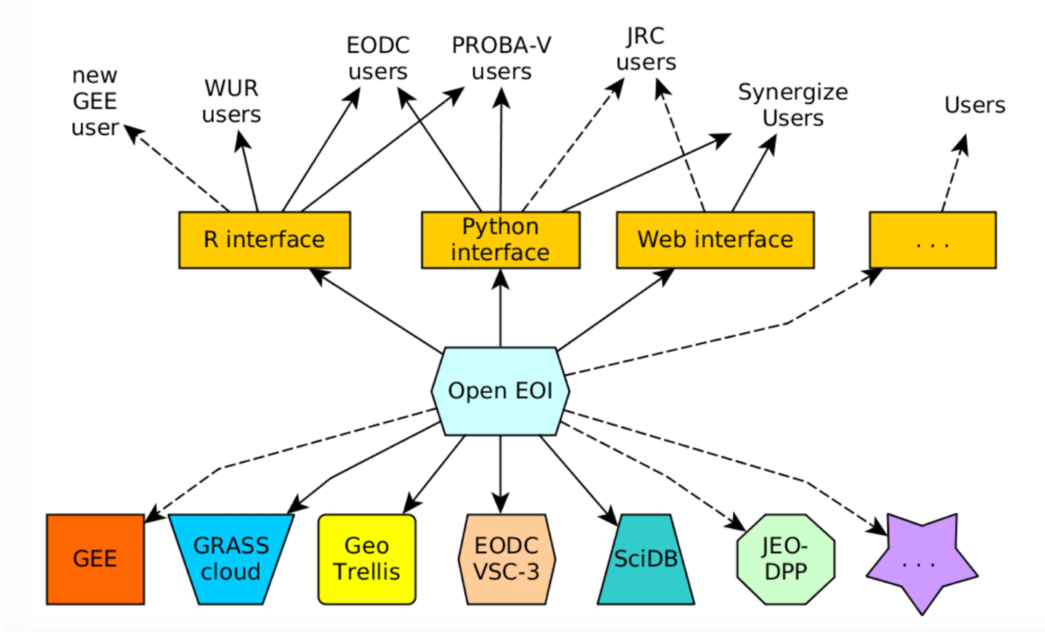
http://r-spatial.org/2016/11/29/openeo.html

10’s



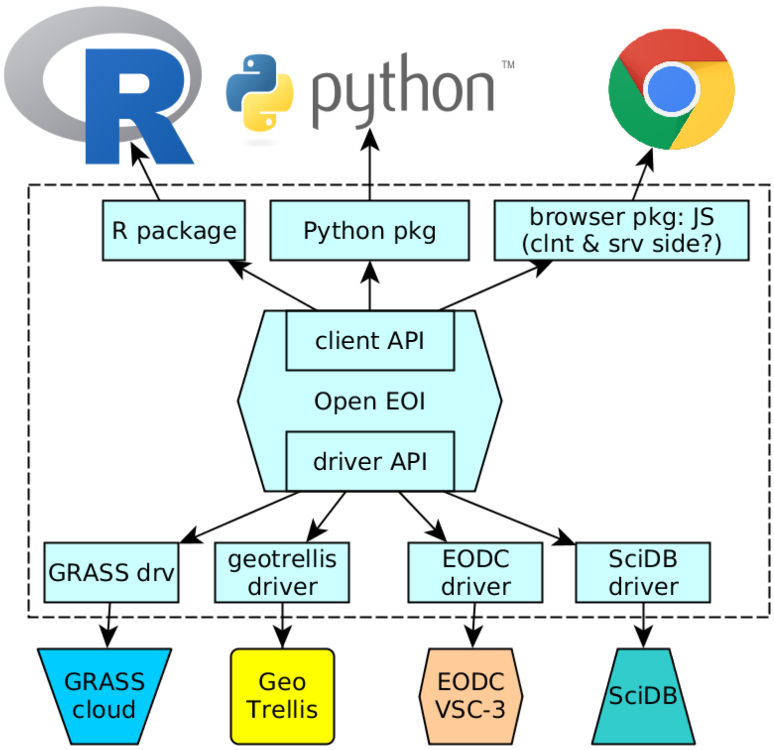
http://r-spatial.org/2016/11/29/openeo.html

20's



http://r-spatial.org/2016/11/29/openeo.html

20's



http://r-spatial.org/2016/11/29/openeo.html

Existing spatial libraries

- sp: https://cran.r-project.org/web/packages/sp/index.html
  Classes and methods for spatial data
- rgdal: https://cran.r-project.org/web/packages/rgdal/index.html
  Provides bindings to Frank Warmerdam's Geospatial Data Abstraction Library (GDAL)
- raster: https://cran.r-project.org/web/packages/raster/index.html
- sf: https://cran.r-project.org/web/packages/sf/index.html :
  Support for simple features, a standardized way to encode spatial data, with bindings to GDAL, GEOS and Proj.4.
- ggmap: https://cran.r-project.org/web/packages/ggmap
  A collection of functions to visualize spatial data and models on top of static maps from various online sources (e.g Google Maps and Stamen Maps).
- maptools: https://cran.r-project.org/web/packages/maptools
  Set of tools for manipulating and reading geographic data, in particular 'ESRI Shapefiles'
- ggplot2: https://cran.r-project.org/web/packages/ggplot2/index.html
  Create Elegant Data Visualisations Using the Grammar of Graphics
- rgeos: https://cran.r-project.org/web/packages/rgeos
  Interface to Geometry Engine - Open Source (GEOS) using the C API for topology operations on geometries
- spatial: https://cran.r-project.org/web/packages/spatial/spatial.pdf:
  Functions for Kriging and Point Pattern Analysis
- spatstat: https://cran.r-project.org/web/packages/spatstat
  Comprehensive open-source toolbox for analysing Spatial Point Patterns

CRAN Task View: Analysis of Spatial Data

CRAN Task View: Analysis of Spatial Data

Maintainer: Roger Bivand
Contact: Roger.Bivand at nhh.no
Version: 2016-11-27
URL: https://CRAN.R-project.org/view=Spatial

Base R includes many functions that can be used for reading, visualising, and analysing spatial data. The focus in this view is on "geographical" spatial data, where observations can be identified with geographical locations, and where additional information about these locations may be retrieved if the location is recorded with care. Base R functions are complemented by contributed packages, some of which are on CRAN, and others are still in development. One active location is R-Forge , which lists "Spatial Data and Statistics" projects in its project tree . Information on R-spatial packages, especially sp is posted on the R-Forge rspatial project website , including a visualisation gallery. Active development of sp is continuing on Github .

The contributed packages address two broad areas: moving spatial data into and out of R, and analysing spatial data in R.

The R-SIG-Geo mailing-list is a good place to begin for obtaining help and discussing questions about both accessing data, and analysing it. The mailing list is a good place to search for information about relevant courses. Further information about courses may be found under the "Events" tab of this blog .

There are a number of contributed tutorials and introductions; a recent one is Introduction to visualising spatial data in R by Robin Lovelace and James Cheshire.

https://cran.r-project.org/web/views/Spatial.html

GDAL

GDAL

Main Page | Related Pages | Classes | Files | Download | Issue Tracker

GDAL - Geospatial Data Abstraction Library

Select language: [English][Russian][Portuguese][French/Francais]

GDAL is a translator library for raster and vector geospatial data formats that is released under an X/MIT style Open Source license by the Open Source Geospatial Foundation. As a library, it presents a **single raster abstract data model** and **single vector abstract data model** to the calling application for all supported formats. It also comes with a variety of useful command line utilities for data translation and processing. The NEWS page describes the October 2016 GDAL/OGR 2.1.2 release.

OSGeo Project

Traditionally GDAL used to design the raster part of the library, and OGR the vector part for Simple Features. Starting with GDAL 2.0, both sides have been more tightly integrated. You can still refer to the documentation of GDAL 1.X if needed.

Master: http://www.gdal.org
Download: http at download.osgeo.org

User Oriented Documentation

- Wiki - Various user and developer contributed documentation and hints
- Downloads - Ready to use binaries (executables)
- Supported raster formats (142 drivers) : GeoTIFF, Erdas Imagine, ECW, MrSID, JPEG2000, DTED, NITF, GeoPackage, ...
- Supported vector formats (84 drivers): ESRI Shapefile, ESRI ArcSDE, ESRI FileGDB, MapInfo (tab and mid/mif), GML, KML, PostGIS, Oracle Spatial, GeoPackage, ...
- Raster utility programs : gdalinfo, gdal_translate, gdaladdo, gdalwarp, ...
- Vector utility programs : ogrinfo, ogr2ogr, ogrtindex, ...
- Geographic Network utility programs : gnmmanage, gnmanalyse, ...
- GDAL FAQ
- Raster and Vector data models and architecture
- GDAL/OGR Governance and Community Participation
- GDAL Service Provider Listings (not vetted)
- Acknowledgements and Credits
- Software Using GDAL

http://www.gdal.org

UNIVERSITÉ DE GENÈVE

r-spatial

Events Projects Blogroll About

Plotting gridded data with sp

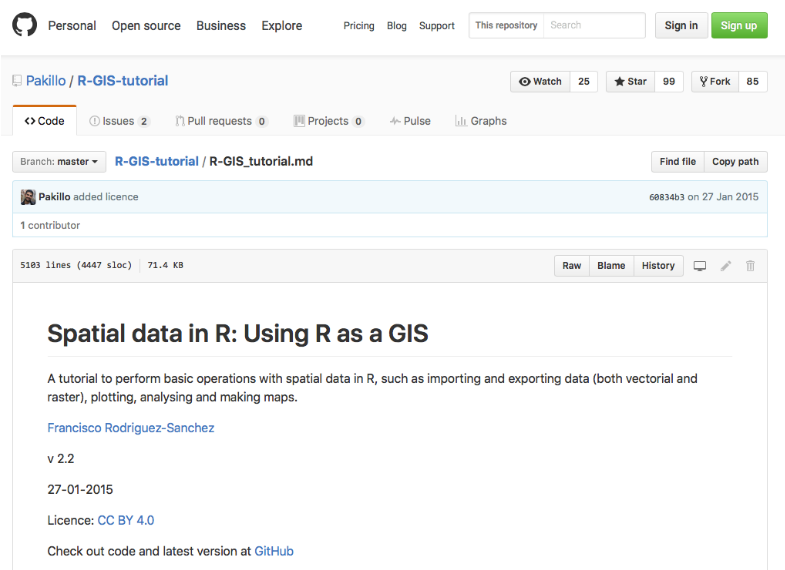
Mar 8, 2016 • Edzer Pebesma

- Introduction
- Incrementally adding elements
- Axes, scale placement and size
- Specifying the scale color breaks, tics and labels
- Adding grid lines
- Plotting scale only, or image only
- Plotting gridded categorical (factor) variables
- Factors: changing the density and width of scales, label width
- Relative size of the categorical scale
- Arranging multiple plots
- Challenges

http://www.rspatial.org

UNIVERSITÉ DE GENÈVE

Tutoriel



1. INTRODUCTION

2. GENERIC MAPPING
•Retrieving base maps from Google with `gmap` function in package `dismo`
•`RgoogleMaps`: Map your data onto Google Map tiles
•`googleVis`: visualise data in a web browser using Google Visualisation API
•`RWorldMap`: mapping global data

3. SPATIAL VECTOR DATA (points, lines, polygons)
•Example dataset: retrieve point occurrence data from GBIF
•Making data 'spatial'
•Define spatial projection
•Quickly plotting point data on a map
•Subsetting and mapping again
•Mapping vectorial data (points, polygons, polylines)
•Drawing polygons and polylines (e.g. for digitising)
•Converting between formats, reading in, and saving spatial vector data
•Changing projection of spatial vector data

4. USING RASTER (GRID) DATA
•Downloading raster climate data from internet
•Loading a raster layer
•Creating a raster stack
•Raster bricks
•Crop rasters
•Define spatial projection of the rasters
•Changing projection
•Plotting raster data
•Spatial autocorrelation
•Extract values from raster
•Rasterize vector data (points, lines or polygons)
•Changing raster resolution
•Spline interpolation
•Setting all rasters to the same extent, projection and resolution all in one
•Elevations, slope, aspect, etc
•Saving and exporting raster data

5. SPATIAL STATISTICS
•Point pattern analysis
•Geostatistics

https://github.com/Pakillo/R-GIS-tutorial/blob/master/R-GIS_tutorial.md

UNIVERSITÉ DE GENÈVE

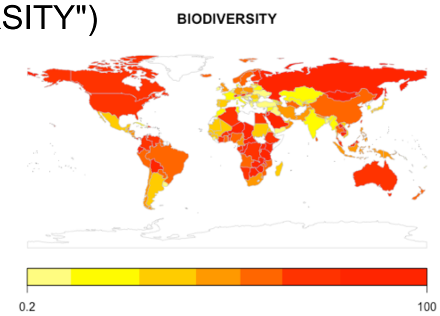
Exercice 1 Generic Mapping

```
# Exercices from Francisco Rodriguez-Sanchez
# Adapted by Anthony Lehmann
# https://github.com/Pakillo/R-GIS-tutorial/blob/master/R-GIS_tutorial.md
setwd(".../STAT 2017/1_intro/Spatial")

install.packages("sp"); library(sp)  # classes for spatial ("raster")
install.packages("raster"); library(raster)  # grids, rasters
install.packages("rgdal"); library(rgdal)

# Retrieving base maps from Google with gmap function i
package dismo
install.packages("dismo")
library(dismo)
install.packages("XML")
library(XML)
mymap <- gmap('Switerland')  # choose whatever country
plot(mymap)
mymap <- gmap('Wallis, Switzerland', type='satellite')
plot(mymap)
```

UNIVERSITÉ DE GENÈVE

Exercice 2 Generic Mapping

```
# RWorldMap: mapping global data

install.packages("rworldmap"); library(rworldmap)

newmap <- getMap(resolution = "coarse")  # different resolutions available
plot(newmap)

data("countryExData",envir=environment(),package="rworldmap")
sPDF <- joinCountryData2Map(countryExData, joinCode = "ISO3", nameJoinColumn =
"ISO3V10")
mapCountryData(sPDF, nameColumnToPlot="BIODIVERSITY")
```



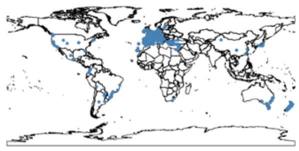
Exercice 3 SPATIAL VECTOR DATA (points, lines, polygons)

```
# Example dataset: retrieve point occurrence data from GBIF
library(dismo)
install.packages("jsonlite")
library(jsonlite)

laurus <- gbif("Laurus", "nobilis")# get data frame with spatial coordinates (points)
locs <- subset(laurus, select = c("country", "lat", "lon"))
head(locs)  # a simple data frame with coordinates
# Discard data with errors in coordinates:
locs <- subset(locs, locs$lat < 90)
# Making data 'spatial’
coordinates(locs) <- c("lon", "lat")  # set spatial coordinates
plot(locs)
# Define spatial projection (http://www.spatialreference.org/)
crs.geo <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
# geographical, datum WGS84
proj4string(locs) <- crs.geo
# define projection system of our data
summary(locs)
```

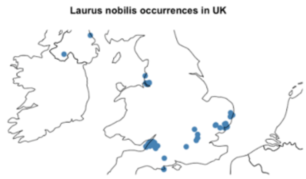
Exercice 3

SPATIAL VECTOR DATA
(points, lines, polygons)

```
# Quickly plotting point data on a map
library(rworldmap)
# library rworldmap provides different types of global maps, e.g:data(coastsCoarse)
plot(coastsCoarse)
data(countriesLow)
plot(countriesLow)
points(locs, pch = 20, col = "steelblue")
```



```
# Subsetting and mapping again
locs.gb <- subset(locs, locs$country == "United Kingdom")
# select only locs in UK
plot(locs.gb, pch = 20, cex = 2, col = "steelblue")
title("Laurus nobilis occurrences in UK")
plot(countriesLow, add = T)
```

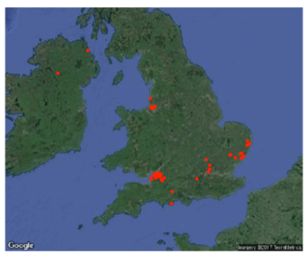


UNIVERSITÉ DE GENÈVE

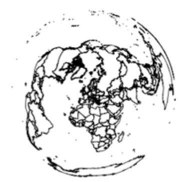
Exercice 4

SPATIAL VECTOR DATA
(points, lines, polygons)

```
# Mapping vectorial data (points, polygons, polylines)
gbmap <- gmap(locs.gb, type = "satellite")
locs.gb.merc <- Mercator(locs.gb)
# Google Maps are in Mercator projection. # This function projects the points to that projection to enable mapping
plot(gbmap)
points(locs.gb.merc, pch = 20, col = "red")
```



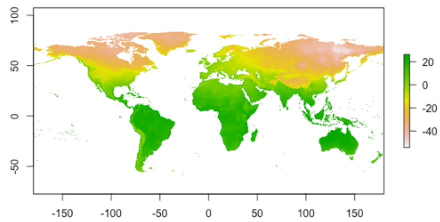
```
# Projecting shapefile of countries
plot(countriesLow)  # countries map in geographical projection
crs.laea <- CRS("+proj=laea +lat_0=52 +lon_0=10 +x_0=4321000 +y_0=3210000 +ellps=GRS80 +units=m +no_defs")
# Lambert Azimuthal Equal Area
country.laea <- spTransform(countriesLow, crs.laea)
 # project
plot(country.laea)
```



Exercice 5

USING RASTER (GRID) DATA

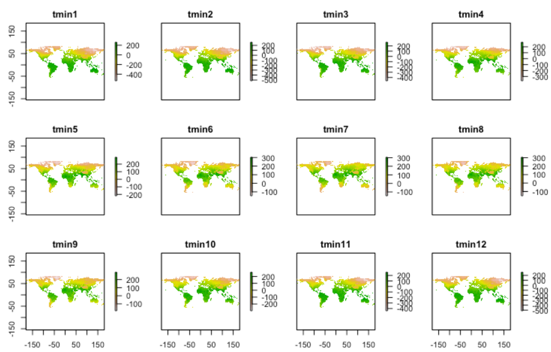
```
# Downloading raster climate data from internet
library(dismo)
tmin <- getData("worldclim", var = "tmin", res = 10)
# this will download

# Loading a raster layer
tmin1 <- raster(paste(getwd(), "/wc10/tmin1.bil", sep = ""))  # Tmin for January
tmin1 <- tmin1/10  # Worldclim temperature data come in decimal degrees
tmin1

# look at the info# global data on minimum temperature
plot(tmin1)
```



Exercice 6

USING RASTER (GRID) DATA

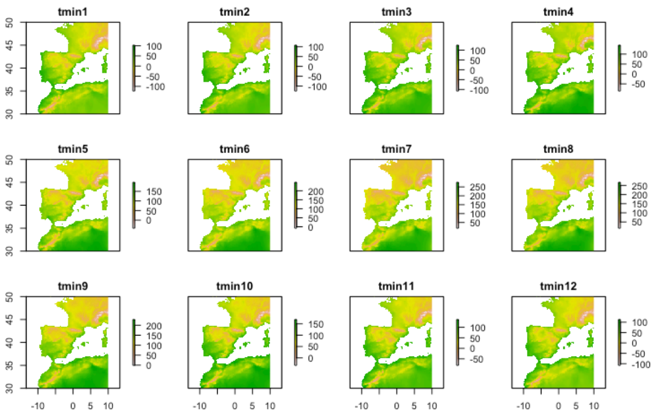
```
# Creating a raster stack
# A raster stack is collection of many raster layers with the same projection,
spatial extent and resolution. Let's collect several raster files from disk and read
them as a single raster stack:

install.packages("gtools")
library(gtools)
file.remove(paste(getwd(), "/wc10/", "tmin_10m_bil.zip", sep = ""))
list.ras <- mixedsort(list.files(paste(getwd(), "/wc10/", sep = ""), full.names = T,
pattern = ".bil"))

list.ras  # a list of the files containing
monthly temperature values
tmin.all <- stack(list.ras)
tmin.all
tmin.all <- tmin.all/10
plot(tmin.all)
```



Exercice 7

USING RASTER (GRID) DATA

```
# Crop rasters
plot(tmin1)
newext <- drawExtent()
# click twice on the map to select the region of interest

tmin1.c <- crop(tmin1, newext)
plot(tmin1.c)
```

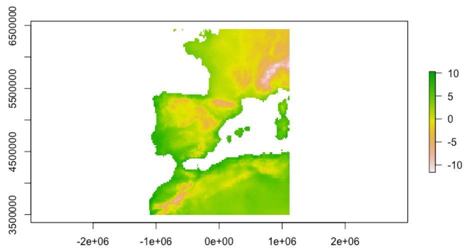



```
# Alternatively, provide coordinates for the limits of the region of interest:
newext <- c(-10, 10, 30, 50)
tmin1.c <- crop(tmin1, newext)
plot(tmin1.c)
tmin.all.c <- crop(tmin.all, newext)
plot(tmin.all.c)
```

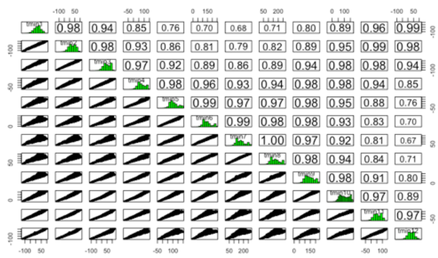
UNIVERSITÉ DE GENÈVE

Exercice 8

USING RASTER (GRID) DATA



```
# Define spatial projection of the rasters
crs.geo  # defined above
projection(tmin1.c) <- crs.geo
projection(tmin.all.c) <- crs.geo
tmin1.c  # notice info at coord.ref.

# Changing projection
tmin1.proj <- projectRaster(tmin1.c, crs = "+proj=merc +lon_0=0 +k=1 +x_0=0
+y_0=0 +a=6378137 +b=6378137 +units=m +no_defs")
# can also use a template raster, see ?projectRaster
tmin1.proj  # notice info at coord.ref.
plot(tmin1.proj)

# Plotting raster data
hist(tmin1.c)
pairs(tmin.all.c)
```



UNIVERSITÉ DE GENÈVE

Exercice 9

MAP YOUR OWN DATA

```
require(rgdal)
install.packages("ggplot2")
require(ggplot2)
install.packages("rgeos")
require(rgeos)
install.packages("ggmap")
require(ggmap)
install.packages("RColorBrewer")
require(RColorBrewer)

# Read shapefile using OGR
setwd("GIS")
shp1 = "wsrhone.shp"
myshp1 = readOGR(shp1, layer = "wsrhone")
shp2 = "RhoneRivers.shp"
myshp2 = readOGR(shp2, layer = "RhoneRivers")

# Convert to lat long
myshp1_proj = spTransform(myshp1, CRS("+proj=longlat +datum=WGS84"))
myshp2_proj = spTransform(myshp2, CRS("+proj=longlat +datum=WGS84"))
```


Exercice 9

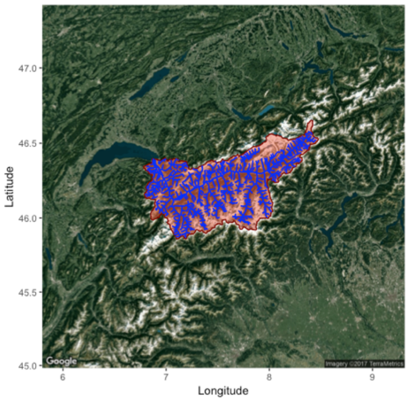
MAP YOUR OWN DATA

```
# Find polygon centroid (This centers the map)
centroid = gCentroid(myshp1_proj)

# Get the Google basemap
mapImageData1 = get_map(location = c(lon = centroid$x, lat = centroid$y),
color = "color", source = "google", maptype = "satellite", zoom = 8)

mapImageData2 = get_map(location = c(lon = centroid$x, lat = centroid$y),
color = "color", source = "google", maptype = "roadmap", zoom = 9)

# maptypes are “roadmap”, “terrain”, “satellite”, “hybrid”

# Convert shapefile to format ggmap can work with
polys1 = fortify(myshp1_proj)
polys2 = fortify(myshp2_proj)

# Define the color scheme for mapping shp
colors = brewer.pal(9, "OrRd")
```

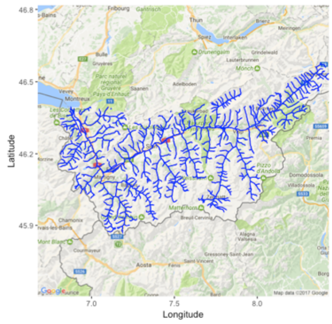
Exercice 9

MAP YOUR OWN DATA

```
# create a first map with your data
ggmap(mapImageData1) +
   geom_polygon(aes(x = long, y = lat, group =
group), data = polys1, color = colors[9], fill =
colors[6], alpha = 0.5) +
  geom_polygon(aes(x = long, y = lat, group =
group), data = polys2, color = "blue", alpha = 0.5) +
  labs(x = "Longitude", y = "Latitude")
```




```
# create another map
ggmap(mapImageData2) +
 geom_polygon(aes(x = long, y = lat, group =
group), data = polys2, color = "blue", alpha = 0.5)+
  labs(x = "Longitude", y = "Latitude")
```



UNIVERSITÉ DE GENÈVE

https://r-arcgis.github.io

UNIVERSITÉ DE GENÈVE

CRAN

CRAN 1.0.2 | downloads 1419/month | downloads 11K

RQGIS establishes an interface between R and QGIS, i.e. it allows the user to access QGIS functionalities from within R. It achieves this by establishing a tunnel to the Python QGIS API via the reticulate-package. This provides the user with an extensive suite of GIS functions, since QGIS allows you to call native as well as third-party algorithms via its processing framework (see also https://docs.qgis.org/2.14/en/docs/user_manual/processing/index.html). Third-party providers include among others GDAL, GRASS GIS, SAGA GIS, the Orfeo Toolbox, TauDEM and tools for LiDAR data. RQGIS brings you this incredibly powerful geoprocessing environment to the R console.

The main advantages of RQGIS are:

1. It provides access to QGIS functionalities. Thereby, it calls Python QGIS API but R users can stay in their programming environment of choice without having to touch Python.
2. It offers a broad suite of geoalgorithms making it possible to solve most GIS problems.
3. R users can use just one package (RQGIS) instead of using RSAGA and rgrass7 to access SAGA and GRASS functions. This, however, does not mean that RSAGA and rgrass7 are obsolete since both packages offer various other advantages. For instance, RSAGA provides many user-friendly and ready-to-use GIS functions such as `rsaga.slope.asp.curv` and `multi.focal.function`.

https://github.com/jannes-m/RQGIS

UNIVERSITÉ DE GENÈVE

UNIVERSITÉ DE GENÈVE

CERTIFICAT COMPLÉMENTAIRE EN GÉOMATIQUE

unige.ch annuaire

Accueil | Plan du site | Portail UNIGE

Université de Genève > TIGERS > Certificat complémentaire en géomatique

- **Accueil**
- **Règlements et plan d'études**
- **Enseignements et horaires**
- **Mémoires du Certificat de géomatique**
- **Informations pratiques et inscriptions**
- **Contact et organisation**

à la Une

Certificat de Géomatique, le film

Inscriptions 2018 closes

Tweets de @TigersUnig

TIGERS a retweeté

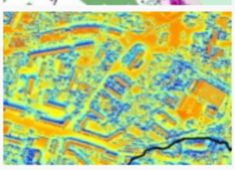
ESA Climate Office
@esaclimate

How has ozone changed over time? The Ozone team in ESA's Climate Change Initiative have harmonised data from different satellites to produce a long climate data record, so we can now identify ozone trends.
esa.int/Our_Activities…

Intégrer Voir sur Twitter

Le certificat complémentaire en géomatique (CCG) est un cursus de formation de base, délivré conjointement par la Faculté des Sciences de la Société (ci-après SdS) et la Faculté des Sciences, et en collaboration avec l'Institut des Sciences de l'Environnement (ISE). Le certificat complémentaire en géomatique est coordonné et administré par l'Institut/Pôle en gouvernance environnementale et développement territorial (Faculté SdS et Institut des Sciences de l'environnement).

L'information géographique a acquis depuis plusieurs années une place fondamentale dans la prise de décision et l'analyse des phénomènes environnementaux et sociaux. De ce point de vue, la conception, la gestion et l'utilisation des systèmes d'information géographique (SIG) constituent aujourd'hui des activités stratégiques pour les organisations et administrations publiques, la recherche, mais également pour les entreprises privées, désireuses de connaître variables et caractéristiques de leurs terrains de compétence ou d'action.



L'objectif du Certificat complémentaire en Géomatique est d'offrir une formation centrée sur les nouvelles technologies d'information à référence spatiale, dans une perspective interdisciplinaire. Les cours visent en particulier à confronter l'étudiant aux questionnements des sciences tant sociales qu'environnementales mobilisant l'information géographique. Ces perspectives sont mises en pratique lors de travaux sur ordinateur - environ la moitié du nombre total des heures d'enseignement - et finalisées par un mémoire.

www.unige.ch/cgeom/

UNIVERSITÉ DE GENÈVE

Spatial thanks

Anthony.lehmann@unige.ch

UNIVERSITÉ DE GENÈVE