Machine Learning as a Framework for Spatial and SpatioTemporal Prediction

ISRIC World Soil Information

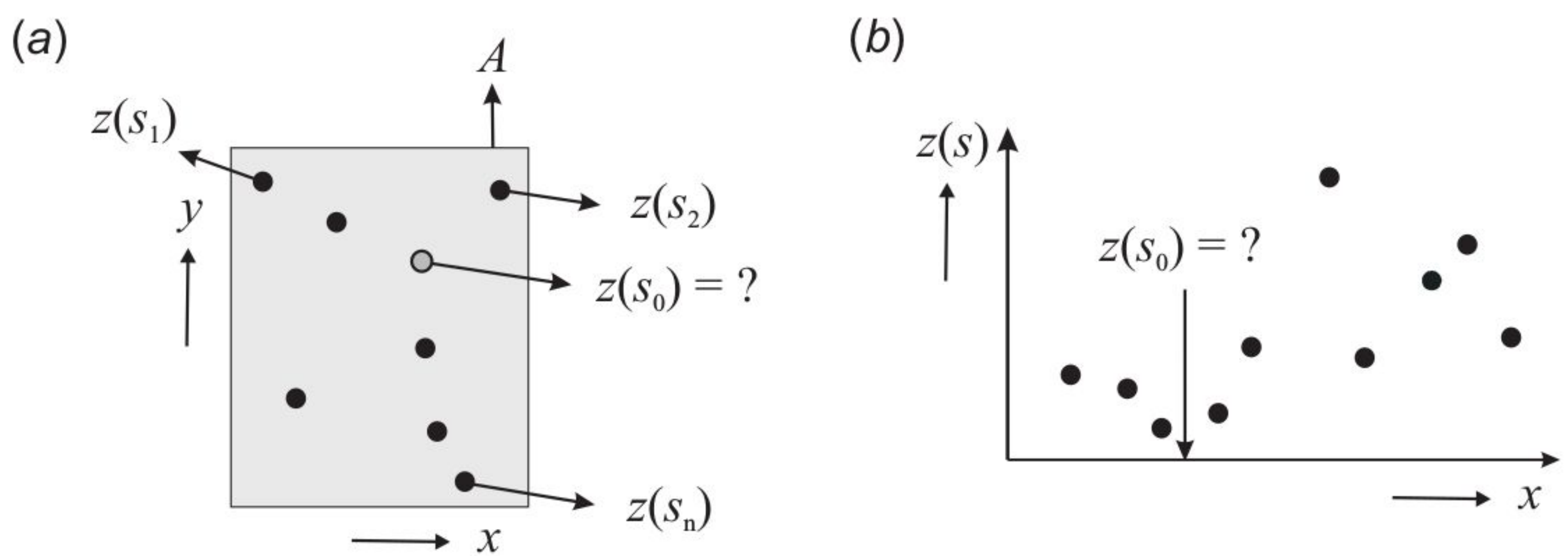
T. (Tom) Hengl <tom.hengl@isric.org>

Could machine learning (completely) replace kriging?

Assuming that the samples are *representative*, *non-preferential* and *consistent*, values of the target variable at some new location $\mathbf{s}_0$ can be derived using a **spatial prediction model**. In statistical terms, a spatial prediction model draws realizations — either the most probable or a set of equiprobable realizations — of the feature of interest given a list of inputs:

$$\hat{z}(\mathbf{s}_0) = E\left\{Z | z(\mathbf{s}_i),\ q_k(\mathbf{s}_0),\ \gamma(\mathbf{h}),\ \mathbf{s} \in \mathbb{A}\right\} \tag{1.1.2}$$

where $z(\mathbf{s}_i)$ is the input point data set, $\gamma(\mathbf{h})$ is the covariance model defining the spatial autocorrelation structure (see further Fig. 2.1), and $q_k(\mathbf{s}_0)$ is the list of deterministic predictors, also known as *covariates* or explanatory variables, which need to be available at any location within $\mathbb{A}$. In other words, a spatial prediction model comprises list of procedures to generate predictions of value of interest given the calibration data and spatial domain of interest.

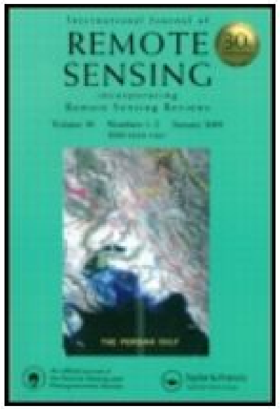


International Journal of Remote Sensing

Volume 32, Issue 15, 2011

Select Language ▼

Translator disclaimer



Original Articles

Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment

DOI: 10.1080/01431161.2011.552923

Robert Gilmore Pontius Jr[a*] & Marco Millones[a]

pages 4407-4429

Publishing models and article dates explained

Received: 27 Aug 2010

Accepted: 20 Dec 2010

Published online: 17 Aug 2011

Preview

Full text HTML

PDF

Access options

Alert me

Journal of Machine Learning Research 7 (2006) 983–999 Submitted 10/05; Revised 2/06; Published 6/06

Quantile Regression Forests

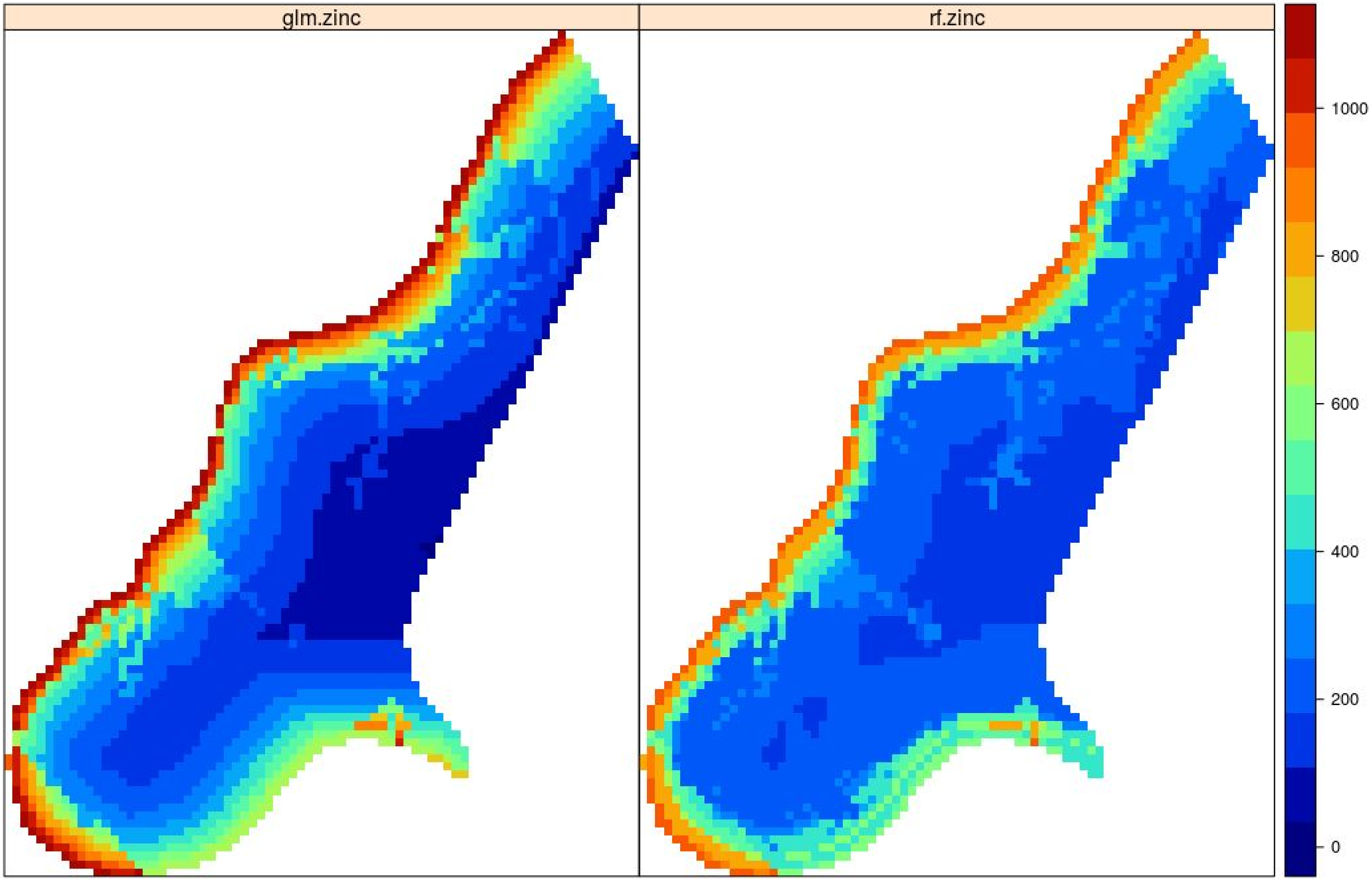
Nicolai Meinshausen NICOLAI@STAT.MATH.ETHZ.CH
Seminar für Statistik
ETH Zürich
8092 Zürich, Switzerland

Editor: Greg Ridgeway

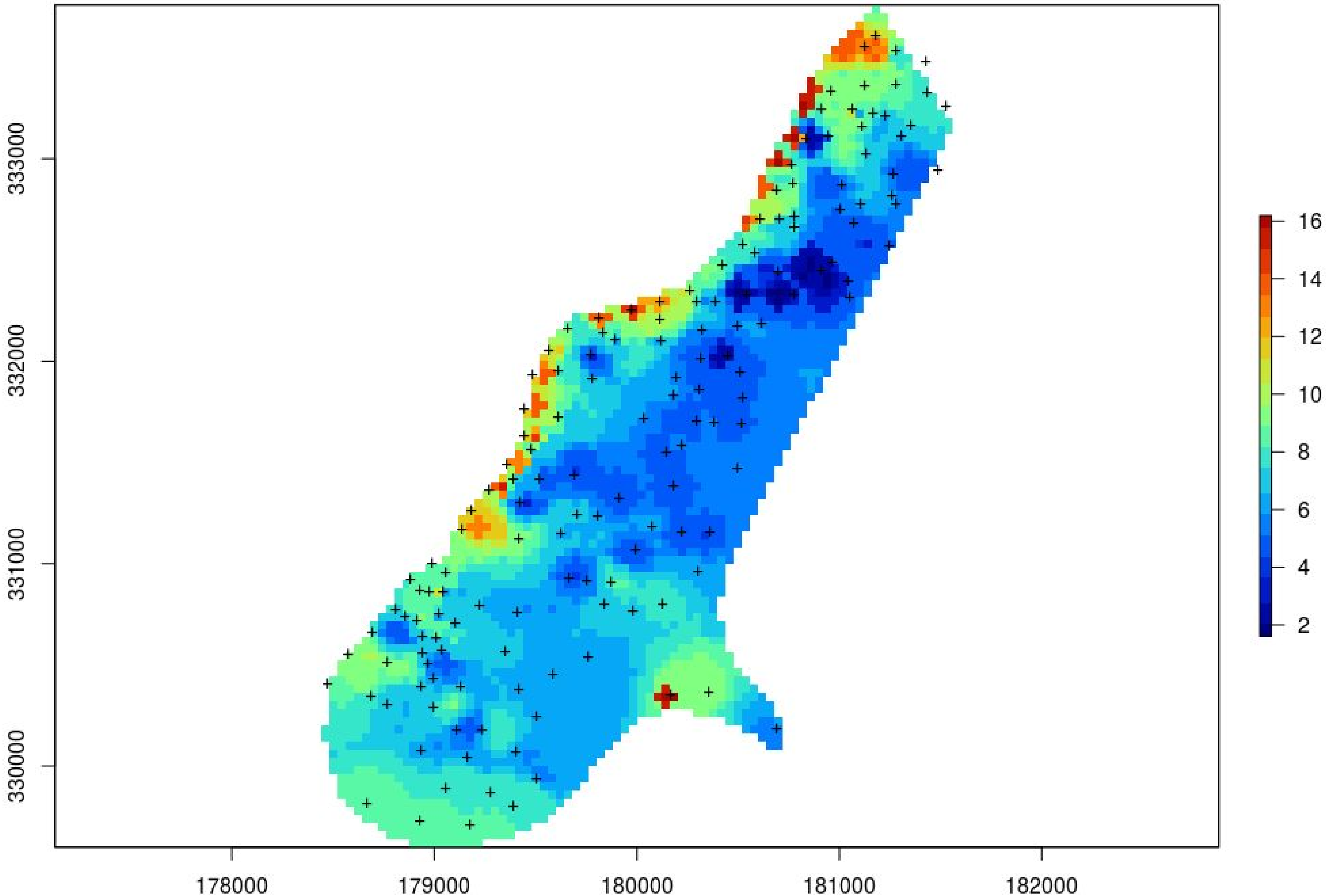
Abstract

Random forests were introduced as a machine learning tool in Breiman (2001) and have since proven to be very popular and powerful for high-dimensional regression and classification. For regression, random forests give an accurate approximation of the conditional mean of a response variable. It is shown here that random forests provide information about the full conditional distribution of the response variable, not only about the conditional mean. Conditional quantiles can be inferred with quantile regression forests, a generalisation of random forests. Quantile regression forests give a non-parametric and accurate way of estimating conditional quantiles for high-dimensional predictor variables. The algorithm is shown to be consistent. Numerical examples suggest that the algorithm is competitive in terms of predictive power.

Keywords: quantile regression, random forests, adaptive neighborhood regression


glm.zinc
rf.zinc
1000
800
600
400
200
0

Can you believe that this map was made using RF?



ISRIC World Soil Information

GLM vs RF

```
m1 = glm(zinc~log1p(dist)+ffreq, meuse,
family=gaussian(link=log))

m2 = quantregForest(x=meuse@data[,c
("dist","ffreq")], y=meuse$zinc)
```

Conclusions

Advantages of using quantile regression forests are: (1) both geographical and feature spaces are considered simultaneously, (2) more complex distances e.g. based on watershed connectivity can be incorporated into model building, and (3) spatial prediction can be speed-up because minimum human interaction is required and most of processes can be parallelized.

Conclusions

Method is also useful for sorting the predictor variables based on importance, for visualizing complex non-linear relationships, and for highlighting possible outliers and blunders in the input data. A fully automated spatial/spatiotemporal prediction framework based on machine learning is a realistic possibility, especially if the input training points are of high quality and representative of both feature and geographical spaces.

Model-based geostatistics

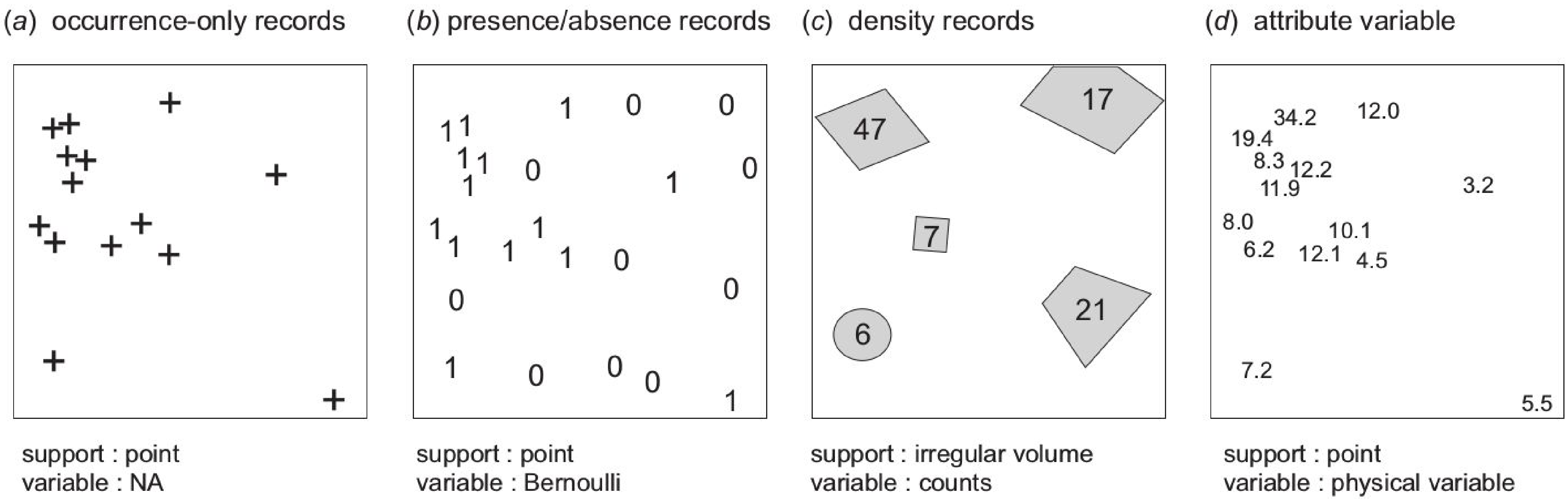


Fig. 1.3: Types of field records in ecology.