Uncertainty in Remote Sensing: Validating Probabilistic Pixel Level Classification

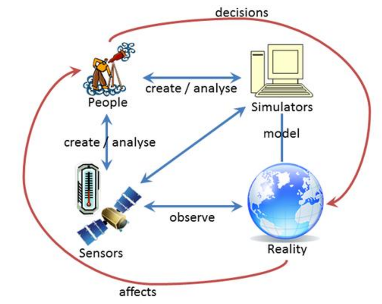
Michail Vrettas[1], Dan Cornford[1], Lucy Bastin[1], Xavier Pons[2], Eva Sevillano[2], Gerard Moré[3], Pere Serra[2], and Miquel Ninyerola[4]

[1]NCRG and Computer Science, Aston University, Birmingham, UK ; [2]Department of Geography, Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain;

[3]CREAF, Cerdanyola del Vallès 08193, Spain; [4]Department of Botanics, Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain

Quality, Uncertainty and the Importance of Validation

Remote sensing plays a key role in Earth observation. However all Earth observation data is subject to uncertainties from a range of sources including instrument error, retrieval error, and resolution error (representativity and scale). Quantifying these uncertainties is essential to rational and optimal use of the remotely sensed data, for example in data assimilation or decision making. In this work we focus on remote sensing of land cover, where the target is to provide a classification into a discrete set of classes. Traditional approaches to classification have provided a single estimate for the most probable class of the pixel and although many also compute class probabilities, these are often not validated, used further or provided to the user.

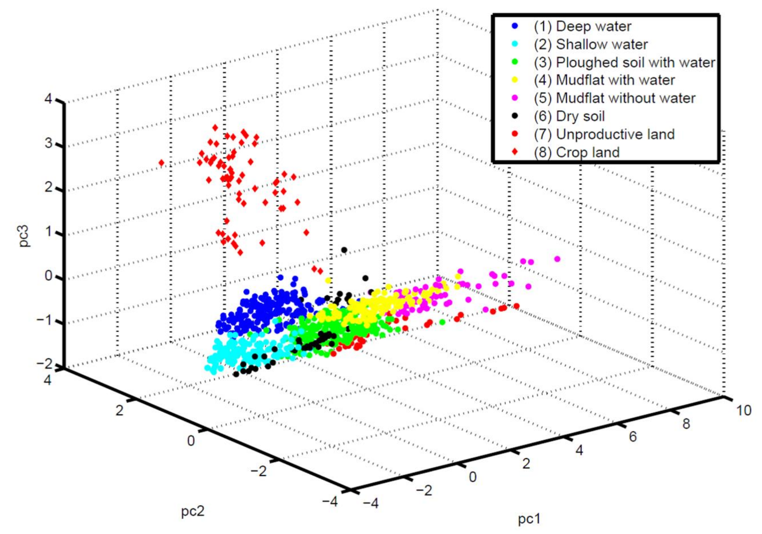


The Context: Flooding of Rice Fields

from Moré *et al*, IJRS 2011

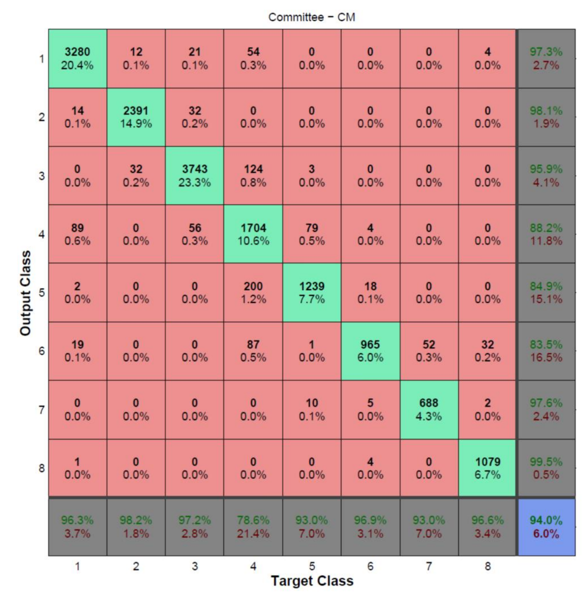
This work considers data derived from the problem of classifying pixel level Landsat data from the Ebre delta in Spain, to determine the degree of flooding in a rice growing region. This is important because management practices have environmental impacts and attract government subsidies. Our focus is on producing an accurate classification, but also a reliable probabilistic prediction of the classification uncertainty. This will then be encoded in UncertML and provided to the users within the GeoViQua quality model.

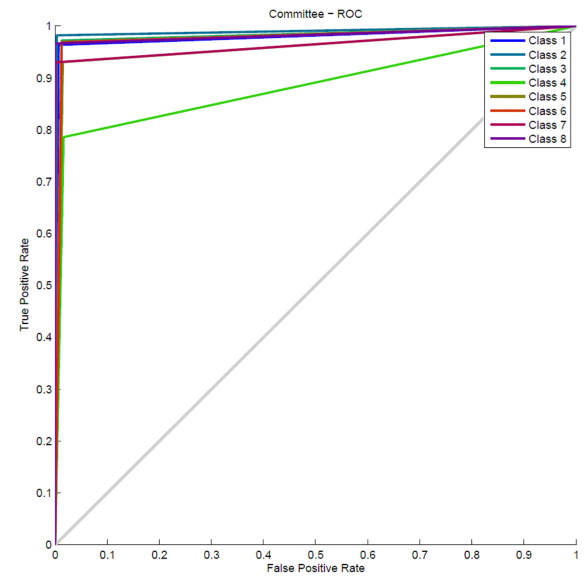
The data consists of radiometrically corrected, georeferenced LandSat data from several times across 3 years. There is also some labelled 'ground truth' data collected from across the time period at specific locations. The land cover is classified into 8 different classes which primarily reflect the extent of flooding in the fields. The data set is split into two parts, one for learning, the other for validation. The plot below shows the projection of the data onto the leading three principal components of the training set – showing there is reasonable, but not perfect class separation. We applied a range of Bayesian classifiers, from simple Linear Discriminant Analysis to a complex variational Gaussian process classifier. All performed reasonably on the independent test set, with the best classifier being an equally weighted committee.



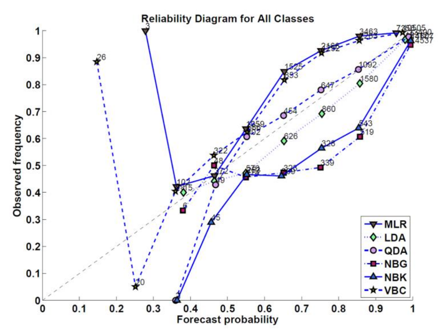
Moré G., Serra P., Pons X. 2011 "Multi-temporal flooding dynamics of rice fields by means of discriminant analysis of radiometrically corrected remote sensing imagery" International Journal of Remote Sensing 31 (7):1983-2011

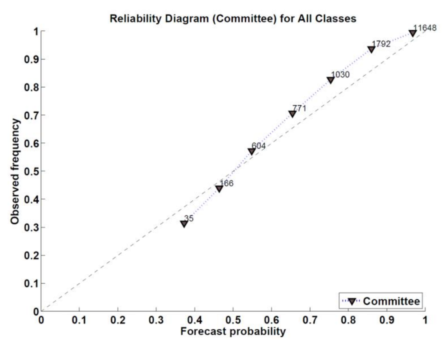
Validation of Probabilistic Models





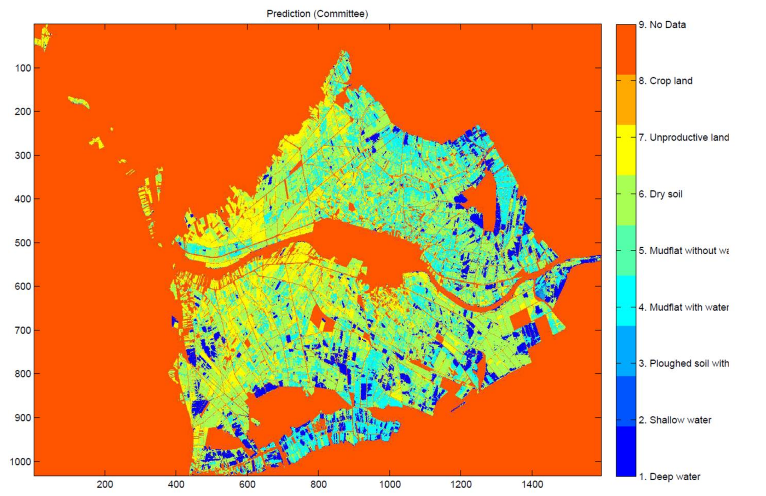
Typically validation of classification methods proceeds by looking at confusion matrices for an independent validation data set. These can be summarised in various ways, for example the widely used kappa statistic. However this is only half the story: for probabilistic methods Receiver Operator Characteristic (ROC) curves are useful, but are most relevant for algorithm comparison. We advocate the use of reliability diagrams which show the predicted probability of class membership against the frequency that the predictions are correct. For a statistically reliable method the points should plot on a one to one line. They are also useful diagnostics, helping to identify why a probabilistic method is performing badly.

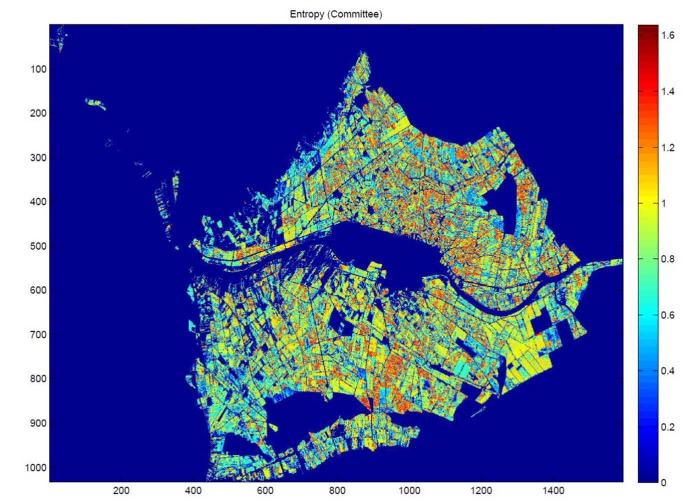




Summary and Outlook

Having validated the probabilistic classifiers, we have applied these to the complete images, as shown below. There are two maps – on the left is the most probable class in each pixel, and on the right is the entropy $(\sum p_i \log[p_i])$. This image clearly shows that some locations are far more reliably predicted (low entropy) that others (high entropy). Users of this data would ideally be provided with information on the uncertainty since this could impact their decision. As a next step we plan to propagate this uncertainty to produce an estimate of the probability a field was flooded for the period required, and use this to inform subsidy payment.





The next challenge is to convey this uncertainty information to the users. They are likely to be interested in two scenarios:

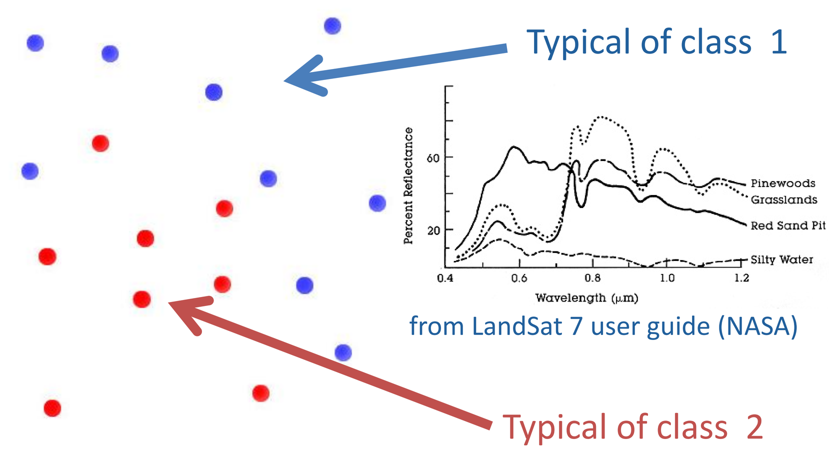
discovery - identify a suitable data set with sufficiently small uncertainty;

use - propagate the uncertainty through some workflow.

At the discovery level summary information is most useful at the data set or product level and might be an overall summary, or a confusion matrix. At the use level the per pixel probability of each class is most useful – all this can be provided using the GeoViQua quality model. There remain interesting questions about mixed pixels, which complicate the interpretation of the probabilities.

Classification in Remote Sensing

Remote sensing approaches to classification often look for prototypical members for each class, then classify in sometimes rather adhoc ways often based on distance (in some space). It is then necessary to produce post-hoc estimates of uncertainty.



from LandSat 7 user guide (NASA)

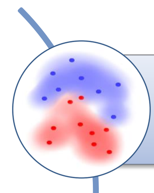
There are some benefits of this approach however – being less model based allows the scientist to use the large amounts of unlabelled data, for example by clustering to find representative prototypes.

The key issue will be that all aspects of the models are validated, including probabilistic aspects. So long as this is done rigorously on an independent validation set, not used at all in training, it is acceptable to use any method for classification.

In general it is much easier to produce probabilities **at the same time** as undertaking the classification. This is also true for continuous variable retrieval.

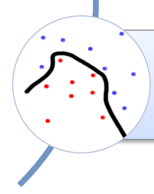
Classification in Machine Learning

Assume we have a target, t (class labels) and inputs, x (reflectance etc.). Our aim is to obtain p(t|x). Two main approaches are considered.



Generative approach

- Estimate p(x|t), use Bayes theorem to invert: p(t|x). Requires a model based approach for p(x|t) – e.g. mixture models.



Discriminative approach

- Directly estimate p(t|x), maximise the likelihood. Typically models the decision boundary,, often in a kernel space

Both have their merits: discriminative approaches tend to be simpler and faster, and can be tweaked to give class probabilities, typically based on distance to decision boundaries, for example in support vector machines. Generative approaches are more naturally probabilistic but inference in these can be a little tricky. Any classifier requires probabilistic validation.

www.geoviqua.org

GeoViQua is funded by the EC under contract 265178 as part of ENV.2010.4.1.2-2: Integrating new data visualisation approaches of Earth systems into GEOSS development