

Slide 1 – Title Slide

Hello and welcome to Week 11, part 6 of EGM310: Accuracy Assessment. In this lesson, we will learn about how we can evaluate how well our classification has done in identifying objects and surfaces in images.

Slide 2 – Evaluating classification

To be useful, we need to have an understanding of how accurate our classification is. Any classification will have errors, but whether we can trust the results, and whether we can use them to draw conclusions about what it is that we're studying, depends on being able to quantify those errors. We have a number of different ways that we can try to evaluate a classification, but perhaps the most common methods works as follows. First, we take a random sampling of points within our study area. Typically, we want to sample around 1-2% of the pixels in an image. We then look at each of these points, and determine what class they should belong to. If we can, we might also go and check these points in the field, in a process known as 'ground-truthing.' Finally, we compare the results of our manual classification with the results of the automatic classification.

Slide 3 – Confusion matrix

When we do this comparison, we most often use something known as a confusion matrix. I've seen these displayed in different ways – sometimes, you might see the 'true' values up here as the columns, while the classified results are in the rows; other times, you might see it displayed as I've done it here, with the columns moving this way representing how the pixel was classified by the computer, and the rows, moving this way, represent the 'actual', or 'true' class for the pixel. So as we move across the top row here, we see that we have pixels that belong to the A class that the machine correctly classified as A, pixels that were incorrectly classified as B, and pixels that were incorrectly classified as C. And, we can keep going like this.

The overall accuracy – that is, the percentage of our sample pixels that were correctly classified by the computer, is the sum of the diagonal – the true A, true B, and true C – divided by the total number of sample pixels. This is one way of reporting the accuracy of the classification, but it is not the only way.

For instance, it doesn't tell us anything about the pixels that were incorrectly labeled as a different class – the false B and false C pixels in the top row here. This is what is known as an error of omission – the classification has omitted these, or left them out, of the correct class. It's also known as a false negative, or a type-II error. We can calculate the error of omission for our A class by adding the number of false B results to the number of false C results in the top row here, and dividing by the number of pixels that should be classified as A. And, we can do this for the other classes as well.

A second type of error that we might want to know about are errors of commission – that is, pixels that have incorrectly been included in the wrong class. These are also known as false positive results, or type-I errors. The way that we calculate this for our A class is by taking the pixels that should have

been classified as B but were classified as A, and adding these to the pixels that should have been classified as C but were classified as A, and dividing by the number of pixels in the A class.

Next up, we have the Producer's accuracy – and this is the probability that the class has been correctly classified. In other words, it tells us how well our classification scheme has done classifying a particular class. This is calculated as either the number of correctly classified A pixels divided by the number of pixels that are actually A, or as 1 minus the error of omission.

Finally, we have what's called the User's accuracy – this is the probability that a given pixel on the map is correctly classified. In other words, it gives us an idea of how well we can trust that the classified map actually represents the truth. We can calculate the user's accuracy as the number of correctly classified A pixels divided by the number of pixels that are classified as A, or as 1 minus the error of commission.

Slide 4 – Kappa coefficient

But, this isn't everything. It is possible that our classification only looks good as a result of random chance – which doesn't mean we've done a good job classifying the image, it just means that we got lucky. One way that we can check this is by calculating something called a kappa coefficient, or Cohen's Kappa coefficient, or sometimes just kappa. Whatever you call it, this is an indicator of whether the accuracy is due to random chance or not. It's calculated by subtracting the probability of chance agreement from the observed agreement, and dividing by 1 minus the probability of chance agreement. And of course, we can calculate the probability of chance agreement: it's just the sum for each class of the percentage of our sample pixels that belong to that class times the percentage of our sample pixels that are classified as that class. We'll work out an actual example on the next slide, so don't worry too much about this just yet.

Kappa typically has values between 0 and 1. We can have values less than zero, but this is really bad. It means that our classification algorithm has done worse than random chance – not a good sign. If kappa equals zero, it means that we haven't done any better than random chance – again, not great. It suggests that our classification isn't particularly reliable. Values greater than about 0.5 are considered moderate or good, depending on who you talk to. To put this another way, a value of 0.5 means that we've done 50% better than we would expect just by random chance.

Slide 5 – A worked example

This is a lot to take in, so I'll try to go through a worked example. To make the maths much simpler, I'll stick to 2 classes, land and water, and say that we've drawn a sample of 30 pixels from both land and water. From our confusion matrix, you see that we've correctly classified 24 pixels as water, and 21 pixels as land. So, our overall accuracy can be calculated as: $24 + 21$ divided by $30 + 30$, or 0.75. An overall accuracy of 75% is generally considered acceptable to good, again, depending on who you talk to. Next up, let's calculate the User's accuracy. For water, this is the number of correctly classified pixels, 24, divided by the total number of classified pixels, which was 33. And $24 / 33 = 0.73$. For land, this is 21 divided by 27, which gives us 0.78. This tells us that a pixel on the map that is classified as

land is slight more likely to be correct than a pixel classified as water. To calculate the producer's accuracy for water, we divide the number of correctly classified pixels, 24, by the number of sample pixels, 30, to get 0.8. For land, we have 21 divided by 30, which is 0.7. So this means that we are doing a better job classifying water pixels than we are classifying land pixels, but with the user's accuracy, we also have a higher rate of false positives for water.

To calculate the Kappa coefficient, we can start with this formula here. For water, the chance agreement is the % of our sample pixels that are actually water, which is 30 divided by 60. We multiply this by the % of our pixels that are classified as water, which is 33 divided by 60. The % actual land is also 30 over 60, and the % classified land is 27 divided by 60. If we put all of this together, we should get 0.5. So we have a 50-50 chance of correctly identifying a pixel, based on our sample.

If we then plug this into our formula for kappa, we get the overall accuracy, 0.75, minus the chance agreement, 0.5, divided by 1 minus 0.5. All of this gives us a kappa value of 0.5 – our classification scheme is 50% better than what we would expect by random chance. So, this is a moderately successful classification. Hooray!

Slide 6 – Summary

In this lesson, we have learned that accuracy assessment is the key to understanding how reliable our classification is. If we don't know how good a job we've done in our classification, we can't know how useful our results are.

To do this, we compare our classified image to either ground-truth data that we've collected in the field, or by manually classifying test points or test regions in our image. Often, we might use a combination of these two approaches.

We want to assess both the overall and by-class accuracy – this can help us to understand where we might need to improve our training samples, or our input data, to help us improve the classification.

Finally, we learned how to compare our results to the likelihood that our results are due to random chance, by calculating the kappa coefficient.

Slide 7 – Additional resources

Once again, you can read more about the concepts we've covered in this lesson in the textbooks, Chapter 7 of Lillesand, Kiefer & Chipman; and Chapter 14 of Campbell & Wynne. I've also included links to information about performing accuracy assessment for both of the two software packages we've used in this class, ArcMap and ERDAS Imagine. There's also a link to a shorter video that shows how to do this in ERDAS. That's all for this lesson – I hope you found it interesting, and if you have any questions, please don't hesitate to e-mail me or post in the discussion forum on blackboard. Bye!