Project Development Manual For Collect Earth Online

M.S. Patterson

Kimberly Mccallum

Produced by the U.S. Forest Service (USFS) Geospatial Technology and Applications Center (GTAC) with support from the United Nations Food and Agriculture Organization (FAO), the SERVIR Global Program, and the SilvaCarbon Global Program.

Contents

I. List of Figures

II. List of Tables

III. List of Equations

Acknowledgements

The authors and the U.S. Forest Service (USFS) Geospatial Technology and Applications Center (GTAC) would like to thank the United Nations Food and Agriculture Organization (FAO), the SERVIR Global program, and the SilvaCarbon Global program for providing essential funding and support for this manual. SERVIR is a joint development initiative of the U.S. National Aeronautics and Space Administration (NASA) and the U.S. Agency for International Development (USAID). SilvaCarbon is an interagency capacity-building initiative funded by USAID and the U.S. Department of State and implemented by the U.S. Forest Service (USFS), the U.S. Geological Survey (USGS), and other partners. The authors would also like to thank Dr. Karis Tenneson for her efforts to get this project off the ground and her continued efforts to move Collect Earth Online forward as a platform.

1. Introduction

Collect Earth Online (CEO) is a powerful online platform for performing sample-based image interpretation. CEO enables users to view global coverage from multiple high resolution imagery sources and assign classifications to geo-referenced points on the earth's surface. A sound theoretical base is required to effectively use this tool. This manual provides an introduction and overview of the body of knowledge needed to develop rigorous projects and perform accurate image interpretation using CEO. It also provides an introduction to the logistical and technical steps required to prepare and complete a CEO project.

The project planning and execution process is broken up into steps, which are discussed in the text and further explored through hands-on technical exercises, and are organized in the following manner:

> Chapter 1 provides an overview of the approaches or types of projects you might use CEO for;
>
> Chapters 2 and 3 cover the essentials of developing a land use and land cover system that can support your project's needs;
>
> Chapter 4 introduces image interpretation, and explains the fundamental components of performing robust image interpretation;
>
> Chapter 5 teaches the concepts of sample design and response design, and how to apply them to your project; and
>
> Chapter 6 presents the essential components of Quality Assurance/Quality Control, and how to incorporate them into your project to insure good quality data.
>
> The appendix contains a project planning worksheet that will help you organize all the information needed to plan and execute your project.

Though this text presents these steps in a particular order, in reality the planning of a project is an iterative process. Each step that you perform will be informed by, and in turn inform, all the others and they can't be considered in isolation from each other. Therefore it's best to plan to move through the process at least twice, creating a rough draft that you will then refine. The exercises presented in each chapter will help with this process, and you'll find a project planning worksheet in the appendix of this text.

The remainder of this chapter provides an overview of the sorts of projects that might make use of CEO.

1.1 Inventory of Land Use, Cover and/or Change

Overview

CEO is very useful for projects with the goal of estimating the amount of a feature (or set of features) on the landscape. Typically we organize these elements into two broad categories: land use and land cover. Land cover refers to the biophysical structure of the landscape. Land use, in contrast refers to how people organize or make use of the landscape. CEO can be used to support the production of numerical summaries (estimates) of the landscape (non-spatially explicit data), or in the production and validation of maps (spatially explicit data). As mentioned in the introductory material, there are two primary ways of producing this data: Point sample-based estimates of landscape elements and model-based estimates.

Point Sample-Based Estimates

A sample or point-based inventory involves: 1) generating a representative, statistically valid sample of features of interest from the landscape; 2) interpreting those sample points; and then 3) producing a summary of the interpreted data. This method can be used to provide estimates of tree canopy cover, impervious surface cover, abundance of crop types, or other land covers and uses. With a sufficiently large, well-constructed sample size, a point-based estimate can be as accurate as more complex methods[1], and can qualify as *bona fide* estimates, consistent with good practices[2]. A sample-based inventory can provide IPCC Approach 1 data (summaries of total land cover), as well as IPCC Approach 2 data (summaries of total land cover, plus information on conversions between classes) if properly structured[3].

When to Use Point Sampling

Point sampling approaches have the distinct advantage of being very simple to undertake. A simple random sample or a systematic grid of samples can be used to place points on the landscape. Because they are focused more on individual landscape elements than complex assemblages, they also often don't require the process of designing a complex Land Use/Land Cover (LULC) classification system. Simple random samples and systematically gridded samples may also be particularly suited to estimates of a single type of cover, such as percent impervious surface or tree canopy cover. These approaches can also be used to monitor change over time, simply by revisiting the sample points as new imagery becomes available or by reviewing a sequence of imagery for the points when they are being classified.

However, these approaches cannot produce generalized, spatially explicit data. No maps should be produced from the point data, in the vast majority of cases. Because the data is not spatially explicit, it becomes difficult to use this sort of estimate to investigate the management activates or natural processes that are shaping the landscape. Unfortunately, it's difficult to assess *why* a change is happening if you don't know *where* it is happening.

The short-comings of point-based estimates can in part be addressed by breaking them up spatially, and providing separate estimates for individual study regions. However, this means a rapid increase in sample size, if there is also a desire to maintain reasonable accuracy. It is also possible to collect more complex data about assemblages of elements on the landscape by summarizing point data connected to a plot, and then treating the plots as single points.

---

[1] Richardson, J. J., & Moskal, L. M. (2014). Uncertainty in urban forest canopy assessment: Lessons from Seattle, WA, USA. Urban Forestry & Urban Greening, 13(1), 152–157. https://doi.org/10.1016/j.ufug.2013.07.003

[2] That is, neither under- nor over-estimating, free from unnecessary bias, with uncertainty minimized to the extent practicable. See the IPCC *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories* (2000), Chapter 1.

[3] IPCC 2006, *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, Prepared by the National Greenhouse Gas Inventories Programme, Vol. 4, Chapter 3. Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds). Published: IGES, Japan.

Example/Use Cases

Point sample-based estimates lend themselves to simple binary classifications: for example, forest canopy cover or impervious surface area. However, they can also be used to measure more complex phenomena, like agricultural commodity distribution or more comprehensive LULC classification. Another use case for sample-based estimates is in updating and supporting field data programs that are collected on long intervals, thereby extending the utility of that data at relatively low cost.

An El Salvador pilot study to establish national estimates of LULC will be used an example of a point sample-based inventory project. This example will be explored further in exercises in following chapters. Another example is a project to estimate the abundance of commodity crops that are replacing forest canopy in Southeast Asia.

Model-Based Estimates

When to use Model-Based Estimates

Model-based inventories and estimates are appropriate when project objectives require spatially explicit information regarding the location and distribution of classes on the landscape. In other words, if what you want is a map, then you may wish to use a modeling approach to produce it. These projects are generally much larger in scope, requiring higher level expertise in modeling techniques as well as significantly more data and a greater number of processing and analytical steps. CEO can be used to collect training data for these models.

Example/Uses cases

A project to map land use, cover and change mapping in Ecuador will be used an example of a model-based inventory project. In this project, modeling techniques were used to generate wall-to-wall maps for two different time periods as well as change between the two time periods (2014 and 2016). From these maps, spatial locations and areas of different classes could be estimated and the pattern and distribution of classes could also be visualized.

1.2 Reference Data for Map Production and Validation

After a map has been produced, particularly after using some sort of modeling approach, it is necessary to validate that map by assessing its accuracy. The primary method for assessing accuracy of a map is through rigorously sampling the map, and then comparing the map data against a high-quality reference data source to see if it is correct. Maps produced through a modeling process should be subjected to an accuracy analysis to ensure that the map is acceptable for its intended use.

1.3 Photo interpretation

Photo interpretation (PI) is the process of looking at high or very high spatial resolution imagery (from satellites or aerial photography) and labeling the objects of interest in your sample locations. Photo interpretation is the core skill needed to effectively execute any CEO project. Good photo interpretation is needed to produce an accurate point-based inventory or to generate training data for a model. A well-trained photo interpreter is still considerably more powerful and accurate a tool than modeling techniques like random forest or maximum likelihood. The data set that a photo interpreter produces are considered to reflect the on-the-ground reality of a site more closely than modeled data, and this is why photo interpretation is used to assess the accuracy of map products. This manual presents an introduction to photo interpretation concepts and how to perform effective PI work.

2. Project Planning

2.1 Overview

This section introduces the information you'll need to begin designing a project in CEO. The steps include:

1. Identifying the project's goals, needs, available resources, and constraints;
2. Creating a system of land use and land cover (LULC) for your project;
3. Creating a photo interpretation key;
4. Establishing your sample and response design; and
5. Making a quality assurance and quality control (QA/QC) plan.

If you are working with a pre-existing LULC classification system, the following chapter may help you extend or modify it to more closely align with your projects goals.

This process may be presented as if it were strictly linear, beginning at step one and ending with data collection after step 5. In reality, each of these steps will influence the others. It's often useful to have developed your response design and unit of assessment before producing a key, for instance. However, you may not be able to determine what your response design will be until after you've reviewed a lot of imagery during key creation. Ultimately, it's best to consider the process presented here as one that is iterative; likely you'll want to review each step again after you've gone through them once. You'll arrive at a more sound, thoroughly considered project by iterating through the project creation process.

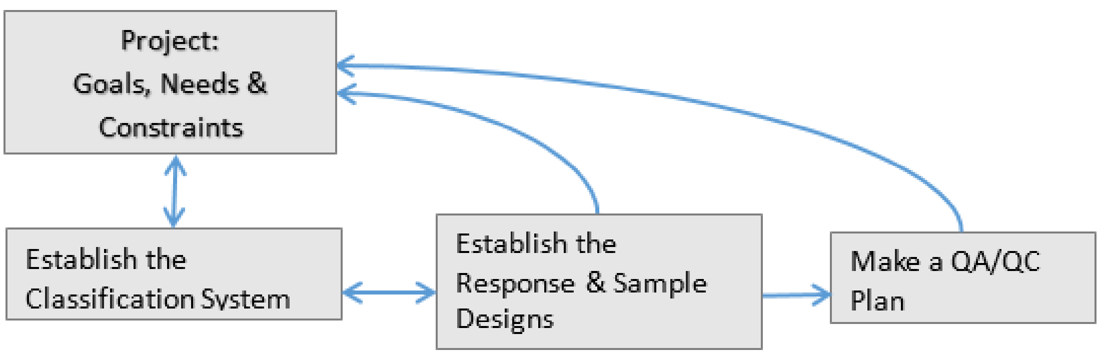

Figure 1. Iterative Collect Earth Online project planning process. At each stage, decisions are made and evaluated in terms of goals, needs and constraints. In some cases, goals may need to be revised.

Because of this, it's extremely useful to design a small pilot for your larger project. Running through the entire project creation process and collecting data will help to reveal any possible problems, unforeseen obstacles, and other considerations that will help you revise your approach to produce the best possible results.

2.2 Objectives, needs, available resources, and constraints

Objectives

The first task when designing a project for use in CEO is to clearly define the project objectives. CEO can be used to support a wide variety of remote sensing, mapping and estimation projects. However, it is primarily a platform for the collection of image interpretation data for plots and points, and therefore its capabilities and limitations should be well understood. Image interpretation is a powerful technique that can be used for a number of very useful purposes. For many remote sensing and mapping projects however, reference data from image interpretation must be used with other data and tools to accomplish the task.

CEO can be used to support both spatially explicit 'wall-to-wall' mapping projects as well as area-based estimations. CEO is well-suited to support estimation of the occurrence or areas of different types of features on the landscape. This could mean estimating tree canopy, area converted from forest to farms, or landscape classification. Each of these objectives could be met in a number of different ways, depending on whether or not the data needs to be spatially explicit. CEO can be used to support the production of numerical summaries of the landscape or in the production and validation of maps.

Do your project's objectives align with what CEO can do? If you want to estimate how much of something exists on your landscape, or need to use image interpretation to support a larger project, then CEO may be a very useful tool for your project.

Project Needs and Constraints

What resources do you need access to in order to meet your project's objectives? Typically for inventory work, map validation, or other photo interpretation-based work the primary need is a source of imagery of sufficient resolution. CEO includes a variety of imagery sources for use by everyone, however, these sources may or may not meet the needs of your project. A few questions to consider:

- Do you need multiple time steps?
- Will the work take place in an area that is frequently cloudy?
- Do you need to make use of false color imagery or other sources of data, such as radar?

A very large volume of high-quality earth observing data is now available to the public for free, and CEO allows for much of this data to be easily incorporated for use (readers are encouraged to investigate the CEO user's manual for information on how to include outside imagery in their projects).

In addition to imagery, there may be other needs for the project, including sufficiently skilled staff and financial resources to conduct the work. The process of clearly establishing the goals of the project, and working through the sample and response designs (discussed in detail in the rest of this manual) needed to meet those goals will help establish how long it will take and how expensive it might be to complete the project.

Obviously, projects with smaller budgets or that must be conducted with small staffing resources may encounter difficulties if they require very large amounts of work, the purchase of additional imagery, field work to support the analysis, or other constraints. However, it should be noted that CEO does allow for public participation, and it easily connects with Google Earth Engine, putting potentially very ambitious projects within the reach of small teams with the appropriate skills.

2.3 Examples

Many projects have successfully used CEO in support of a variety of mapping, estimation and monitoring projects, at a variety of scales. For example, the Central American Commission for Environment and Development (CCAD) and the German Corporation for International Cooperation (GIZ) recently completed a pilot study in El Salvador where CEO was used to characterize the country's historical trends in land use for the years 2000-2018. This project used photo interpretation at sampled locations to estimate areas of different land use categories. In Ecuador, the Ministry of the Environment is collaborating with RedCastle Resources (RCR) the Spatial Informatics Group (SIG), and FAO on a large scale, country wide project to map LULC change from 2014 to 2016. This project has developed wall-to-wall country wide data products and has used CEO at multiple steps to support data collection for model training as well as final map validation.

These two example projects clearly demonstrate the two main analysis modes of CEO: 1) use of photo interpretation at sample locations to support statistical estimation or 2) use of photo interpretation to collect reference data for mapping or validation. While these are just two examples that will be referred to frequently throughout this manual, CEO has been used to support land cover mapping and inventory projects around the world, including throughout Southeast Asia[4], Latin America, within the United States, and elsewhere. Projects have included validation of regional map products, inventories of agricultural commodities, and support of national forest monitoring systems.

Exercise 2.1: Planning your CEO project

Introduction

In this exercise, you will work through the process and considerations required for setting up a project in Collect Earth Online (CEO). You will start by reviewing the goals and photo interpretation objectives of two example projects: one using CEO to support an inventory and monitoring project and another using CEO to support vegetation mapping and map validation objectives. After reviewing these examples, you will check your knowledge before applying what you have learned to your own project. By the end of this exercise, you will understand how to clearly define the objectives for using CEO and be able to plan the desired end product. We will use this information to help guide and plan the development of a CEO classification scheme. This topic will be covered in the next exercise.

Objectives

- Learn how other projects have used CEO to accomplish clearly defined objectives
- Understand what objectives can be met using CEO
- Outline your own CEO project that can meet your project information needs

Part 1: Review example projects using CEO

Image interpretation methods can support both inventory and mapping applications. These two general use types result in very different end products and require different approaches in planning and implementation in CEO. The two examples that you will review in this part of the exercise broadly demonstrate these different application types and will be referred to throughout the exercises in this manual.

[4] http://servir-rlcms.appspot.com/static/html/map.html

A. El Salvador Pilot Study using CEO: Project Summary

In 2018 and 2019, a pilot study was conducted in El Salvador to characterize the country's historical trends in land use for the years 2000–2018. The pilot project was located in a subset of the country encompassing a land area of 265,536 ha. CEO was used to characterize the composition and changes in land use from 2001, 2012 and 2018 imagery. To do this, high resolution imagery was interpreted for plots in a systematic grid across the study area. Interpreters recorded land use classes at points within each of the plots for the three different image time periods. The proportions of the different land uses were estimated for the entire study area by summarizing the data interpreted from all of the plots in each time period. The proportions were then converted to areas and the standard errors of the estimates were calculated. In a similar fashion, changes were estimated by comparing the interpretations between the years at the point level. The long term goal is to scale this pilot project up to support greenhouse gas emissions accounting and provide data for landscape monitoring and restoration applications.

The information needs of this project included:

1) *Photo interpretation of high resolution imagery from multiple time steps;*
2) *Image quality had to allow for the photo identification of multiple land use types of interest outlined in the project classification scheme (covered in the next exercise);*
3) *In addition to distinguishing broad categories of land use such as forested lands, crop lands, settlement, etc., this project aimed to distinguish finer class subcategories such as broadleaf forest, conifer and salt forest (forest types);*
4) *Because the end goal was to generate statistically valid estimates of LULC through time, a robust sampling approach was required to establish photo monitoring plot locations.*

B. Ecuador Land Cover, Use Change Mapping: Project Summary

The overall objective of this project was to map LULC for two points in time as well as detect and map LULC change across Ecuador. This project used geospatial modeling and change detection methods to generate 2014 and 2016 LULC maps as well as a map of change between the two years. Photo interpretation in CEO was used to support this project at multiple steps in process; first, to collect model training data and then to collect map validation data. The end products of this project are multiple validated maps of Ecuador, estimates of LULC class areas, and associated error matrices to guide their use.

A primary concern for the project was imagery availability. Equatorial South America can be a very cloudy place, and at least two years of imagery were needed to validate the maps that the project aimed to produce. In many areas, multiple imagery acquisitions were necessary to obtain imagery of high enough quality for use in photo interpretation. That process involved purchasing imagery from a vendor and hosting it on a GeoServer in order to load it into CEO.

LULC model training data were collected using multiple interpreters working in CEO. This dataset involved generating a stratified random sample of over 8,000 plots in 32 strata. LULC classes were interpreted at points within plots and then summarized at the plot-level. These plot data were then intersected with multiple continuous remotely sensed image resources, and the resulting model dataset was used to fit models and create maps. CEO was then used to collect an independent dataset for the validation of these maps.

The information needs for this project were quite extensive and included:

1) *A comprehensive Ecuador LULC classification scheme of classes that could reliably be used to derive photo interpreted data from high resolution imagery;*
2) *Sufficient high resolution imagery to develop training data to create outputs and validation data to evaluate outputs;*
3) *A robust, continuous modeling dataset comprised of multiple preprocessed remotely sensed images;*
4) *Carefully planned sample designs to optimize the separate processes of model data collection as well as map validation data. Additionally knowledge and expertise related to modeling, sample design and Ecuadorian vegetation was required.*

C. El Salvador project: Key Points

1. <u>Project objectives</u>: To establish methods to characterize overall trends in land use in El Salvador.
2. <u>CEO objectives</u>: To identify land use in a systematic grid of plots across the country.
3. <u>Products</u>: Numerical estimates of land use proportions across the study area as well as changes through time – this is an example of an inventory product.
4. <u>Information needs</u>: Comprehensive land use classification system representing broad and finer scale land use categories that can be identified in high resolution imagery. Multiple time steps of high resolution imagery. A statistically robust sample design to allow the estimation of land use proportions and comparison through time.

D. Ecuador: Key Points

1. <u>Project objectives</u>: To model and map land cover, use and change across all of Ecuador.
2. <u>CEO objectives</u>: 1) To collect PI model training data and 2) to collect PI map validation data.
3. <u>Products</u>: Validated wall-to-wall LULC and change maps of Ecuador including: 1) 2014 and 2016 LULC maps; 2) a map of LULC change from 2014 to 2016; and 3) map validation products (error matrices) to inform the use of map products.
4. <u>Information needs</u>:
   - i. Comprehensive LULC classification capturing the wide range of vegetation and land use types across Ecuador. This had to include LULC classes that can reliably be identified from high resolution imagery.
   - ii. Two time steps (2014 and 2016) of high resolution imagery to allow for the identification of LULC as well as change.
   - iii. Modeling data including processed remotely sensed image data.
   - iv. Two carefully planned sample strategies: 1) to ensure that model training data capture the range of variability across Ecuador for model training and 2) a robust stratified sample to ensure validation of all the LULC classes of interest.

Part 2: CEO project planning: Knowledge check

A. Describe key differences between mapping and inventory projects?

B. Classify the following applications as mapping or inventory projects:

1. Using CEO to interpret forest/non-forest in a systematic grid of plots across a study area and using the data to generate an estimate of forest cover is an example of what type of project?
2. Interpreting vegetation type from high resolution imagery for a sample of plots stratified according to elevation, intersecting plot data with remotely sensed imagery to create a model and generate a map.
3. Interpreting forest/non-forest from high-resolution imagery in CEO in order to assess the accuracy of an existing forest map.

C. Select the general class pairs from the list below which are likely to be distinguishable in typical high resolution imagery (e.g., what you might see in Google Earth). We will cover this topic in more detail later.

1. Tree vs. herbaceous vegetation
2. Water vs. barren ground
3. Grass vs. pasture crop
4. One deciduous tree species vs. another deciduous tree species
5. Snow/glacier vs. beach
6. Housing structure vs. commercial structure
7. Shrub vs. grass
8. Primary growth tree vs. secondary growth tree

D. From the list of examples below, rank the following in order of complexity based on information needs:

1. Estimate the proportion of land in a small country that is developed/non-developed.
2. Create a continuous map of canopy cover for small region (e.g., less than 1 Landsat scene)
3. Estimate broad land use class proportions across a small region.
4. Create a map representing vegetation type change between two time periods.

E. Answers

A. Inventory projects use statistical sampling to estimate proportions of LULC across an entire study area. Mapping projects generate wall-to-wall maps which are spatially explicit estimates of LULC.

B. Classify projects

   i. Inventory

   ii. Mapping

   iii. Mapping

C. Likely distinguishable class pairs

   i. Tree vs. herbaceous vegetation

   ii. Water vs. barren ground

   iii. Snow/glacier vs. beach

iv. Shrub vs. grass

D. Rank in order of complexity (least to most)

i. Estimate the proportion of land in a small country that is developed/non-developed.

ii. Estimate broad land-use class proportions across a small region.

iii. Create a continuous map of canopy cover for small region (e.g., less than 1 Landsat scene)

iv. Create a map representing vegetation type change between two time periods.

Part 3: Plan your own CEO project

A. Briefly describe and summarize your project.

1. Summarize and specify whether the work is an inventory or mapping project.

B. List the primary project objective and describe how CEO will support it.

1. What is the primary project goal?
2. What are the specific goals for using CEO?
3. What constraints exist affecting this project?

C. What needs to be measured, reported and produced by this project?

1. What basic landscape elements do you need to identify from imagery in CEO? This will determine your classification scheme as well as the imagery you select for your project.
2. Is it likely that these can be reliably identified using image interpretation?
3. What are the desired end products? Do you need maps, or will numerical summaries be sufficient? This will greatly affect the scope of your project, expertise and resources required as well as sampling strategy.

D. Outline the basic classification scheme

1. Make a general list of the basic classes that need to be identified from imagery in this project. Keep this very general as we will return to develop this in the next exercise.

E. Outline basic data needs

1. What kind of imagery will you likely need?

1. What time scales are required to address the project needs and why?

2. What spectral scale is necessary in the imagery to allow photo interpretation of the classes?

F. Optional – Identify data that may already available to support the project?

1. In addition to publically available datasets in CEO, is there any additional imagery available for your project?
2. Is there any available reference data to help train interpreters and develop the interpretation key?

G. Summarize all the information you have put together

1. Refer to section 1 Appendix 1. CEO Project Planning Worksheet as you will likely want to complete this after you have reviewed the rest of the exercises.

2. This information will be used in the ensuing exercises and provide the basis for setting up and completing your CEO project. If you are unable to answer the questions above and/or do not have the information to complete the first section CEO Project Planning Worksheet, you are likely not ready to move forward with you project.

3. The Basics of Land Use and Land Cover

Once the objectives and information needs of your project have been defined, the next step is to produce a set of clear definitions for the landscape elements you wish to identify in your project. These definitions must be in line with your project objectives. For example, if your goal is to monitor forest recovery, you'd want to define a set of LULC (discussed below) definitions that clearly identify landscape-level forest change and recovery. This ensures that you can gather data which are relevant to the monitoring effort. If instead you wanted to use CEO to monitor changes in what crops are being grown in a region, you'd want a classification schema that differentiates different types of crops.

The classification scheme you create will provide the framework for your image interpretation. Once you settle on clear classification definitions, the next step is to produce a formal image interpretation key to support your work. Each of these steps is described in the following sections.

3.1 Land Cover vs. Land Use

When developing a classification, it is vital to keep in mind the difference between *land cover* and *land use*. Land cover describes the biophysical elements of the landscape, whether they are natural or man-made: trees, patches of grass, fields of crops, buildings, roads, ponds, rivers, and all the other physical objects that we see when we look at the landscape. In contrast, land use refers to the way that people make use of the landscape: agriculture, settlement, or park are examples of land uses. Put simply, land cover describes what elements make up the landscape, while land use describes how those elements are organized in terms of human systems[5,6]. For example, while a feature on the landscape appears to physically be herbaceous vegetation, it may be *used* as a pasture. It can be quite difficult to mentally separate the two, as we tend to frame the landscape in terms of how we interact with it. This can lead to some confusion, such as having a classification system that is a mixture of land use and land cover – if both herbaceous vegetation and pasture existed in a classification scheme used to classify the feature referenced above, this area could theoretically fall into both classes even though it can only be classified as one in reality, causing inconsistencies and confusion in your final map product or inventory. It's important to keep the two separate as much as possible so that you can address what is on the landscape separately from how those elements are organized in human terms.

3.2 Classification Systems

A LULC cover classification system needs to meet certain criteria in order to be useful and properly defined. The most important criteria are that the classification be exhaustive and exclusive; that is, all areas in the study area can be assigned to a category (exhaustive) and can only be placed in one category (exclusive). If a classification is not exhaustive, then some areas will be left unclassified at the end. This will create problems when it comes to estimating error and uncertainty. If the classification is not exclusive, then it may be ambiguous what category an area should be placed in, and this becomes a matter of individual interpretation.

[5] FAO/UNEP, 1999: Terminology for Integrated Resources Planning and Management. Food and Agriculture Organization/United Nations Environmental Programme, Rome, Italy and Nairobi, Kenya.

[6] FAO, 1997a: State of the World's Forests. Food and Agriculture Organization, Rome, Italy, 200 pp.

Classes should also be defined so that they are consistent with the purpose of the work. If forest extent is being studied, then it may not be necessary to have classes for different types of agriculture, but recording the difference between different forest types may be very important.

Class Definitions

To create a classification that meets the needs of your project, first consider what LULC types are essential to identify in order to achieve the goals of the project. For some projects, creating a simple classification that has a small number of classes is sufficient. For example, when mapping forest extent in a temperate country, appropriate classes might be Woodland, Deciduous Forest, Evergreen Forest, and Non-forest. Non-forest would contain all other LULC types.

However, if your project is part of a larger system, or might be used for more than one purpose, defining a larger set of classes may be the more appropriate choice.

Classifications can have a single level or be hierarchical with multiple levels. In a hierarchical classification the top level includes a set of broad, inclusive classes, each of which contain sub-classes with more specific descriptions (see table 1). The International Panel on Climate Change (IPCC) six class land use system may be the most widely used example of a hierarchical classification. It uses six land use categories (Forest Land, Cropland, Grassland, Wetlands, Settlements, and Other Land), inside of which a country may define multiple levels of more specific classes, tailored to that country. Readers are encouraged to consult the *Good Practice Guidance for Land Use, Land-Use Change and Forestry* and the *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol 4* for an in-depth discussion of this system.

Note also that your classification does not need to be of uniform detail. If it is being developed to support a National Forest Monitoring System, and intended to be compatible with UN reporting, it might use several levels of classes inside of the Forest Land class, but keep the other five land use classes defined only at the highest level.

Table 1. Example of using a hierarchical classification, based on the IPCC six class system, for a project focused on defining forest types. Only Forest Land has second level classes.

Level 1	Level 2	Definition
Forest Land	Deciduous Forest	Areas with >= 30% tree cover and tress > 3 m in height, with > 50% of tree cover deciduous.
	Evergreen Forest	Areas with >= 30% tree cover and tress > 3 m in height, with > 50% of tree cover evergreen.
	Woodlands	Areas with >= 10% but < 30% tree cover, tree cover may be deciduous or evergreen.
Grassland		Non-range or pasture areas, may include non-grass vegetation such as herbs and brushes. < 10% tree canopy cover.
Cropland		Area > 50% covered with agricultural land, including rice fields.
Wetlands		Land that is > 50% covered or saturated by water for all or part of the year (e.g., peatlands) and that does not fall into the Forest Land, Cropland, Grassland or Settlements categories.
Settlement		Areas with >= 30% developed land, including transportation infrastructure and human settlements.
Other Lands		Bare soil, rock, ice, and all land areas that do not fall into any of the other five categories.

Class definitions should be sufficiently descriptive and detailed that they could be used by people not affiliated with your project to produce data that would match your own. That is to say, they should be detailed enough to support consistent and accurate interpretation of elements on the landscape.

Consider the examples below:

> *"Forest" is defined as land covered in large trees.*
>
> *"Forest" is defined as continuous land area of at least one hectare, with more than 30% canopy cover, with trees more than 5 meters in height. Trees may be deciduous or evergreen. Trees planted by humans for forestry or agricultural purposes are included. Areas which may be seasonally or permanently inundated with water are not included.*

The first definition is vague and unspecific, and two people could disagree as to whether or not a particular area should be classified as forest. The second provides clear, measurable thresholds that define what should be classified as a forest and what should not. Classes should be defined as clearly and as specifically as possible. Attempting to apply these class definitions helps to find places where they might fail. For example, when using the more specific definition of forest above, we might run into problems with mangroves. Because they are inundated with water, they would not be included as forest. If it's important to separate flooded forests from non-flooded forests a "Mangrove" or "Flooded Forest" class should be created. In a hierarchical classification system, each new level below the broadest top classes adds new details that are used to make the new classes distinct from each other. It is important to keep in mind that some classes may be difficult to separate and distinguish from one another during photo interpretation. This issue will be discussed further in section 4.2 Photo Interpretation; however, you can plan for and handle some potential class separability issues with a carefully crafted hierarchical classification scheme.

Classes: Points vs. Plot

It's important to note the difference between defining your classes for use with *points*, or with *plots*. Points correspond to individual landscape elements or features, whereas plots must refer to groups of landscape elements, most likely one that is not completely homogenous. For example, a single point on the ground can't logically be classified as closed or open forest. The necessary information for that distinction can't be contained within a single point location or landscape element, because it's inherently about how different components of the landscape relate to each other. That is, it's the balance of points classified as tree and not-tree within a plot that determines if the plot should be called closed or open forest.

CEO can be used for both of these approaches, as it can summarize the data collected from points within a plot into plot-level information. This also has advantages in terms of simplifying the statistics involved in working with your data, as will be discussed in the sample design section.

This point/plot distinction means that you need to consider what point classes can be created that will produce the desired plot classes. Mostly this will mean that your point classes may be simpler than your plot classes, and will be focused on individual landscape elements (tree, shrub, grass, bare soil, etc.), rather than an assemblage of them (forest, grassland, agricultural land, etc.). Then, using the summarized point data at the plot-level, a class can be assigned to the plot.

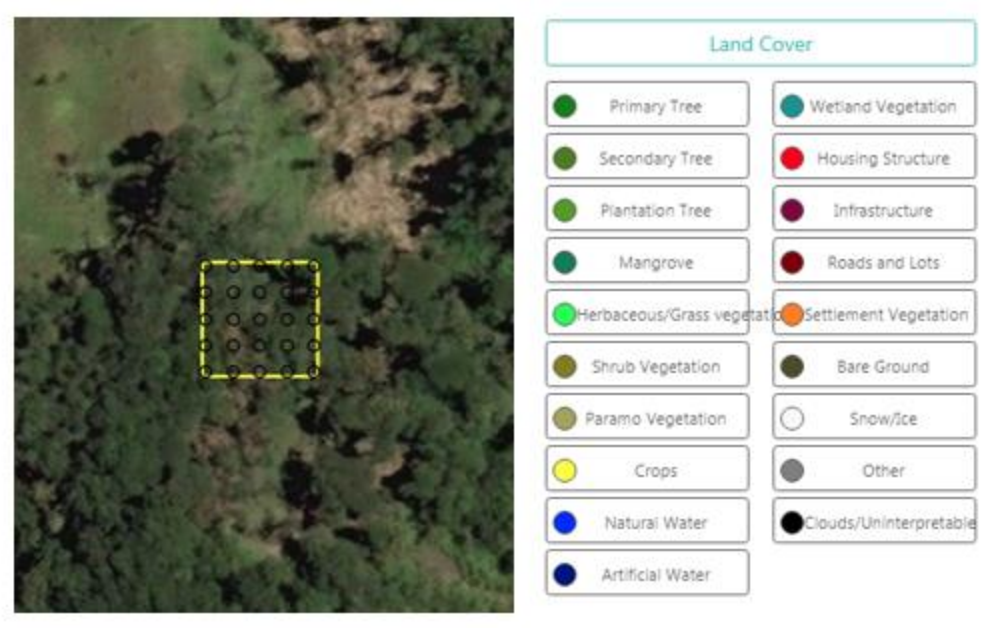

Figure 2. Example plot containing points to be interpreted with the available list of available classes. In this case, point-level classifications will be summarized at the plot-level for analysis.

It's generally a good idea to think about your plot-level classes first, and then define the point classes that you would need to identify in order to make sure a plot was described well enough to be placed in one of your plot classes.

A Brief Introduction to LCCS

If you want to create a robust, extremely granular classification system that uses international standards and recognized best practices, there are tools available to help you undertake the task. The Land Cover Meta Language (LCML)[7] is a standard international classification system developed by the UN Food and Agriculture Organization (FAO). The LCML allows the user to create a standardized set of attributes that are used to describe land cover features. A particular hierarchical cluster, or grouping, of these attributes is then assigned to a land cover class. The top-level division in LCML is biotic (natural environment) and abiotic (human constructed), and each additional attribute branches underneath these two, as appropriate. A land cover class is defined by the attributes (and thus, landscape elements) it contains and how those attributes relate to each in other in time and space. The LCML is implemented in the Land Cover Classification System (LCCS) software, which is freely available from FAO[8]. Using LCML and LCCS to generate a classification system is a deep exercise and outside the scope of this manual, but readers are encouraged to look into these tools to see examples of how to construct robust and comprehensive LULC classification systems.

[7] http://www.fao.org/geospatial/resources/standards/en/
[8] http://www.fao.org/geospatial/resources/tools/land-cover-toolbox/en/

3.3 Interpretation Keys

Clear, precise class definitions also make it easier to produce a key. Keys are the collection of rules, imagery, and guidance used to classify land cover elements in earth observing imagery. Well made, comprehensive keys are the cornerstone of effective image interpretation work, and they allow for much more consistent and reliable results.

Keys most often include example images representing each class. These should be selected to show the variability of that class within the range of different areas the project will cover. These images may be annotated to highlight the class in question and may be created for the various types of imagery that the project will make use of (see Figure 3). Keys can also include written guidance explaining the possible difficulties an interpreter might encounter, detailing landscape elements that commonly occur in a given class, how the class may appear differently in false color imagery, and other helpful tips for separating closely-related classes.

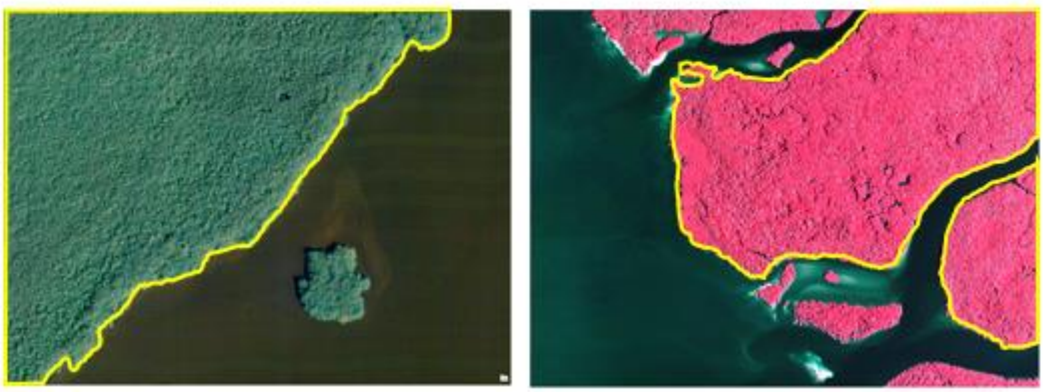

Figure 3. Example images from a photo key. These two images highlight the Mangrove class, outlined in yellow. The image on the left is a true color image, and the image on the right is a Color Infrared image (discussed below). In this example, Mangrove is defined using the following rules: includes Forested land restricted to coastal locations interfacing between Forest land and Sea. Typically on the landward fringe of forest communities but occasionally grows on lower elevation, more frequently flooded sites. It also readily colonizes disturbed sites, where it may form nearly pure stands. Must have 10% or greater canopy cover, must be 1.0 acre in size and 120 feet wide or greater. Images courtesy USFS.

3.4 Examples

In the previous section, we introduced two example projects: a LULC inventory and monitoring pilot project in El Salvador and a LULC and change mapping project in Ecuador. In both of these projects, a classification system was developed according to project goals. The pilot project in El Salvador focused on monitoring trends in land use through time and looked at proportions of land use classes related to carbon emissions and landscape restoration. The Ecuadorian mapping project focused on landscape and vegetation change. In both cases, hierarchical classification systems were developed according to the project objectives. These examples will be explored further in the following exercise.

Exercise 3.1: Develop a Classification Scheme and Interpretation Key

Introduction

In this exercise, we will start by reviewing the process of establishing classification systems and developing keys in example projects. You will then check your knowledge before you are ready to develop a classification scheme and interpretation key for the project that you began mapping out in Exercise 2.1.

Objectives

- Evaluate example classification systems
- Ensure that classification systems are exhaustive and exclusive
- Appreciate the importance of developing interpretation keys with example images

Part 1: Review example classification schemes

A. El Salvador

1. Project Objectives – Estimate proportions of land use types across the pilot study area and characterize trends through time.
2. Classes of interest – Land use classes related to greenhouse gas emissions and landscape restoration.
   i. These classes were defined using expert knowledge and organized in a hierarchical structure to accommodate flexibility in analysis as well potential difficulty in class separability.
3. Unit of assessment – This was an inventory project and photo interpretation was done at the point-level (each point representing an area within a plot). This topic and will be discussed further in chapter 5. Sample and Response Design.
4. Results – Plot-level data labeled according to the classification system detailed below for 2001, 2012 and 2018.

Table 2. Land use classification used in El Salvador's pilot study

Level I	**Level II**
Land Use Category	*Land Use Sub-category*
1. Forested lands	11. Broadleaf forest
	12. Conifer forest
	13. Salt forest
2. Cropland	21. Sugar cane
	22. Coffee
	23. Orchards
	24. Other crops (grains, vegetables, rice, flower crops and similar)
3. Pastures	31. Pasture
4. Wetlands	41. Water bodies (oceans, seas, lakes, lagoons and rivers)
	42. Aquaculture and mariculture ponds, including the salt plants
5. Settlements	51. Continuous urban zones and other artificial infrastructure
	52. Discontinuous urban zone
6. Other lands	61. Beaches, sandy areas and dunes
	62. Rock and lava
	63. Shrub vegetation (scrubland)
	64. Clouds and/or shadows
	65. Barren land and road/rail network
7. Other treed areas	71. Riparian trees
	72. Trees in agricultural or livestock areas
	73. Trees in continuous urban areas
	74. Trees in discontinuous urban areas

B. Ecuador

1. Project Objectives – Map land cover and land use across the country and look at changes. Focus on existing vegetation, land use and change related to forest conservation.
2. Classes of interest – Vegetation types and land uses representing the diversity of the Ecuadorian landscape and conservation concerns.
   i. These were selected using expert knowledge and organized in a hierarchical structure to accommodate flexibility in analysis as well potential difficulty in class separability.
3. Unit of classification – Square plots 900 m^2 in size, containing a grid of 25 points. Points on the landscape were first labeled using the point classification system described in the table below. Level II classifications were then summarized at the plot level to generate the classification described in the Level I system. This topic will be discussed further in chapter 5. Sample and Response Design.
4. Results - Class definitions and photo examples
   i. Plot level class definitions based on point classes

Table 3. Example classification used for point photo interpretation for land cover land use mapping in Ecuador.

Level I	Point Class	Description
Forest Lands	Primary Tree	A native tree, growing in an unmanaged location, more than 5 m tall.
	Secondary tree	A native tree, growing in an unmanaged location, more than 5 m tall, after a recent disturbance.
	Plantation Tree	A native or non-native tree, grown in a managed location, for economic production purposes.
	Mangrove Tree	Native woody plant growing in or near brackish water near the coast, or estuary environments, more than 2 m tall.
Shrub/Grassland	Herbaceous vegetation/Grass	Low lying, non-woody vegetation, including non-bamboo grasses and other types of herbaceous plant cover.
	Shrub vegetation	Woody vegetation less than 5 m in height.
	Paramo vegetation	Tropical high Andean vegetation characterized by dominant non-tree species that include fragments of native tree endemic to the area.
Cropland	Crops	Cultivated crop vegetation, whether row crop or other agricultural system.
Wetlands	Natural Water	Static or moving water, of natural origin. May be dry part of the year.
	Artificial Water	Static or moving water, body of water created by human intervention. May be dry part of the year.
	Wetland Vegetation	Submerged or emergent vegetation associated with inundated land.
Settlement	Housing Structure	A building primarily used for housing or commercial purposes.
	Infrastructure building	Structures used for industrial purposes, such as power stations or dams.
	Roads and Lots	Paved and unpaved roads and lots.
	Settlement vegetation	Vegetation in a settlement area (parks, yards, etc.).
Other Lands	Bare Ground	Natural ground, not subject to seasonal inundation, with no vegetation present, could include sand bars, rocky slopes, mud flats, etc.
	Snow/Ice	Non-seasonal snow or ice.

5. Plot-level class definitions and photo key examples were created for each plot-level class. The photo key was developed using the same imagery that was used for the interpretations
6. Example of plot-level classification
   i. Primary Forest or Secondary forest - Plant community of at least one hectare, dominated by trees (>= 30% tree canopy cover), where majority of tree cover represented by trees > 5 m in height, characterized by the presence of native tree species (of different ages and sizes) with one or more stratum. Areas covered with bamboo and native palms are

included, as long as they reach the minimum limit established in terms of minimum area, height and canopy cover.

Figure 4. Examples of primary forest (left) and secondary forest (right), with the Collect Earth Online plots used in the project.

ii. Settlement - Areas principally occupied (> 50%) by homes and buildings for public use.

Figure 5. Examples of the settlement classes, with roads, structures, and urban vegetation. The right panel shows clearly the necessity of including an urban vegetation class, as this plot could otherwise be labelled as forest.

iii. Grassland/Herbland - Lands with herbaceous cover, derived from spontaneous growth, that don't receive special care, used for sporadic grazing, wildlife, or protection purposes. The aerial extent of trees and shrubs is each < 30%.

Figure 6. Examples of herbaceous and shrub cover types. The left panel shows grassland or herbaceous cover, while the right shows a mixture of shurbs and grass or herbaceous plants.

Part 2: Classification System & Key: Knowledge check

Many of the questions in this section will not yet have definitive answers at this stage. Do your best to answer the questions and if you can't, think about what information you are lacking. Because the process of planning and setting up a CEO project is often iterative, it is important to consider how decisions will play out and affect decisions and outcomes further on in the process. Frequently, we will have to return to previous topics to revise and refine.

A. Review the classification systems for the two examples and answer the following questions:

1. Do the classification systems reflect the project goals?
2. Do both classification systems appear to meet the essential criteria of being comprehensive and exclusive?
   - i. Comprehensive/exhaustive - Can you imagine a landscape feature in either of these regions that would not fit into classification systems? In order to meet this criterion, every feature should belong to one class.
   - ii. Exclusive – Can you imagine a landscape feature in either of these examples that could belong to more than one class? In order to meet this criterion, every class must be mutually exclusive.
3. Do you think that it is reasonable to expect the point classes to be distinguishable using photo interpretation? Why and why not?
   - i. It is likely that you won't be able to answer this question fully yet; however, it is important to start thinking about interpretability at this stage. We will learn much more about this in the next three exercises.

B. Examine the example plot class definitions and photo key images for the Ecuador project and answer the following:

1. Can you see how the images in each pair represent variability within plot classes based on point level labels?
2. Why or why not?

Part 3: Create a classification scheme for your project

This part is critical and must be done right so that the data collected meet the project needs. Additionally, it is important to take into consideration whether your CEO project will be shared with other projects and/or possibly be reused in future projects, as this may affect the classes that are defined. In general, the classification must be comprehensive, meaning that every point or plot can be assigned to a class. Additionally, the classes must be exclusive, meaning that each point or plot can only be assigned to one class.

A. Refer to the previous exercise and review the list of classes that you outlined.

1. Are the classes in this scheme exhaustive and exclusive? What additional classes need to be included to make it so?
2. Is there a need for hierarchical classes?

B. Define classes

1. For each class in your classification, write a clear definition for the class.
2. Review the list of class definitions.
3. Determine whether classes should interpreted at the plot-level or point-level (land use classes will often be at the plot level, but land cover classes could be at either).
   i. *Note-This and the following question will likely require reviewing imagery. You might want to return to these questions after you have completed the next two exercises in this manual.*

C. Find examples of classes in imagery for your key

i. Add additional information for interpretation such as additional tips (expert ecological knowledge as well as image interpretation tips). You may want to return to this question after reviewing imagery.

4. Image Interpretation

Remote sensing platforms currently collect an enormous amount of imagery data at various spectral and spatial resolutions and different temporal frequencies. Many datasets have become publicly available, and now remotely sensed data is more abundant and more easily accessible than ever before. In order to make use of available image data for image interpretation, we must have a good understanding of the basics of remote sensing. The first section below provides a brief introduction to the uses of remotely sensed imagery from earth-observing platforms. The following sections deal with the types of imagery, how to interpret it, and the final sections provide more details on how to rigorously design projects for two different approaches in CEO.

4.1 Uses of Imagery Interpretation

Remotely sensed imagery can be used to support inventory projects as well as the production of maps. For inventory type projects, image interpretation at sampled point locations can be used to generate estimates of LULC types. In mapping projects, it can be used to obtain datasets for calibrating models that predict the composition of the landscape. Similarly, image interpretation can be used to collect reference data that can be used to validate maps such as the outputs of those models.

Point sample-based inventories

Producing a map is not always the final product from remote sensing analysis. Producing maps often requires the use of statistical modeling or machine learning to make predictions about the landscape, and a substantial investment of time, resources, and expertise is necessary to ensure that the model and resulting map are reasonably accurate. If modelling is not used, manual classification and digitization of landscape features may be performed, which requires an even larger investment of effort. Many of the same processes can also be measured by using sample-based inventories and simply summarizing the results into a series of numbers. For instance, you can use a sample-based approach to estimate the coverage of tree canopy in a region or the change in impervious surface over time. Providing these numbers may be sufficient for your task. Not all questions about the landscape and how it is changing over time need to be addressed using maps.

Training Data

The primary drawback of a point-based estimate of landscape characteristics is that it does not provide a spatially explicit estimation. It's possible to estimate how much of a landscape characteristic exists, but it is impossible to determine where it is located inside the study region. To produce a spatially explicit estimate, people often use various types of statistical modeling or machine learning. These models need a sample of landscape data to be trained on, and image interpretation provides an easy way to collect an accurate and reliable set of training data.

For many methods, this training data does not need to be created using a statistically rigorous sampling process. Data can be selected purposively, targeting areas of importance or where there is likely to be confusion. This can make sample generation a very simple, iterative process. Other methods may require a statistically valid sample to train on, in which case a more rigorous approach to sampling will need to be implemented. While sample size determination will be discussed in section 5.3 Sample design strategies, machine learning and statistical modeling applications are beyond the scope of this manual.

Validation Data

The third (and most common) primary use for image interpretation is validation, or assessing accuracy of a map product. After a model is trained and tuned to be as accurate as possible, the outputs must be assessed for error and the level of uncertainty must be quantified. One of the standard methods for producing a reference data set is through the use of image interpretation. It is good practice to make use of imagery with higher spatial resolution for producing the reference data set, when possible[9]. It is also important to make use of good quality assurance and control methods during and after collection to ensure the production of unbiased and accurate reference data. These methods apply to nearly all image interpretation work, and will be discussed in chapter 6. Data Collection and Quality Assurance and Control.

4.2 Photo Interpretation

Overview

Image or photo Interpretation is defined as the "determination of the nature of objects on a photograph or image and the judgment of their significance".[10] Photo Interpretation (PI) is the central skill needed for successfully executing a project in CEO. Photo interpretation is a skill that is acquired through hours of practice with photos, ideally combined with ground visits to check the accuracy of interpretations. This work is not an exact science due to the complexities of the real world, but the human visual system (the combination of the detail the eye is capable of detecting, and the pattern matching systems of the brain) is incredibly powerful and efficient, and is still considered to be a more accurate image processing system than any computerized or automated systems. Human interpreters are capable of dealing with unusual data much more effectively than trained algorithms (computers are far more consistent, however). Visual interpretation is the standard for evaluating LULC classifications. Proper validation of a classification depends on accurate and consistent human inputs.

Photo interpretation work relies on a properly constructed project framework. The project needs a clearly defined classification scheme, well-made keys that can be referenced by interpreters, a system of samples to interpret, and of course imagery.

Seven Characteristics of Images

When computers classify images, they make use of the brightness values of the bands in the data, or whatever other spectral information is available (normalized differences, radiometric statistics, etc.). Object-oriented classifiers can also use spatial or topological information such as shape and association. In contrast, humans use any combination of seven image characteristics for visual interpretation: **tone/color**, **shape**, **size**, **association**, **shadow**, **pattern**, and **texture**. This is a much more efficient use of the available information in an image.

The relative importance of each of the seven image characteristics varies with different imagery and subjects. The importance of any characteristic can depend significantly on the scale of the imagery and the properties of the features of interest. For example, the shape of a tree crown is important indicator

---

[9] Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015

[10] Paine D., 2012, Aerial Photography and Image Interpretation, John Wiley & Sons Ltd

of tree species in very high resolution (VHR) imagery, but in coarser resolution data, individual crowns may not be distinguishable. In other circumstances, such as a closed canopy forest, color or tone may be the most important indicator for separating features when other characteristics are similar or shared between the two features.

Tone is used to refer to the differences in brightness (relative lightness or darkness) in a black and white image, while *Color* refers to the different hues and shades of color that appear. Tone and color can be extremely useful for distinguishing between landscape objects. This effect can be enhanced by making use of false color imagery, where wavelengths of light the human eye cannot detect are assigned to a color. This is often done with near infrared data in what are called color infrared images, where near infrared is assigned the color red, while the red band is assigned to the color green, and the green band is assigned to the color blue (as seen in the image below). There are many possible combinations of bands that can be used to produce false color imagery, suitable to many different purposes.

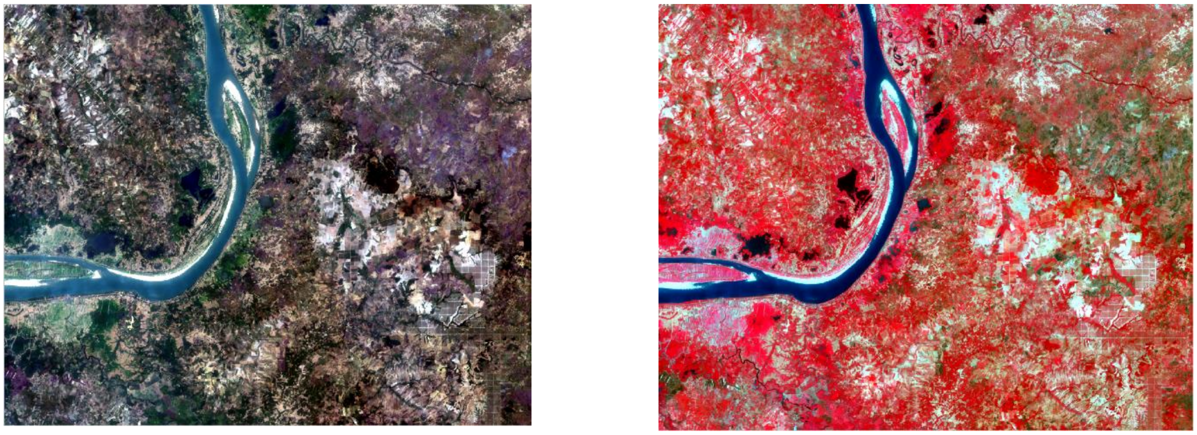

Figure 7. A pair of Landsat images, using natural color (left) and false color (right) representations. The false color image uses Near Infrared (NIR) light as the color red. Because live vegetation reflects NIR very strongly it shows up as bright red in the false color image, making it easier to separate from other landscape elements.

When developing your project, it is recommended that you investigate which types of false color imagery are most suitable for meeting the goals of your project.

The importance of *shape* varies with imagery resolution and the goals of the image interpretation. For instance, distinguishing characteristic tree crown shapes is impossible except with VHR, and thus if identifying individual trees by type is necessary, VHR imagery will be needed. However, shape is often very useful in identifying man-made objects even in moderate resolution imagery, due to the rectangular and linear forms infrastructure and buildings tend to display from above.

Size is usually evaluated by looking at objects that the interpreter is familiar with and comparing their relative size to less familiar objects. Relative size of objects is useful in helping distinguish between things with otherwise similar structures, such as trees and shrubs (as in the case of red and green circles, below, which illustrates the difference between the size of trees and shrubs). If the scale of an image is known, true size of objects in the image can also be estimated.

Figure 8. Aerial image demonstrating the usefulness of size (both relative and absolute) in distinguishing between objects on the landscape. The red circle encloses a tree, which has a crown larger than the vehicles in the image, which are of a size that will be familiar to most people viewing the image. The green circle, in contrast contains shrub cover. Without the presence of objects of known size in the image, it might be difficult to divide trees and shrubs in the image.

Pattern refers to the arrangement of individual objects into distinctive, recurring forms that permit recognition. Note that different patterns may become more or less apparent at different viewing scales.

For example, the checkerboard pattern of rectangular shapes often occurs in forest areas with a history of clear-cutting along regular boundaries. In general, man-made patterns can be easily distinguished in at many different resolutions. Natural patterns that occur within landscapes can also provide information that will aid the interpreter's decision.

Figure 9. An example of how pattern can provide information to the interpreter. The hard edges within this forest suggest a history of clear cutting trees.

Texture refers to the apparent roughness and smoothness of an imaged area and it is created by the frequency of tonal change. Texture can offer the interpreter clues about density, age, and type of vegetation present. Texture becomes increasingly more important as the scale of the imagery becomes smaller, and individual objects are harder to discern. To properly interpret texture, some knowledge of how individual landscape elements appear when they aggregate is necessary.

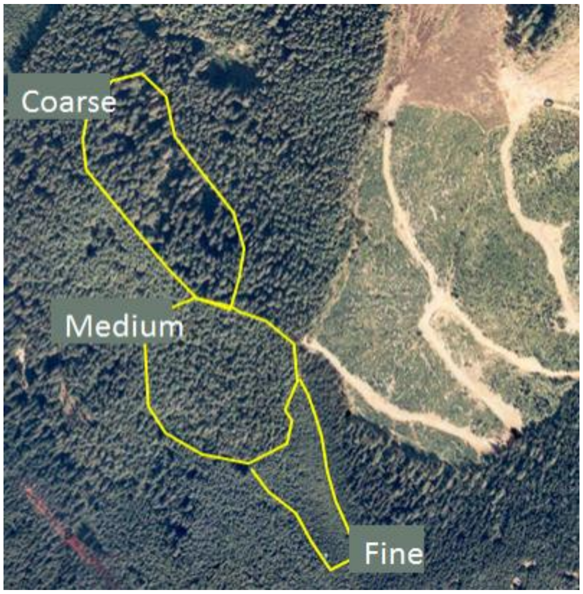

Figure 10. These differing forest stands show how texture can be useful in distinguishing between types of landscape. Here, differing sizes of trees produce very different visual textures, with large trees producing a coarse texture and smaller trees producing a finer texture.

Shadows can be inconvenient, as they obscure a portion of the landscape, and imagery is usually collected close to solar noon to minimize them. However, shadows cast by features on the landscape can provide information about the size and shape of those features that would otherwise be difficult to assess. Shadows are particularly useful for determining the relative heights of objects in the image, which can be vital in separating trees from shrubs. This is especially true when working with orthophotos, as the lack of radial displacement makes judging the height of objects extremely difficult.

Figure 11. This image demonstrates how shadows can be useful in determining the relative heights of objects. Taller trees in the image cast much longer shadows that trees of smaller stature, which would be difficult to otherwise distinguish.

Landscape objects exist in *association* with other objects. By observing objects in the image and observing others in the surrounding area, the interpreter can make inferences about what the objects really are. In many instances biological association is particularly important in making vegetation type maps. In other cases, the presence of a dam would suggest that a water body is a reservoir rather than a natural lake. Without that association, the water body would not be able to be further classified as a reservoir or artificial waterbody.

Figure 12. To determine whether this waterbody is natural or artificial requires either knowledge of the area, a catalog of artificial waterbodies that is known to be reliable and complete, or the use of other landscape features to infer the difference. The sharp, linear southern edge of the water strongly suggests the presence of a dam, meaning this would likely be a reservoir.

Convergence of evidence

When interpreting images, we often have a **convergence of evidence**, where multiple observations regarding an image combine to allow us to draw a conclusion about its contents. Convergence of evidence is usually a subconscious summation of all the interpretable image characteristics. It is convergence of interpretable image characteristics that makes image interpretation an inherently inductive logic (subjective) exercise that is difficult for a deductive logic (objective) algorithm to duplicate. This difficulty has made automation of the image interpretation process elusive for pattern recognition software, and it has remained a steadfastly human endeavor. While the results of image interpretation are often treated as if they are 100% accurate, this rarely the case[11]. It's therefore important to remember that there is a difference between interpretation and identification. Interpretation is about the judgment and experience of the interpreter and is a subjective decision. On the other hand, identification is an objective, accurate measure, a definitive answer.

Interpretation guidelines

Image interpretation is a skill that requires practice and upkeep to achieve consistent, accurate results. Ideally, interpreters regularly take prepared imagery into the field, with a classification, and visit targeted locations to classify them, comparing ground conditions with imagery. Similarly, features should be studied using imagery from multiple sources to understand how features on the landscape (and their attributes) translate into imagery at various scales.

One of the key challenges of image interpretation is to arrive at consistent results across multiple interpreters. It's important to be consistent across interpreters in order to reduce the variability in response and avoid biases introduced by differences of opinion between individuals reviewing the same imagery. There are a number of useful tools that help reduce this effect. The first, creating a good **interpretation key**, was discussed above in section 3.3 Interpretation Keys. A comprehensive and accurate key provides a foundation for agreement between interpreters. A second useful component is **ancillary information**, or data other than the primary imagery used for the interpretation. This could be other sets of imagery, topographic data (elevation, aspect, and slope), vegetation maps, cadastral data, or virtually any other source of data that could help to interpret the images for the desired purpose. It is possible to add many types of imagery and data objects to a CEO project, allowing for viewing the same scene with multiple false color composites, vegetation maps, road data, etc.

The final component to arriving at consistent and reliable image interpretation results is using training, cross validation, and other forms of Quality Assurance and Quality Control to actively verify that interpreters are producing similar results. It is possible to measure and describe differences in interpretation results using a variety of tools. These methods will be introduced and discussed in chapter *6. Data Collection and Quality Assurance and Control*.

[11] Mann, S., & Rothley, K. D. (2006). Sensitivity of Landsat/IKONOS accuracy comparison to errors in photointerpreted reference data and variations in test point sets. International Journal of Remote Sensing, 27(22), 5027–5036. https://doi.org/10.1080/01431160600784291

Exercise 4.1: Image Interpretation Basics

Introduction

The focus of this exercise is to learn and explore the basics principles of image interpretation. We will explore a collection of example images using the classification system and response design developed for the Ecuador LULC and change mapping project. We will see how the essential image elements of tone, shape, size, pattern, texture, shadow and association can be used to identify different landscape classes. More information on image interpretation and image selection will come in later exercises.

Objectives

- Learn how to use seven essential visual image characteristics to interpret imagery
- Gain experience with very basic image interpretation

Part 1: Understand Seven Image Characteristics

A. Tone and color

In the three example images below, tone (in the black and white image on the left) allows easy distinction between roads, forests, water and bare-ground. Color (in the middle image) allows easy distinction between coniferous trees, deciduous trees, senesced grasses and road surface types. Color (right) in this case, color infrared, allows for added distinction between coniferous and deciduous forest, as well as between dead/dying and healthy forest. The brighter red vegetation is indicative of deciduous and the darker (finer textured, too) red vegetation is indicative of conifers.

Figure 13. Three different examples of the effectiveness of color and tone in helping to distinguish between landcover elements.

Figure 14. A pair of Landsat images, using natural color (left) and false color (right) representations. The false color image uses Near Infrared (NIR) light as the color red. Because live vegetation reflects NIR very strongly it shows up as bright red in the false color image, making it easier to separate from other landscape elements.

B. Shape

1. The graphic and image below demonstrate the importance of shape.

Figure 15. Shape examples, graphical examples from a key (left) and a stereo image (right). The crown shape can indicate tree type (deciduous/conifer) or even species of tree.

C. Size

1. Size is usually evaluated by looking at objects that the interpreter is familiar with and comparing the relative size to less familiar objects. At first glance the vegetation in the lower part of the image below may look like trees, but in comparison to the size of the vehicles, the vegetation is likely shrubs.

Figure 16. Aerial image demonstrating the usefulness of size (both relative and absolute) in distinguishing between objects on the landscape. The red circle encloses a tree, which has a crown larger than the vehicles in the image, which are of a size that will be familiar to most people viewing the image. The green circle, in contrast contains shrub cover. Without the presence of objects of known size in the image, it might be difficult to divide trees and shrubs in the image.

D. Texture

1. Texture is the apparent roughness or smoothness of an image region created by the tonal changes of the image. Texture becomes increasingly more important as the scale of the imagery becomes finer.

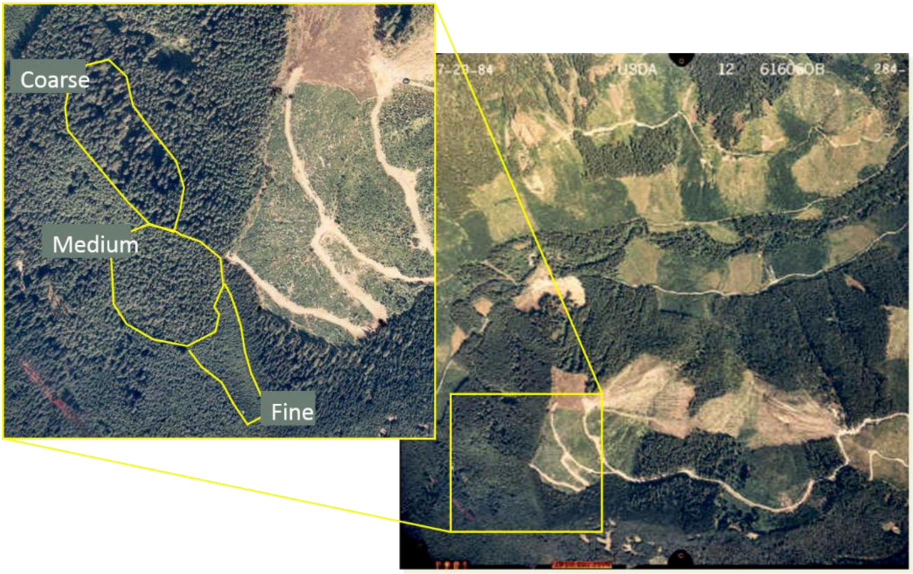

Figure 17. These differing forest stands show how texture can be useful in distinguishing between types of landscape. Here, differing sizes of trees produce very different visual textures, with large trees producing a coarse texture and smaller trees producing a finer texture.

E. Pattern

1. Pattern (in image interpretation) is a recognizable repetition of particular shapes. In this example, the checkerboard pattern of rectangular shapes suggests that this once-forested area has been harvested along regular boundaries.
   i. In general, man-made patterns can be easily distinguished. Natural patterns can also lead to information that will aid the interpreter's decision.

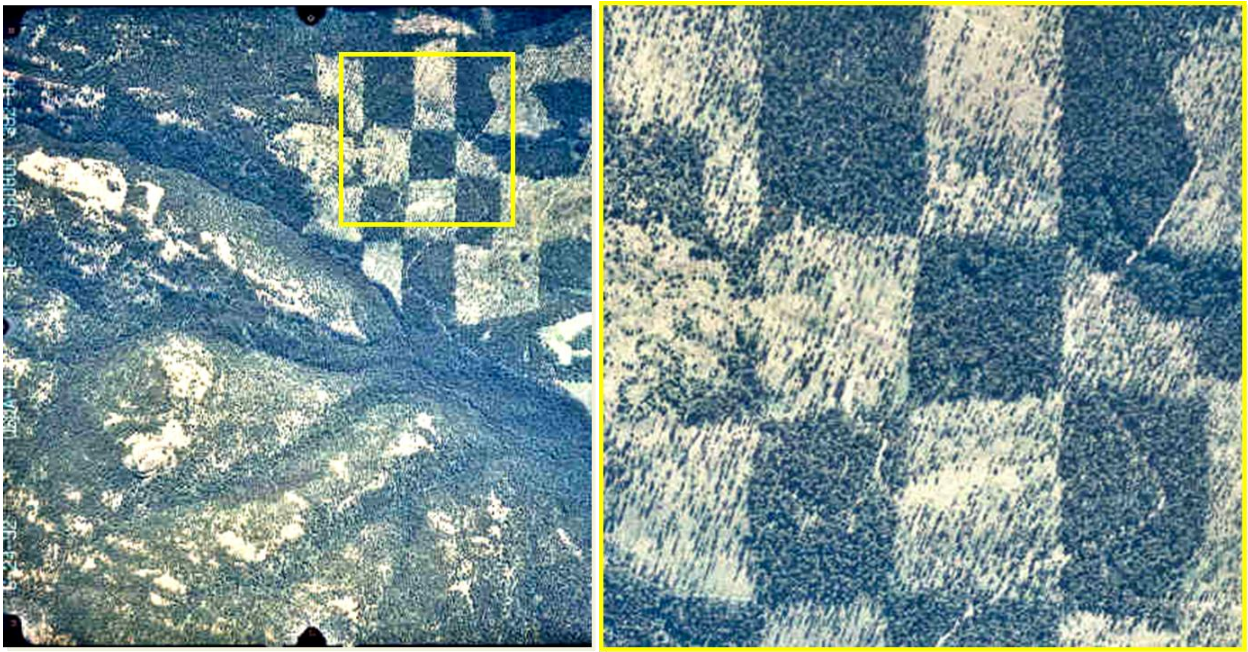

Figure 18. An example of how pattern can provide information to the interpreter. The hard edges within this forest suggest a history of harvesting trees.

F. Shadow

1. Shadows cast by objects in the image can give the interpreter information about the shape and size of certain features in the image.

Figure 19. This image demonstrates how shadows can be useful in determining the relative heights of objects. Taller trees in the image cast much longer shadows than trees of smaller stature, which would be difficult to otherwise distinguish.

G. Association

1. By observing objects in the image and observing the objects in the surrounding area, the interpreter can make inferences about what the objects really are.
   i. In the following example, the presence of a dam suggests that this water body is a reservoir. Without that association, the water body cannot be further classified as a reservoir.

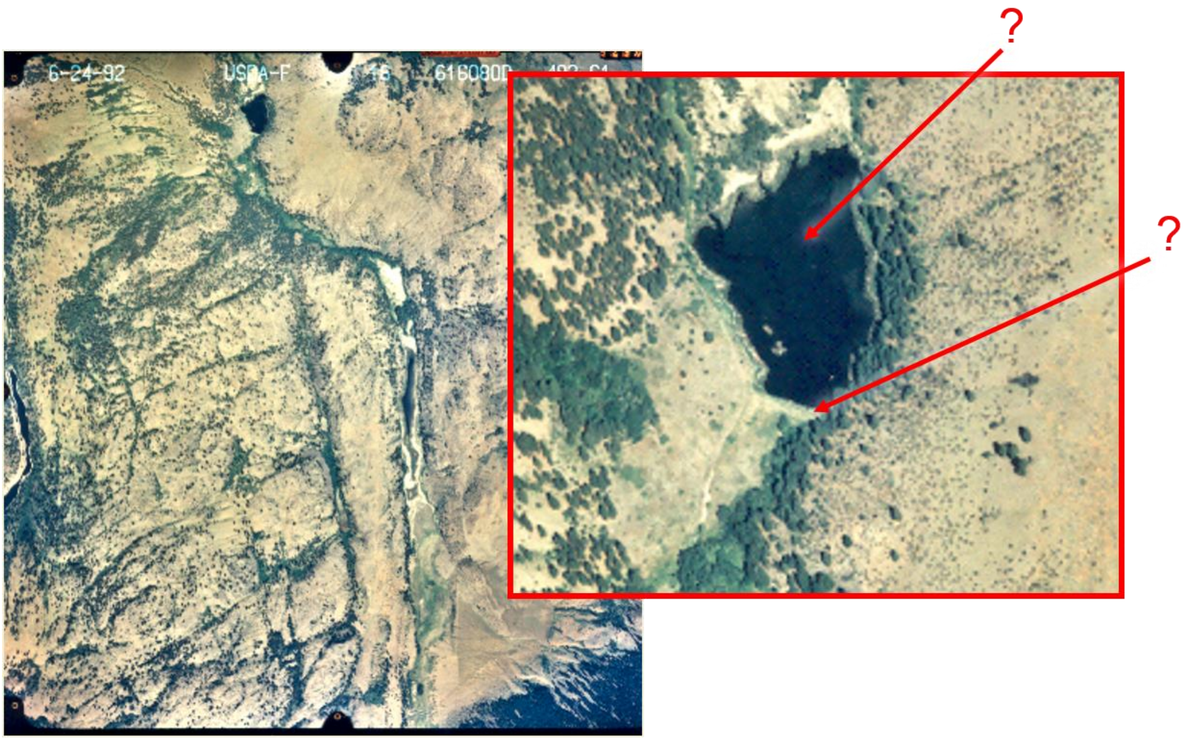

Figure 20. To determine whether this waterbody is natural or artificial requires either knowledge of the area, a catalog of artificial waterbodies that is known to be reliable and complete, or the use of other landscape features to infer the difference. The sharp, linear southern edge of the water strongly suggests the presence of a dam, meaning this would likely be a reservoir.

H. Convergence of Evidence

1. Bringing together of several types of information so that a conclusion may be drawn. This uses inference from all of the image elements.

Part 2: Image Characteristics – Knowledge Check

A. Observe the example image below and select the best class for each point labeled in the image

1. Choose from the classes in the table below
2. For each point, ask yourself how each of the visual elements supports your interpretation (tone, shape, size, pattern, texture, shadow and association

Figure 21. Image interpretation example. What classes should be applied to the four labelled circles?

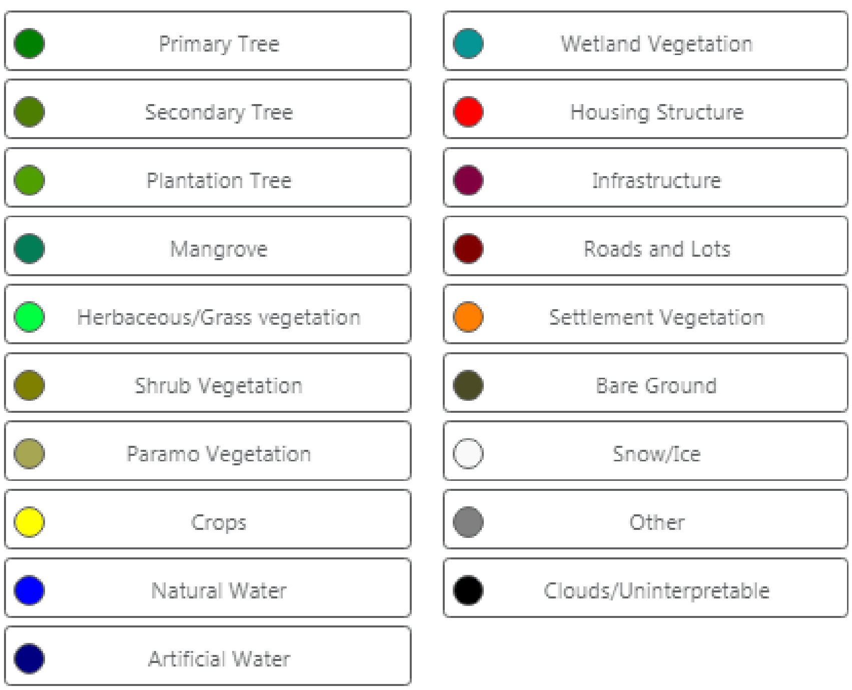

Figure 22. Land cover element classes to use for labeling the above image.

4.3 Overview of Imagery for Photo Interpretation

CEO is primarily used to view remotely sensed imagery of various sorts. It could be collected by a low earth orbit satellite, a manned airplane, or an unmanned aerial vehicle (or drone). Remotely sensed imagery is often based around photography in the visual spectrum, but can include other wavelengths of light, particularly those in the infrared portion of the electromagnetic spectrum.

When discussing aerial or space-based imagery, many unfamiliar terms may be used. We will briefly discuss some of the most important of these terms, though each is a substantial area of study on its own:

- Focal length
- Viewing angle
- Footprint
- Scale
- Resolution
- Perspective

Focal length describes the internal relationship of the lens elements of a camera. Each lens will have a distance at which light passing through the lens will be refocused into a sharp image. This distance will correspond to the distance from the center of the lens to the plane of the sensor (or film) in the camera. Focal length is related to field of view, viewing angle, and magnification. Long focal lengths lead to narrow fields of view and larger magnification. Small focal lengths lead to wide fields of view, and lower magnification. Focal length can also generate artifacts and distortions in imagery.

Viewing angle describes the total angle of incoming light (or field of view) that is captured by an imaging device, often a lens. It depends on the combination of film or sensor size and lens used for collecting an image and functions very similar to regular handheld photography. As just discussed, short focal length lenses produce wider viewing angles, long focal length lenses produce narrower viewing angles. However, the same lens on a smaller sensor will produce a narrower field of view than on a larger sensor, due to cropping of the image. This effect is seen often on digital cameras with interchangeable lens, where smaller sensor cameras increase the effective focal length of a lens.

Footprint describes the area of ground covered by the lens and captured by the film or sensor in the camera. The footprint of an image depends on the focal length of the lens, the size of the sensor, and the elevation of the camera. Changing any one of these elements will alter the footprint of an image. Higher elevations capture more area, increasing footprint. Longer focal lengths decrease angle of view, and thus reduce footprint. Larger sensors increase footprint.

Scale also describes the area of ground covered by the lens and captured by the film or sensor in the camera, similar to the footprint. This idea relates directly to the idea of map scale, as it is the ratio of a distance on the image to the corresponding distance on the ground. Scale also depends on the focal length of the lens, the size of the sensor, and the elevation of the camera. Scale is impacted by the topography of the ground beneath the camera; higher elevations will have a smaller scale than lower elevations, which means that an image that contains large changes in elevation and that has not been corrected for topography will have an inconsistent scale. Two images may have the same footprint, but different scales, if they were captured with different equipment. Typically, larger scale images will have smaller footprints, and smaller scale images will have larger footprints.

Resolution refers to a number of possible image characteristics. Generally the resolution of an image is taken to refer to the size of the smallest element, typically a single pixel, as when we say that Landsat images have a spatial resolution of 30 m (this corresponds to the length/width of the pixel). However, this is only one of several different types of resolution associated with aerial imagery. The types of resolution are: Spatial, Spectral, Radiometric, and Temporal.

Spatial resolution is the size of cells or pixels in the image, and determines how much detail the image contains and hints at what type of objects may be identified from it. It’s usually a good idea to select data with a spatial resolution that is somewhat smaller than the size of landscape elements you want to be able to identify.

Figure 23. An example of how spatial resolution impacts what landscape objects can be resolved in an image. The 5 m data would be sufficient for identifying stands of trees, the 1 m data would allow for identifying individual trees, and the 10 cm data makes discerning the difference between trees and shrubs easier.

Spectral resolution describes the size, position, and number of the imaging bands being recorded. By size, we mean the width of the electromagnetic spectrum being captured. Position refers to where on the electromagnetic spectrum the band is located. Most point and shoot digital cameras only record three bands of visible light (red, green, and blue, also known as RGB). Many digital sensors can capture more than the visible bands. It is very common to see the NIR band collected with the RGB bands in aerial image acquisitions. Satellite imagery usually contains even more than these four bands, which can be very useful in mapping vegetation types and conditions, identifying clouds and cloud shadows, and many other tasks.

Radiometric resolution refers to the number of values that can be recorded in each band of data; a higher radiometric resolution means more gradations of intensity can be captured by the sensor. Common values are 8-bit (2^8 = 256) and 16-bit (2^{16} = 65,536). The more gradations are available in the data, the more likely the data will be useful for detecting differences in the landscape. However, larger radiometric resolution also means much larger data sizes.

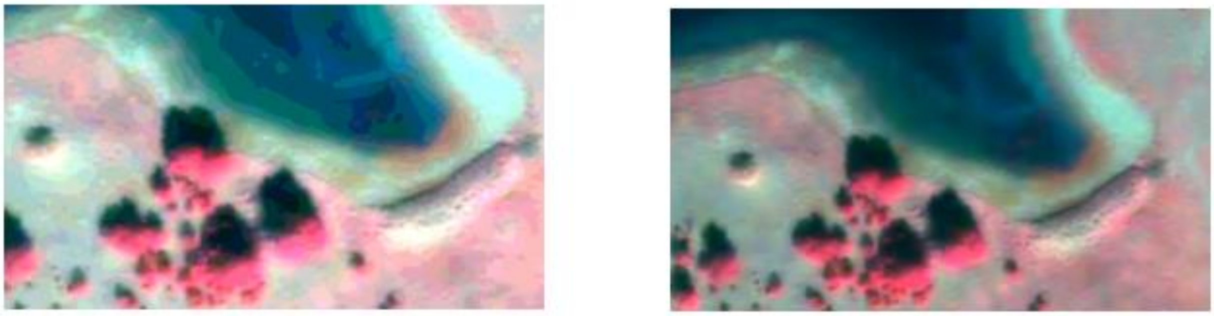

Figure 24. Two false color images showing the differences introduced into images by differing radiometric resolutions. The image on the left is a 6-bit image (64 values available), while the image on the left is an 8-bit image (256 values available). Note the artificial blocks of single values in the image on the left – these are introduced by forcing the complex gradations present in reality into a small number of representative values.

Temporal resolution describes how often the same area of ground is imaged. Many satellite systems have fixed intervals of recurrence. For instance, Landsat recaptures the same area every 16 days. Imagery flown for special purposes may have a variable temporal resolution. Generally, a finer temporal resolution is desirable, as it allows for capturing faster changes occurring on the landscape.

Perspective defines the position of the camera relative to the ground. Most aerial imagery is considered vertical, that is the camera is < 3 degrees from vertical, relative to the ground. However, that doesn't mean an aerial photo is like a map. Unaltered aerial photos have a single-point perspective and are not orthographic. This means only objects directly under the camera are seen directly from above, and features further away from that appear at an angle – this phenomenon is known as radial displacement. As the elevation of the camera increases, this effect becomes less noticeable. The existence of perspective in images can be both a good and bad thing. Single-point perspective allows for the generation of stereo imagery, if images overlap properly. This greatly enhances our ability to correctly interpret features within the imagery, and it allows us to accurately measure heights.

However, single-point perspective imagery is not planimetrically correct (in other words, features on the image are not in their true map location)—to make it planimetrically correct, we have to orthorectify or orthocorrect the imagery. This is necessary for most GIS applications.

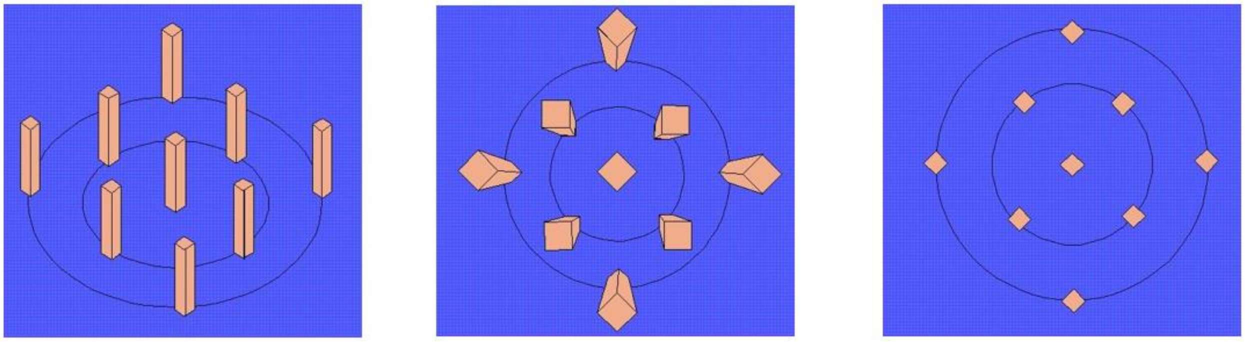

Figure 25. Three images showing the effects of perspective on the same scene. The left image shows an oblique image, that is, it is being viewed from the side. The center image shows the effects of single-point perspective. The right image shows an orthographic view of the objects, as if taken from an infinite distance away. Note the radial displacement of the columns, and their uneven apparent sizes in the center image.

To orthocorrect aerial imagery requires that points on the ground be captured from multiple angles, typically from multiple, overlapping passes of image acquisition. These overlapping images are combined with a digital elevation model (DEM) to generate orthocorrected or orthorectified images. Put simply, the overlapping images allow for correction of radial displacement, and the DEM is used to correct for elevation effects. The output images then have uniform scale and no radial displacement. In true orthophotos all objects appear as if viewed from directly overhead. Popular web mapping platforms like Google or Bing Maps use orthophotos for most scales of imagery displayed.

(Very) High Resolution Satellites Imagery and Aerial Photography

A large variety of high and very high resolution satellite and aerial imagery is now available, with spatial resolutions down to 30-50 centimeters publicly available. Currently available data include that from the Pléiades, IKONOS, QuickBird, WorldView series, KOMPSAT series, TripleSat, SPOT, GeoEye-1, and other satellites. However, most of this data is not publicly available, and must be purchased for the time and area of interest, and in many cases the historical record for a location of interest may be fragmented or non-existent. In addition, high resolution imagery tends to have more limited spectral resolution than other satellite imagery with fewer bands available, typically visible red (R), green (G), blue (B), and near infrared (IR or NIR).

Many countries also regularly collect aerial imagery (collected with airplanes, helicopters, or UAVs) through their agricultural programs. Aerial imagery can have even higher resolutions, down to the 2.5 centimeter range, an entire order of magnitude finer than satellite imagery. In the United States, this is the National Agricultural Imagery Program (NAIP). NAIP data is acquired roughly every three years at the state level, with a spatial resolution of 1 meter. Historically most NAIP data was acquired as RGB, but some states have opted to collect 4-band data, including the near infrared band. Web Mapping Services are available for much of this data. A wide variety of other aerial imagery is also available, as collected for other projects over time.

A number of VHR imagery services may be available in CEO free of charge, and it is advisable to investigate them and determine if they meet your project's needs. If you find the available imagery in CEO is not sufficient for your project, you will need to locate additional data to use. Determining what aerial imagery is available in your project area can be a time-consuming task. Don't underestimate the effort that might be required.

VHR imagery makes identification of smaller objects on the landscape much easier, and can allow for finer separation of LULC classes. Depending on the purpose of the project, it may be necessary to use VHR imagery to conduct reliable and accurate photo interpretation (see the previous section on image interpretation). However, relatively simple or coarse LULC classes may be readily interpreted from much coarser resolution data. In some cases the additional bands and consistent historical records of a moderate resolution sensor—as is present on the Landsat satellites—are more useful than higher spatial resolution data.

Moderate Resolution Imagery

A number of very important and widely used moderate and low resolution satellites also exist. The most important are likely MODIS (which produces both low and moderate resolution imagery), the new Sentinel satellites (which include Synthetic Aperture Radar), and the most widely used source of satellite imagery, Landsat.

MODIS (the Moderate Resolution Imaging Spectroradiometer) is one of a large suite of sensors aboard the Terra and Aqua satellites, and provides coverage of the entire earth every one-to-two days. MODIS produces a large variety of data products at varying resolutions and different time scales. Those likely to be the most important for use in CEO are the vegetation indices, though others might be useful for very specific purposes. A wealth of details is available via the National Aeronautics and Space Administration (NASA) website for the MODIS program[12]. The historical record for the MODIS mission dates back to 2000.

The European Space Agency (ESA) has launched a series of Sentinel satellites, each with an individual mission. The two of primary interest for CEO users are likely Sentinel-1 and Sentinel-2. Sentinel-1 is a Synthetic Aperture Rader (SAR) platform. SAR is useful for detecting changes in LULC under persistent cloud cover. Sentinel-2 is a multispectral earth-observation platform that captures 13 bands of data at varying spatial resolutions (10, 20, and 60 meters), and a fast revisit interval of five days. The complete Sentinel-2 is available from June, 2015. Sentinel-1 SAR data is available from October, 2014. Additional information and tools for working with Sentinel data can be found on the ESA website[13].

The most widely used earth-observing satellite data is produced by the Landsat series. Landsat data is collected in a variety of bands, with the most recent sensor (Landsat 8) collecting 11 bands, ranging from the deep blue to thermal infrared. Seven bands are collected at a spatial resolution of 30 meters. A panchromatic (all visible light) band is collected at 15 meters, as is a band for detecting cirrus clouds. The two thermal bands are collected at 100 meters and resampled to 30. The full Landsat record goes back to 1972, however, significant changes in sensors make comparison for data before 1982 difficult. Landsat 4 archives begin in August of 1982 and are largely compatible with current data. The Landsat 5 satellite collected data continuously from 1984 to 2012, giving it the longest operational lifetime of any earth-observing satellite so far. Early Landsat data coverage may not be complete for all areas, however, as images were only retained for areas of active study during part of the program's history. Landsat data has a revisit interval of 16 days. The long historical record, frequent revisit time and moderate spatial resolution of Landsat data have made it very widely used in remote sensing applications generally, and in land cover classification and change detection in particular. More information about the Landsat program can be found the USGS Landsat page[14].

Sourcing Imagery

Part of designing a project using CEO will be determining what imagery you need to meet the project's goals. Depending on what is being measured or the type of interpretation being performed, various types of data may be needed. While more bands of data are very useful, unless the project task requires the use of near or shortwave Infrared, visible light (RGB) data may be sufficient. Synthetic Aperture Radar data can be extremely useful for observing below persistent cloud cover to look for disturbances on the landscape, but only a short historical record is available, and interpreting SAR data may be less straightforward. A large volume of high quality data is now available for free public use, which greatly

[12] https://modis.gsfc.nasa.gov/
[13] https://sentinel.esa.int/web/sentinel/home
[14] https://landsat.usgs.gov/

reduces the cost of performing remote sensing analyses. Much of this data is easily accessible through Google Earth Engine[15] and can be imported into CEO.

Exercise 4.3: Evaluating Imagery for your CEO project

Introduction

Selecting imagery appropriate for your CEO project is one of the most critical steps to ensure a successful photo interpretation. Often the availability of high resolution imagery for photo interpretation can be a major limitation and feed back to influence classification systems (and in some cases, even the CEO project objectives). In this exercise, we will learn about imagery available in CEO as well as options for additional image sources. We will evaluate image characteristics in terms of project objectives and classification systems. The Ecuador LULC mapping project will provide an example of how the problem of image selection has been tackled. You will then check your knowledge before moving on to evaluating and selecting imagery for your own CEO project.

Objectives

- Understand how objectives and classification systems affect image selection
- Learn how to evaluate and select imagery for photo interpretation
- Determine whether available imagery is appropriate to support your CEO project

Part 3: Evaluating available imagery for a CEO project

A. Imagery datasets available in CEO

CEO readily provides access to several image datasets. These imagery datasets are collected by a suite of different onboard satellite sensors. Some imagery in CEO comes from free and publically available sources, such as the well-known Landsat datasets, while others, such as some high resolution datasets from World View satellites, come from other commercial sources such as Digital Globe and Bing Maps. CEO is working to incorporate more and more imagery as it becomes available, and the suite of available image datasets will likely expand.

1. Examine the example scene as shown in the readily available imagery datasets in CEO. Notice that all images are shown in natural color (RGB). This is the default view of these image datasets in CEO.
   i. Note the features that can be distinguished in each using the seven essential image characteristics.
   ii. *Optional* - Click the links to learn more about each dataset.

[15] https://earthengine.google.com/

Figure 26. Example of [Bing Aerial imagery](#) (variable sensors). Spatial: 1-5 m; Spectral: RGB; Temporal: Infrequent/Variable.

Figure 27. Example of [Digital Globe](#) imagery (variable sensors); Spatial: 1-5 m; Spectral: RGB; Temporal: Infrequent/Variable

Figure 28. Example of ESA Sentinel-2 imagery; Spatial: 10 m; Spectral: RGB; Temporal: Frequent: 5 days.

Figure 29. Example of [USGS Landsat](#) (8, 5 and 7) imagery; Spatial: 30 m; Spectral: RGB; Temporal: Frequent: 16 days

B. Accessing additional imagery in CEO

Availability of high resolution imagery is variable in different areas of the world. Many projects require higher spatial resolution than what is provided by the readily available global datasets such as Landsat and Sentinel-2. Often, more frequent imagery (temporal resolution) is also required. It is crucial to evaluate the availability of suitable imagery for your project in your study area early in the project planning process. Some countries support ongoing programs that regularly collect high resolution imagery such as the National Agricultural Imagery Program (NAIP) in the United States*. If your project requires additional imagery, you may have to investigate and acquire other data to incorporate into your CEO project.*

1. If the imagery in CEO is not sufficient for your project, it is possible to bring in outside imagery by linking CEO to outside image repositories. For information on how to do this, refer to the CEO technical manual.
2. Review the list of outside image repositories below and click the links for more information:

Table 4. A brief list of possible sources of earth observing data for use in your project.

Data Provider	Documentation
Google's Earth Engine	Description of Available Datasets
Committee on Earth Observation Satellites (CEOS)	CEOS Database Handbook
Group on Earth Observations (GEO)	GEOSS Portal Documentation

Part 4: Imagery characteristics for interpretation

A. Spatial resolution determines what landscape features are identifiable

1. Review the example scene below shown at varying resolutions from different sensors.
   i. Observe how the seven essential image characteristics are affected by scale (spatial resolution).
   ii. Which characteristics are most important at each resolution?

Figure 30. US National Agricultural Imagery Program data (1 meter): Many features are larger than the image scale and are therefore relatively easy to distinguish. Shadows are visible and individual structures are recognizable. Image texture provides information on vegetation.

Figure 31. Sentinel-2 (10 meter): Some features are obscured at this scale. Structures and roads are vaguely clear. Color become increasingly important and allows us to distinguish stands of trees from grass vegetation.

Figure 32. Landsat (30 meter): At this scale, many of the features are obscured within the relatively large pixels. Fine-scale pattern and shadows are lost and color becomes the most important characteristic allowing us to still distinguish large features.

2. From the examples, which imagery would be suitable to differentiate or identify the following:
   i. Individual trees
   ii. Housing and structures
   iii. Roads and lots
   iv. Forest vs. non-forest
   v. Managed vegetation vs. unmanaged vegetation
   vi. Bare ground vs. herbaceous vegetation
3. Answers
   - 1ii. At fine image scales, texture, size and shadows are easily observed. As the resolution becomes coarser, association and tone become increasingly important. At Landsat scale, tone is the most important of the characteristics.
   - Suitable imagery: *Note that answers will depend somewhat on the study area*
     - 2i. Individual trees – 1-meter
     - 2ii. Housing and structures – 1 to 10 meter
     - 2iii. Roads and lots - 1 to 10 meter
     - 2iv. Forest vs. non-forest 1 to 30 meter though small patches will be obscured at 30 meter scale.
     - 2v. Managed vegetation vs. unmanaged vegetation- 1-meter
     - 2vi. Bare ground vs. herbaceous vegetation - 1 to 5 meter

B. Spectral (color) characteristics also determine identifiable features

We can obtain additional spectral information from multispectral imagery to help differentiate features. This can be particularly important in cases where high resolution imagery is not available. In the case of vegetation, near infrared reflectance is helpful for distinguishing vegetation types as well as live and dead vegetation. We can also compute vegetation indices from multispectral images which further enhance our ability to differentiate vegetation features. Though not available by default, CEO does support the integration of additional image views of multispectral imagery.

1. Review the false color composite (Near infrared displayed as red) image pairs of the same scene below as viewed by different sensors (aerial NAIP in Figure 33 and Landsat in Figure 34 below). Observe the utility of near infrared reflectance for distinguishing vegetation.

Figure 33. A set of true color (red, green, blue), and false color (near infrared, red, green) composites for the same location, based on very high resolution imagery from the US National Agricultural Imagery Program.

Figure 34. A set of true color (red, green, blue), and false color (near infrared, red, green) composites for the same location, based on Landsat imagery.

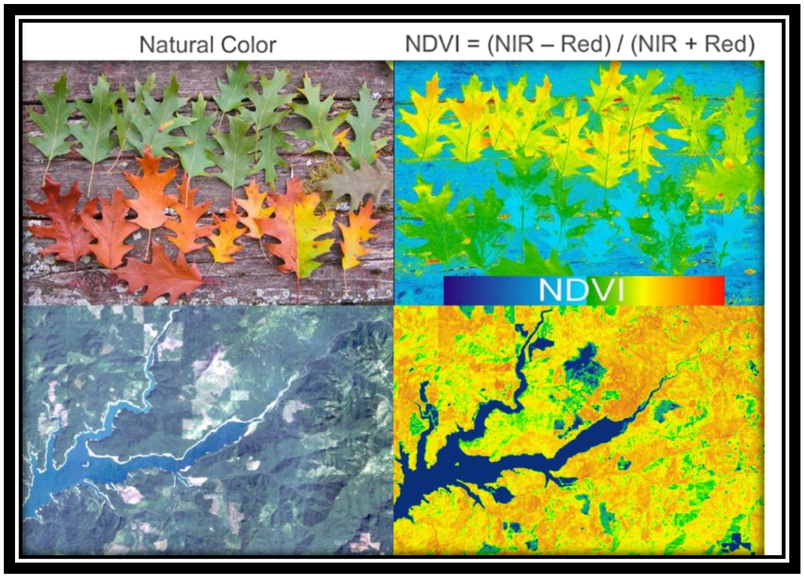

Figure 35. Natural color images (left) and normalized difference vegetation index (NDVI) images (right) demonstrate the ability of NDVI to distinguish healthy, live vegetation from senesced vegetation or background.

2. Examine the images above in Figure 35. Natural color and NDVI image pairs show how healthy vegetation (greenness) can be enhanced by viewing the normalized difference vegetation index (NDVI). The top photos are taken from a high resolution camera while the bottom

C. Image timing and interpretation of change

1. If your project is focused on identifying landscape changes, multiple image dates representing the time periods of interest are likely required. Examine the example two-date image pair below and note the changes in the plot.

Figure 36. A US National Agricultural Imagery Program image taken in 2013, showing a stand of forest pre-harvest.

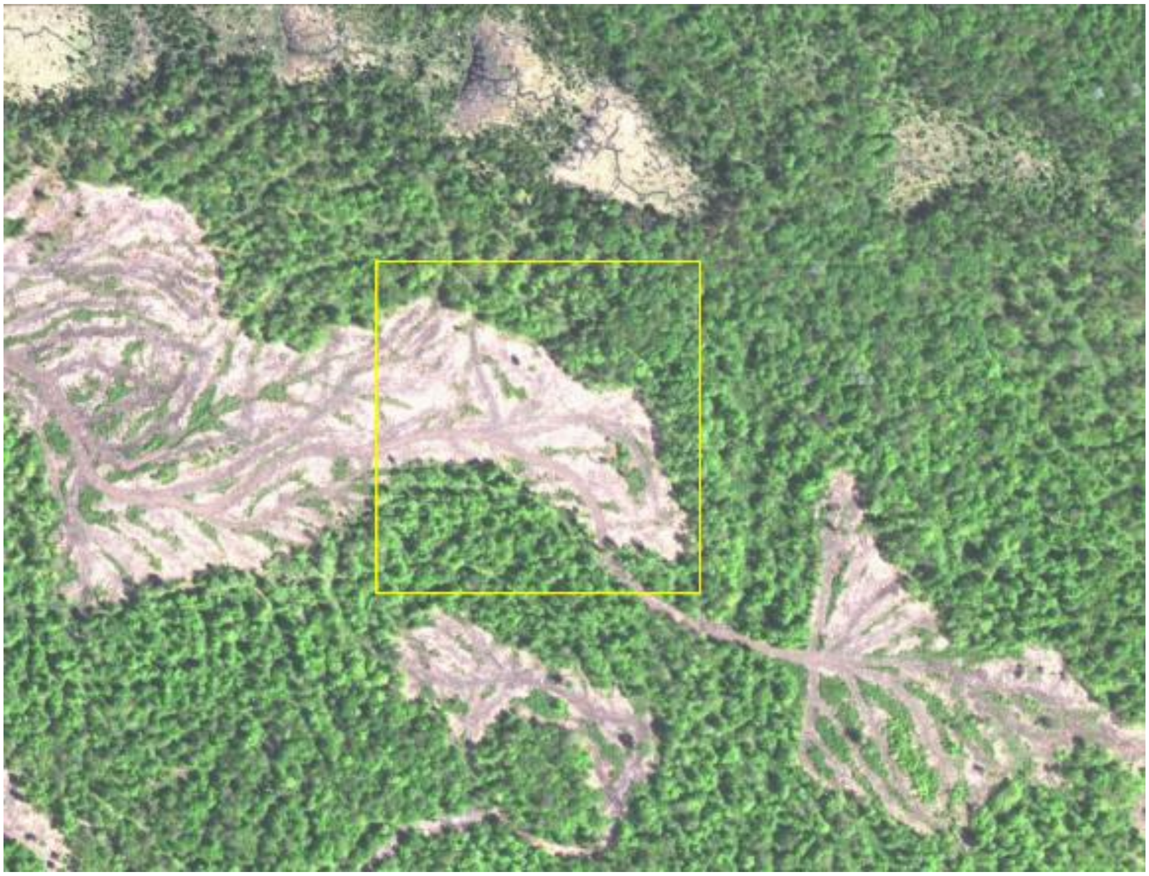

Figure 37. A US National Agricultural Imagery Program image taken in 2015, showing the same stand of forest as in Figure 36 post-harvest.

2. In addition to enabling change interpretation, image timing can be important to consider if the features you are interested change seasonally. In some cases, looking at imagery from a specific season or from multiple seasons can help to identify targets. Review the example image pair below in Figure 38 illustrating how seasonal effects can make it easier to distinguish certain types of vegetation.

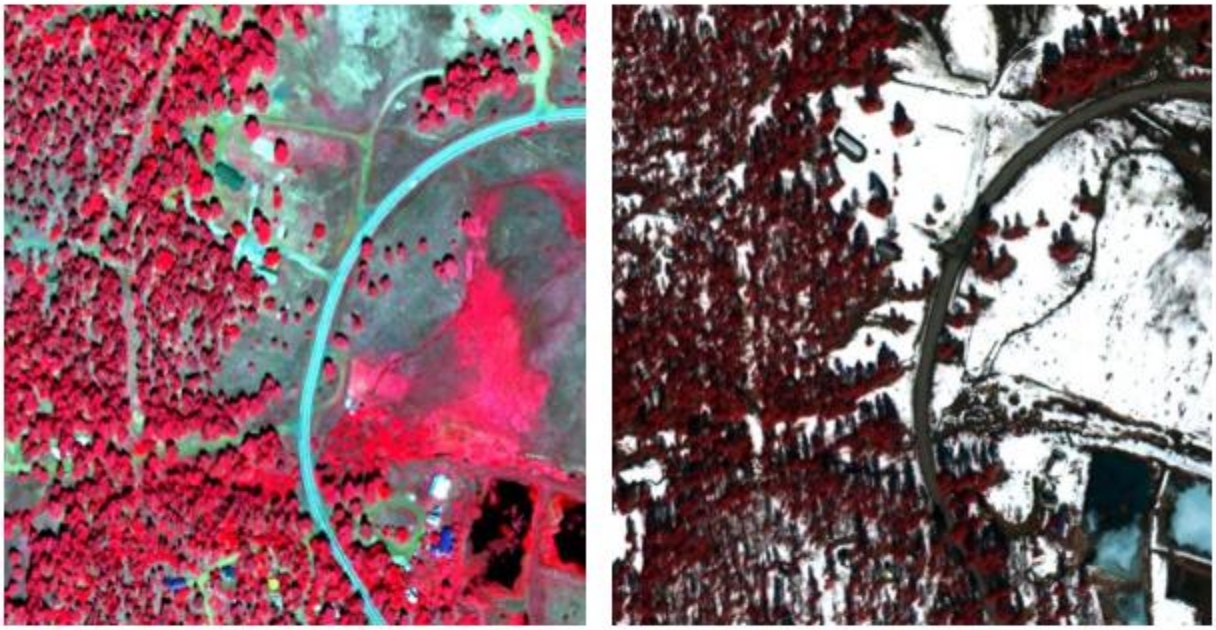

Figure 38. Left panel: A WorldView-2 (0.5 m) taken in midsummer – The summer color infrared image clearly shows herbaceous and understory vegetation along with the evergreen trees. Right panel: the same area in a winter color infrared image. The evergreen trees are easily distinguished due to their 'greenness' and the snow cover.

3. Interpreting change between multiple image dates can be complicated as seasonal changes and true landscape changes can both be present in the same image simultaneously. Examine the scene pair below in Figure 39 and note the significant landscape changes as well as the seasonal changes in vegetation.

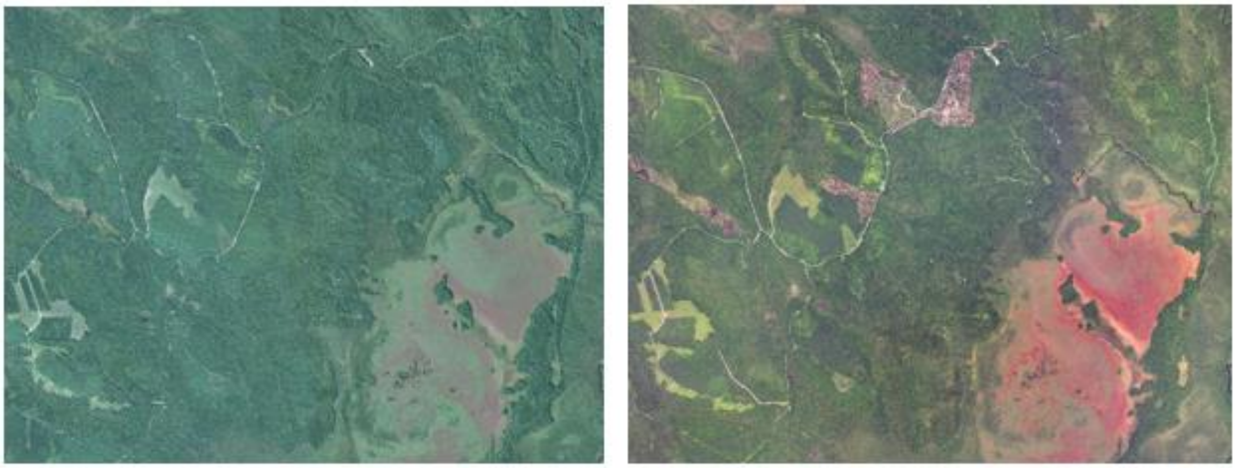

Figure 39. Left panel: A 2013 US National Agricultural Imagery Program image (1 m) – This image is taken early in the growing season (greener) but also before several harvest activities. Right panel: A 2015 National Agricultural Imagery Program image (1 m) – This image is taken later in the growing season and contains senesced vegetation. Multiple harvest activities have also occurred.

Part 5: Explore CEO imagery for Ecuador LULC

A. Review project information needs

1. Recall that the goal of the Ecuador LULC project was to map cover and use in two time periods and create a map of change. Specifically, this project wanted to create a comprehensive 2014 LULC map of Ecuador, a 2016 LULC map and a map of change classes.
2. The cover and use classes of interest were detailed and described in the classification system that we looked at in Exercise 3.1. The classification system included several fairly detailed vegetation classes. This meant that high resolution imagery was required for interpretation.
3. Given these needs, this project required high resolution imagery from both 2014 and 2016. This imagery would need to be available across the entire country.

B. Explore and evaluate available imagery – Identify problems

1. The first step was to explore the imagery already available in CEO. To do this, a small test project was created and all the available and potentially relevant image layers were added. The test plots were then examined with the classification system on hand.
   i. The imagery included the following layers in CEO: *DigitalGlobeRecentImagery, BingAerial, DigitalGlobeWMSImagery, PlanetGlobalMosaic* as well as several different views of Landsat 8 and Sentinel-2 mosaics.
2. This process quickly revealed two key things: 1) high resolution imagery was required to distinguish many of the classes and 2) the available high resolution imagery was insufficient for the project.
   i. Specifically, there was a serious lack of high resolution imagery coverage representing 2014 conditions.

C. Imagery selection and acquisition for Ecuador LULC - Solution

1. The readily available high resolution imagery in CEO was selected, as it was sufficient for interpreting 2016 conditions.
2. Additional imagery had to be purchased to allow interpretation of 2014 LULC. Two meter resolution Rapid Eye imagery was available from 2014 and was purchased for the project. Due to persistent cloud cover across much of Ecuador, it is very difficult to get complete cloud-free coverage, and this Rapid Eye acquisition was also incomplete across the country.
   i. This presented a significant additional cost to the project and it was still not totally sufficient, as complete coverage of the country was not available, and the lower spatial resolution made it more difficult to interpret compared to the available 1 meter 2016 Digital Globe imagery.
3. To mitigate the issues with interpretation of 2014 Rapid Eye imagery as well as assist in areas where Rapid Eye was not available, additional spectral information was brought into CEO to illustrate multiyear, seasonal greenness (NDVI) and wetness (normalized difference water index [NDWI]) trends from available Landsat imagery. These data were used to help infer change. This topic will be discussed further in the next exercise.

Part 6: Evaluate and select imagery for your CEO project

Now that you have learned more about evaluating and selecting imagery that can meet photo interpretation project goals, you can begin the process of evaluating and selecting imagery for your own CEO project. This section simply guides you through the questions that you will need to answer in order to complete this part of the process.

A. Review information needs and define imagery requirements

1. Given your project goals and classification system, what type of imagery is likely to be required:
   i. What spatial resolution is necessary to identify your target features?
   ii. What spectral resolution is required to identify features? Will the default natural color image views suffice, or is it crucial to incorporate additional band information to identify features?
   iii. What temporal resolution do you need? What time period or periods must be represented by the imagery to capture the features of interest? *This is crucial if you need to represent a specific time period or characterize change.*

B. Check out available imagery and identify gaps

1. Review the available imagery. You can do this by setting up a test project allowing you to examine the imagery in light of your classification system.
2. Does the available imagery meet the needs of the project?
   i. It is crucial that you have imagery that allows you to see in fine enough detail to be able to identify the landscape feature of interest to your project.
3. If the answer to the last question was 'no', can you access additional information that will meet these needs? If not, is there other information that you could use to fill in the gaps?

C. Select appropriate imagery for your project

1. Use the information you obtained through the questions above to identify the imagery you will use for your project.
2. Set up your CEO project and configure it to incorporate the imagery and information you have selected.

Exercise 4.3: Image Interpretation

Introduction

In this exercise, you will learn more about image interpretation and review several example interpretations taken from one of the Ecuador LULC mapping project. You will learn more about how the principles of image interpretation that you explored in Exercise 4.1. Additional information can also be used to provide more context and tone information, and there are tools available in CEO to incorporate ancillary information from other sources. You will be introduced to these tools and learn how they can be used to improve your image interpretation.

Objectives

- Apply your knowledge of image characteristics to example images
- Get practice interpreting points and plots
- Learn how CEO tools can provide more information
- Be able to plan your own CEO image interpretation setup

Part 1: Review CEO examples-Ecuador LULC Mapping

Review the examples to observe how the seven image characteristics you learned about in Exercise 4.1 can be helpful in distinguishing features in the example plots from the Ecuador LULC mapping project.

A. Color easily differentiates barren points from those on a glacier

1. While this example may seem obvious, these points might be difficult to separate without the color information.

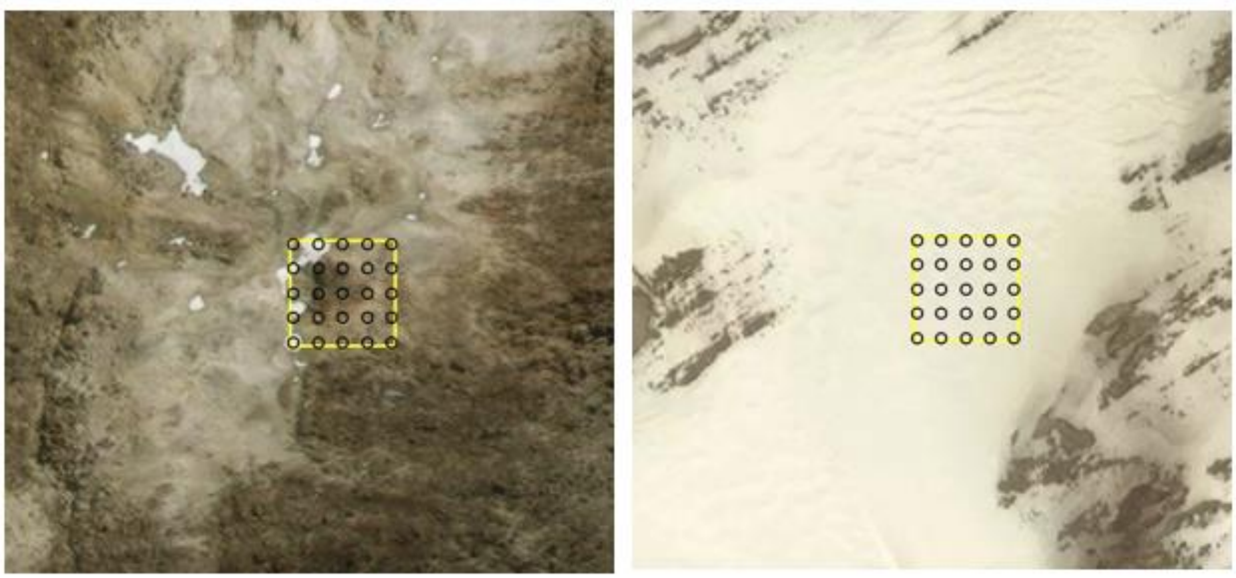

Figure 40. Left: a high elevation area in the Ecuadorian Andes, with small patches of snow and bare rock. Right: A high altitude area of the Ecuadorian Andes, covered in snow and ice. Without color/tone information, these areas would likely be appear very similar.

B. Image pattern and texture can be used to differentiate classes similar in color or size

1. Primary Forest vs. Plantation – Though somewhat similar in tone, the plantation trees are organized in clear, cultivated rows, and the difference in pattern and texture is apparent.

Figure 41. Primary forest (left) and forest plantation (right) in Ecuador. Note that the presence of distinct pattern and texture helps separate these classes.

2. Herbaceous/Grassland vs. Shrubland – Observe how shrub vegetation presents coarser texture in the image than the lower lying, comparatively smooth, herbaceous/grassland.

Figure 42. Grassland or herbaceous cover (left) and shrubland (right) plots. The smoother texture of the herbaceous cover, combined with patterns associated with tufting and bunching of grasses help distinguish grasslands from the rougher, more clustered texture and pattern of shrubs.

3. Cropland vs. Herbaceous/Grassland – In the following example, you can see how the unnatural smoothness of the cultivated cropland distinguishes it from the unmanaged grassland.

Figure 43. Croplands (left) and grasslands or herbaceous cover (right). The extremely uniform texture and straight lines of the crops make it easy to distinguish from similar plants in a natural setting, which have little regular pattern or texture.

C. Observe how association, size and shadows can be used to extract information

1. Forest type by location context – Mangrove trees grow in close association with water, and this key clue can be used to identify mangroves and differentiate them from other coastal trees.

Figure 44. Mangrove (left) and primary forest (right). Mangrove can only often only be distinguished from other types of forest by context and association with open water.

2. Vegetation height from shadows and buildings – Observe how shadow length and the association with man-made objects can be used to determine that the vegetation in these examples are indeed trees (> 3 m in height).

Figure 45. Urban vegetation (left) and rural vegetation (right). The presence of structures makes identifying the height of the vegetation in these scenes easier.

Part 2: Explore Additional Data – Spectral Time Series

In addition to examining the high-resolution imagery for interpreting landscape features, it can be helpful to look at additional remotely sensed data to provide additional spectral (color) information. Other image sources that may lack the spatial resolution for photo interpretation may be available at finer (more frequent) temporal resolutions and thereby provide insight into landscape features. Spectral time series data are particularly useful to incorporate in projects where it is important to interpret landscape change.

A. Spectral time-series data may help distinguish vegetation types

1. Forest

   i. In the example below, clouds partially obscure the plot and hinder interpretation. In addition to looking at the high resolution imagery, we can look at the NDVI (greenness) time series for this plot. From here, we can see high and seasonally stable greenness values, which supports our interpretation of forest.

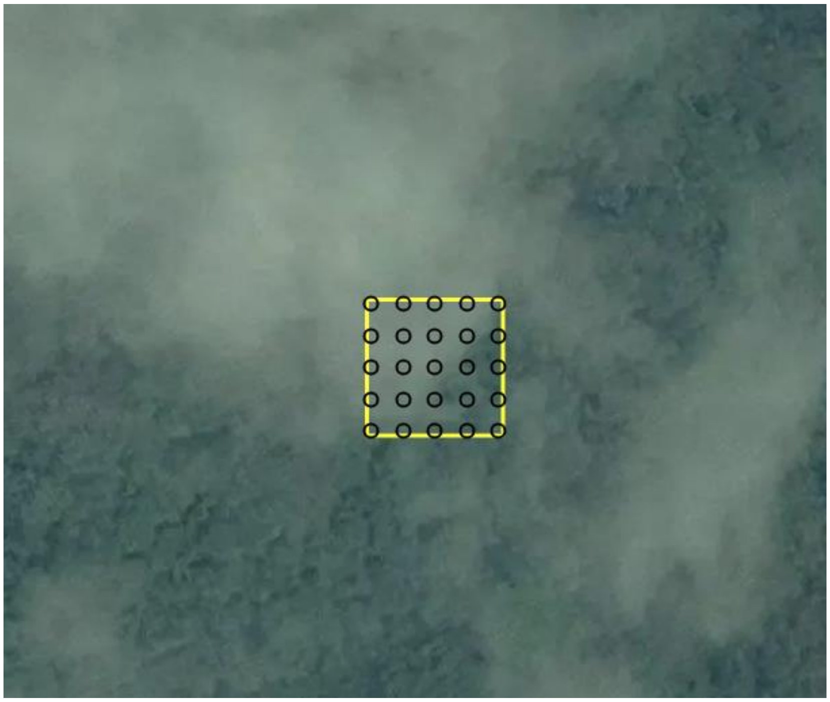

Figure 46. A plot from the Eastern Amazon in Ecuador. Many parts of the world have frequent cloud cover, making direct interpretation of imagery more difficult.

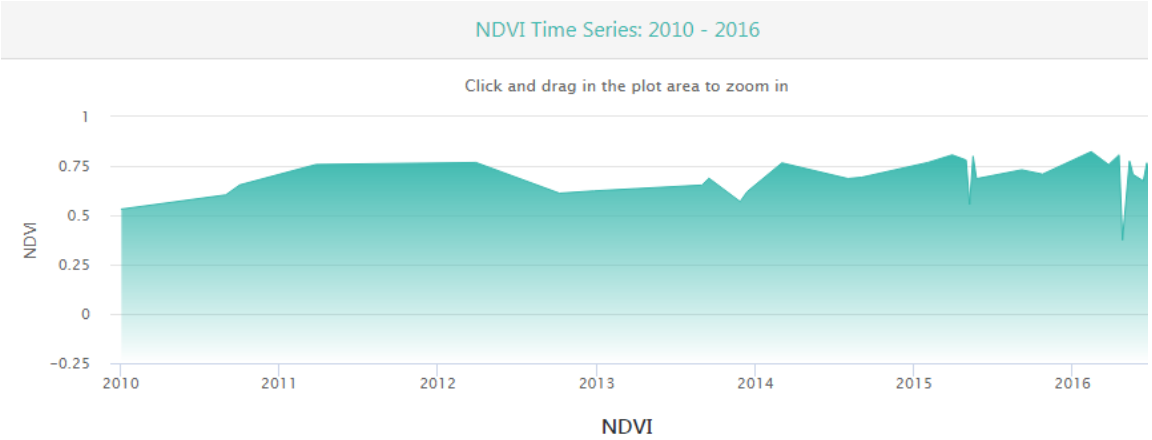

Figure 47. A Normalized Vegetation Difference Index time series. NDVI is a very useful proxy for vegetative cover. Uninterrupted high NDVI values suggest continuous living vegetative cover.

2. Herbaceous/Grassland
    i. In the example below, we can see a high elevation grassland with some shrub cover (Paramo). In addition to color, texture, etc., we can see the seasonal peaks and valleys in the NDVI time series indicating a seasonal green-up and browning following the seasons. This is consistent with our expectations for grasslands.

Figure 48. A high altitude vegetation area in Ecuador, containing a mix of grasses, herbaceous plants, and shrubs.

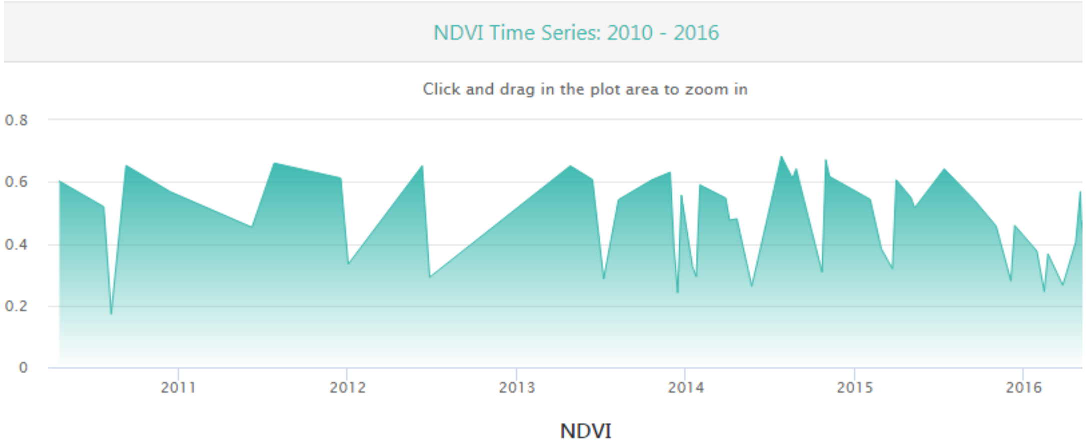

Figure 49. The Normalized Vegetation Difference Index for the plot show in Figure 48. Note the large changes that occur; many of these are likely related to seasonal changes in vegetation.

Part 3: Photo Interpretation of Example Plots

In this section, you will examine several example point and plot interpretations from the Ecuador LULC mapping process.

A. Mangrove and natural water in an area with aquaculture

1. Examine the plot and interpretations below. The plot contains 25 points that have been interpreted using the high resolution 2016 Digital Globe imagery. Of the 25 points, 18 of these intersect the dense mangrove canopy. Therefore, in this interpretation, this plot has about 72% mangrove cover.

Figure 50. An example plot from Collect Earth Online, showing the labelling of points within the plot; 18 are the 'mangrove' land cover element.

B. Herbaceous and grass vegetation with shrubs

1. Examine the interpretation of the example plot below. Using interpretation rules that will be discussed further in later exercises, this plot has been interpreted to contain

herbaceous/grass vegetation (light green), shrubs (olive green) and bare ground (gray in the bottom left).

Figure 51. An example plot from Collect Earth Online, showing the labelling of points within the plot; the plot is a mixture of grassland/herbaceous cover, shrub cover, and bare ground.

C. Tricky classes - Differentiate between trees and shrubs in a plot?

1. In the Ecuador project, trees and shrubs were separated using a 5 meter height cutoff. This made it somewhat difficult to differentiate trees from shrubs.
2. In the example below from the Ecuador mapping project, do you think the points within the plot fall primarily on small trees or large shrubs?

Figure 52. An example plot from Collect Earth Online; this plot highlights some of the difficulty in telling the difference between trees and shrubs.

3. Use of shadows and association of other objects to differentiate between trees and shrubs
   i. To address the problem of differentiating shrubs from trees, we zoomed out and found some nearby structures for comparison.
   ii. We can see that the vegetation in question is higher than the structures

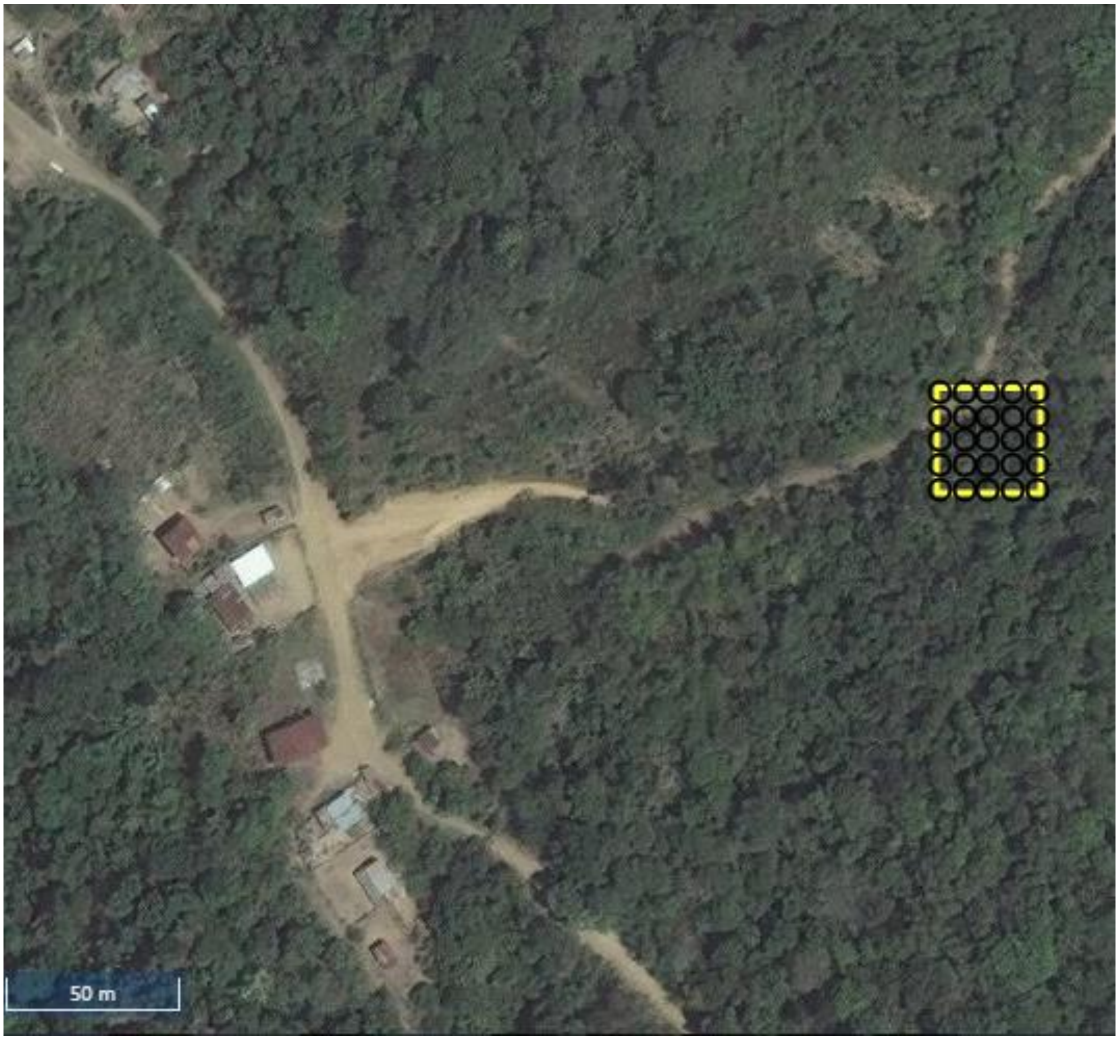

Figure 53. An example plot from Collect Earth Online; this zoomed out view of the plot in Figure 52 shows context that we can use to help determine what the vegetation in the plot is composed of. In this instance, we can clearly see that the individual crowns of plants are bigger than some of the man-made structures in the image, suggesting that they are trees.

4. From this, we can infer that these are indeed small trees and make the following interpretations: most points fall on trees, a few fall on the herbaceous/grass understory and some on the bare ground dirt road.

Figure 54. An example plot from Collect Earth Online, showing the labelling of points within the plot show in figures 52 and 53.

Part 4: Apply and check your knowledge

A. Review your classification scheme and imagery selected for your CEO project

1. If possible, simply review the list of classes in your system while viewing some example imagery that you will use for interpretation.

B. Use your knowledge of imagery and interpretation to answer the following questions.

1. Of your classes, which ones do you think will be most easily interpreted from the imagery?
2. Which problems will be most difficult and why?
3. What additional information will you use to assist interpreters with difficult classes?
   i. This could be additional data or simply expert knowledge and advice.
4. If additional data is required, how could you incorporate it into CEO?
5. If expert knowledge and advice is required, can this information be included in the photo interpretation key?

5. Sample and Response Design

5.1 Overview

Any project where the goal is to describe aspects of a real, physical population may make excellent use of design-based inference. This approach is not the same as statistical model-based inference. Design-based inference is often used to

> *"address sampling problems, where a sampling problem is defined as one in which (1) interest focuses on collective properties (parameters) of a real, existing population, (2) the collective parameters do not need to be known with certainty, and (3) the attributes of all individual elements of the population are not required to be known."*[16]

In other words, you are looking at the real world, and you don't need or want to perform a perfect census of the thing you are interested in (which would provide a completely certain estimate of the population's parameters and also tell you about all the individual elements). In these situations, a well-prepared sample can provide a robust estimate of the parameters of interest for the population (percent forest cover, for example). This approach corresponds to most projects for which you might want to use CEO.

A project that uses design-based inference will be composed of two parts: the response design and the sampling design, each of which will be discussed below.

The response design describes much of the structure of the process: what will be sampled (the unit of assessment), what will be considered when interpreting the objects sampled (the spatial support unit), and how values will be assigned to interpreted objects (the rules of agreement and labelling protocol).

The sample design describes the manner in which the samples used in the response design will be located and collected and how many of them to collect. The choice of sample design also influences what equations will be used to estimate the population values and the uncertainty of those values.

It should be noted that a design-based inference approach does have limitations or drawbacks. In particular, it does not allow for estimation of the value a given member of the population of interest, unlike a statistical modeling process, nor does it necessarily allow for one to discuss the distribution of the members included in the population. The results of a design-based approach are also strongly contingent on the sample and response design. This is why it is extremely important to create a robust and valid response and sample design.

5.2 Response design

The *response design* details how data will be collected, interpreted, and related to the map data being assessed. The key component of your response design will be your labeling protocol, or how samples will be assigned to your LULC classes.

[16] Stehman, S. V. (2000). Practical implications of design-based sampling inference for thematic map accuracy assessment. Remote Sensing of Environment, 72(1), 35–45. http://doi.org/10.1016/S0034-4257(99)00090-5

Unit of Assessment

The unit of assessment is the object or area of the landscape that you will be interpreting (using the labeling protocol, below) to either compare with the output of your map making process, or as sample in your inventory, etc. In an accuracy assessment context, the unit of assessment may be the same as the output unit of your map, but it does not need to be the same. The unit of assessment can be composed of pixels, blocks of pixels, or polygons/segments/objects. Each of these units brings with them distinct advantages and disadvantages (a subject the reader is encouraged to investigate more deeply)[17]. For accuracy assessment or model training purposes, the pixel (or an area corresponding to a pixel in your map product) is likely the most commonly used unit of assessment. This provides the advantage of simplicity, straightforward interpretation and relatively easy statistical methods.

Note here that the unit of assessment is set *relative to the data being investigated*. In the case of a map validation, this means it's a property of the map having its accuracy assessed. In this case, a 'pixel' will be a pixel of that map. In a map derived from Landsat data that would be a 30 meter square. In CEO this would be a 30 meter square plot with some number of points placed in it. Each point may be considered to represent a certain percentage of the plot area, and the more points you add to the plot, the more completely and accurately described it will be.

When performing inventories, a polygon or block of pixels (a plot) is a more natural choice, as higher resolution data is will likely be used, and a pixel in that context will likely be smaller than the landscape elements you will wish to identify.

Spatial Support Unit

The unit of assessment is used in combination with a spatial support unit. The spatial support unit is an area around the unit of assessment that provides context for classifying it. In some cases the spatial support unit is identical to the unit of assessment, but often it will be defined as a larger area. This is necessary when you need to consider the context of a particular sample. For instance, a patch of herbaceous vegetation the size of your unit of assessment may exist in the middle of a forest, in an agricultural context, or in the middle of a city. Depending on how your classes are defined, that patch of vegetation may be described as open forest, croplands, settlement, "cultural vegetation" or some other class, depending on how the context in which it exists relates to your classification system. Even when the spatial support unit is not formally defined, you will find that photo interpreters naturally attempt to use the context around the area of interest. It's therefore a good idea to establish guidelines for how they do so.

Labeling Protocol

Choosing LULC Classes

The first component of your labeling protocol is setting up the class structure of the data that will be collected about the unit of assessment. It may seem obvious that the LULC classes that you will use in your response design will directly correspond to those of the map being validated – this is generally the case. However, as discussed in Classes: Points vs. Plot, it sometimes may make sense to collect data at

---

[17] Stehman, S. V., & Wickham, J. D. (2011). Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment. Remote Sensing of Environment, 115(12), 3044–3055. https://doi.org/10.1016/j.rse.2011.06.007

the level of points, and aggregate that data into classes for the plot. In this case, the classes used to label the points will differ from those used to label the plot. Essentially a second classification system is created to produce a set of landscape elements that can be assembled into your LULC classes, with an associated set of rules.

Dealing with Heterogeneity

If you choose not to structure your labeling protocol around extremely basic land cover elements, but instead will identify broader areas, then you should have rules for dealing with heterogeneity. For instance, suppose you have a plot with five points. Each point “represents” 20% of the plot, and you may wish to interpret the area around the point, rather than simply the area directly under it. In this case, you will likely have several different types of land cover elements included in the point label. Creating a set of rules for dealing with various situations helps produce consistent results. For instance, you might create a rule that trees are given preference over other vegetation (due to canopy), or that the presence of water in the vicinity of the point should lead it to be classified as water.

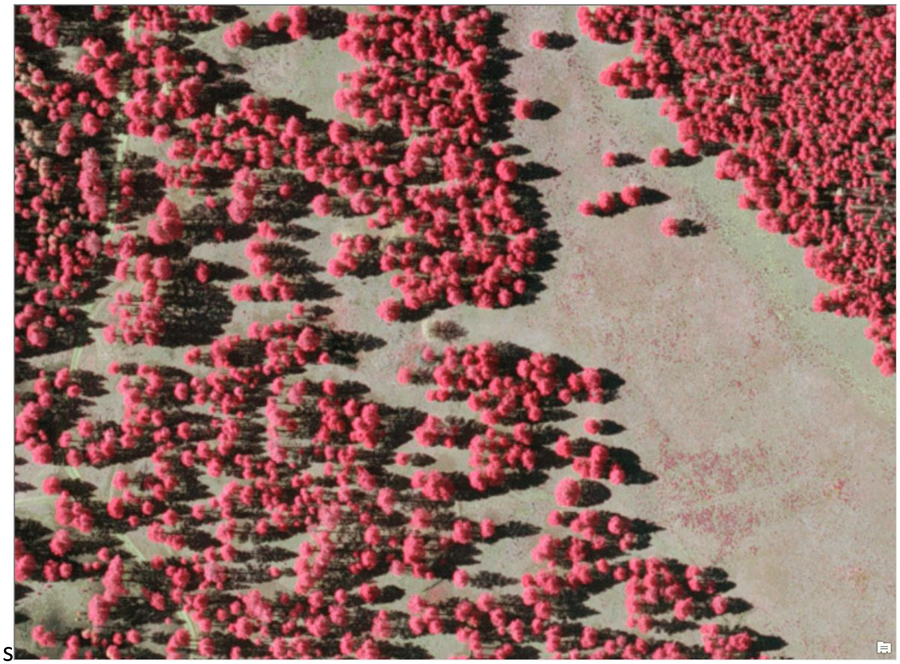
s

Figure 55. An example color infrared image showing a heterogeneous landscape. Different labelling approaches could lead to his area being classified a variety of ways.

In Figure 55 for instance, a point falling between two trees on the left hand side of the image could be interpreted several ways. If treated as a pin-point, it would likely be labelled as something like herbaceous cover, or something similar. If treated as a small area, or if trees are given priority over other vegetation, it would likely be labelled as tree canopy. These two approaches for labelling points would produce very different estimates of canopy cover or a plot-level class for this area. Establishing guidance for dealing with these types of heterogeneous landscape is important for producing consistent results.

Using Rules

Rules can be established that allow for consistent labeling in ambiguous situations. For instance, points will often land on the edge of two different land cover elements; edges are a particularly common problem in developed or highly disturbed areas. This situation can be dealt with by defining a "right hand" or "north" rule, where the land cover to the right or north of the point is the class that is assigned.

Rules become especially important when looking at more than one period of time, or otherwise attempting to determine change. When looking at multiple times of VHR imagery, it is quite likely that there will be some displacement of objects just due to changes in angle of view, position of the sun, or other changes inherent in the imagery. It may also be that certain kinds of changes are of particular interest. If, for instance, the loss of primary forest were of interest, then simply noting the presence of trees in two times for the same plot would not be sufficient if there had been some disturbance between the two images and the tree in the later image was regrowth.

Exercise 5.1: Planning the Response Design

Introduction

With clear objectives, classification system, imagery and interpretation resources assembled in your photo interpretation, we are almost ready to set up a CEO project. First you need to know exactly what you are assessing (unit assessment) and how you are going to label it by carefully planning your project response design. In this exercise, we will review the example LULC mapping project in Ecuador, review the response design and rationale, check your knowledge and help guide the response design process for your CEO project.

Objectives

- Understand the importance of the response design
- Know how to choose the unit of assessment
- Understand the spatial support unit
- Be able to designed labeling protocol and rules

Part 1: Review the Ecuador LULC Response Design

A. Ecuador LULC mapping – Response design summary and key points

Recall that the objectives of this project were to create LULC and change maps for all of Ecuador. CEO was used at multiple stages in this process but primarily to collect training data for modeling and reference data for final map validation. These objectives and considerations largely determined the response design. Because mapping was done at the scale of 30 m pixels, the unit of assessment was a 30 m plot. As you have seen from the examples shown from this project, a combination of points and plots were used in this project. The unit of assessment was actually the 30 m plot, and the grid of 25 points *within each plot were simply sub-labels to assist in the interpretation process. The use of points within plots streamlined the interpretation process and allowed for more precise and consistent interpretations, as individual features were interpreted at a finer scale within the plot and then summarized at the plot scale. Another way to think of this is that because each plot contained 25 points and each was labeled individually, plot cover was interpreted in roughly 4% increments. Tie breaking rules were implemented for cases in which points directly intersected more than one class. Precise definitions in the LULC classification system were used to summarize point classes to plot classes. Along with these definitions, a priority system was required in order to assign classes in cases where multiple criteria were met.*

B. Unit of assessment – 30 m^2 plot

1. The plot as labeled by summarizing and applying point cover thresholds.

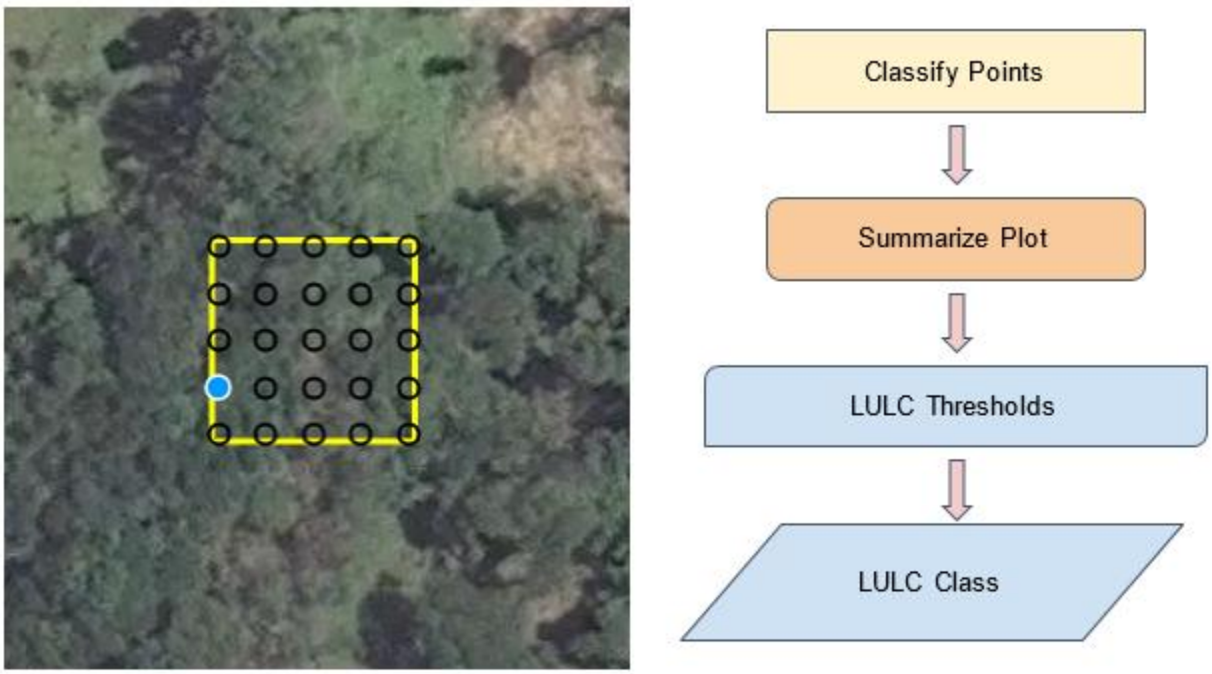

Figure 56. Example plot with points showing the unit of assessment for the Ecuador land use and land cover mapping project. Rules and thresholds guided the assemblage of point labels to generate plot labels.

C. Spatial Support unit – Visible area extent at a specified scale

1. The area of land around the plot visible in CEO at the default viewing scale (20 m). This area provides context and essential image information that assists interpretation.

Figure 57. In this example image, the unit of assessment (UA) is the plot outlined in yellow while the spatial support unit (SSU) is the area extent visible in Collect Earth Online at the 20 m scale.

D. Labeling protocol and rules:

1. Point labeling rules - If a point directly intersected more than one class, t*he "north rule" was used to break ties in cases where points were located on the boundary of two or more classes. In cases where this was insufficient, the "right hand rule" was also implemented.*
2. Imagery use rules – Multiple imagery datasets were used to support interpretations. These were ranked in terms of quality and reliability, and priority was given to the 'best' imagery, with lower ranked imagery used to support or when 'better' imagery was not available.
3. Point to plot summary thresholds – Each plot and map class was defined according to the point cover thresholds and a simple priority system was implemented to ensure consistency.
   i. Example 1 – Plots with at least 30% Primary Tree cover are labeled as Primary Forest (see below)

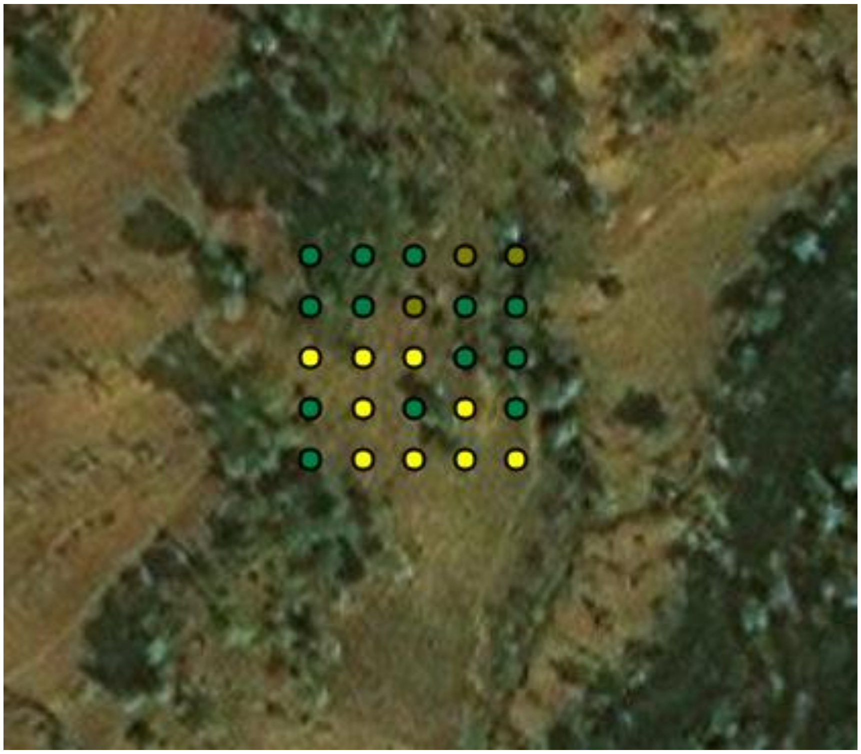

Figure 58. A primary forest plot. It contains 52% tree cover, 36% crop cover, and 12% shrub cover.

ii. Example 2 – Forest was given priority over agriculture (read example below).

(a) In a plot with 25 points, 8 of which (32%) were labelled as Primary Tree, 4 (16%) were labelled as Shrub Vegetation, and the rest (13, or 52%) were labelled as Crops, then the plot could have been labelled as either Cropland or Primary Forest as both criteria were met. Forest Land must have tree cover greater than 30%, and Cropland must have crop cover greater than 50%. To deal with this, Forest was given priority over Cropland as a rule.

Part 2: Check your knowledge of response design

A. Review theoretical examples related to canopy cover interpretation in CEO below

1. Canopy cover inventory project:

   i. Objective – Use CEO to obtain a single inventory estimate of tree canopy cover across a relatively large study area. Recall that this approach won't produce a map, rather a statistical summary of the tree canopy cover.

2. Canopy cover mapping training data project:

   i. Objective – Use CEO to gather a reference dataset that will be used to train a model to map continuous canopy cover (i.e., not cover categories) values at the pixel scale using Landsat imagery (30 m) across a large area.

3. Canopy cover map validation project:

i. Objective – Use CEO to obtain a reference dataset that will be used to validate a continuous pixel canopy cover map.

B. Answer the following questions related to each example above

1. For the canopy cover class inventory project, what do you think would be the simplest and most appropriate unit of assessment and why?
2. For the canopy cover mapping training data, how would you design the response and why?
3. For the canopy cover map validation project, how would you design the response and why
4. For each example, what might be an appropriate spatial support unit (SSU) and why? If you are unsure, what additional information about the project would you want know before being able to make this decision?

C. Answer review

1. B1. For the inventory project, you could use a simple plot with one to five points in it. The size of the plot in this case is somewhat arbitrary without further information about the project goals. In this example, using a circle might be best to reduce edge effects in interpretation, especially if multiple imagery sources are used.
2. B2. To produce training data for the canopy cover mapping project, the unit of assessment should be a plot at the same scale as the map source data (30 m for Landsat data). The number of points within the plot is determined by how finely canopy cover needs to be mapped. More points lead to a finer canopy cover map; however, interpretation time is increased.
3. B3. The response design of the map validation project is likewise dependent on the map itself and will likely be the same as the training data response design.
4. B4. Due to the simplicity of the classification, the spatial support unit is less critical in these examples and could probably simply be the plot itself. This is likely all that is required to provide context for interpreting tree canopy cover.

Part 3: Planning your response design

A. Review your project objectives and classification system, and answer the following questions

1. Is your project an inventory or mapping project, and how does that determine the unit of assessment?
2. How do the objectives dictate the size of the unit of assessment?
3. Can your classes be reliably interpreted at the plot-level, or are multiple points required?
4. What level of precision is required in your interpretation, and how does this determine the number of points?
5. How will you define your spatial support unit for your objectives?
6. What labeling rules are required to ensure consistency?
7. Are labeling priority rules required?

5.3 Sample design strategies

Overview

Your sample design will determine what areas are included in the data you collect and is thus one of the key components of designing an effective and robust project. A properly designed sample is often the most important factor in producing a reliable inventory or accuracy assessment[18]. It is also a deep subject on which a great deal has been written, with entire books and hundreds of research papers available that contain advice on how to design a sample for a particular purpose. Below is a brief introduction to some of the principles of sample design, as they relate to LULC mapping.

There are a variety of approaches that can be taken when designing a sample for inventories or accuracy assessment of remote sensing products, ranging from the simple and direct to the very sophisticated and complex. These include simple random sampling, systematic sampling, cluster sampling, stratified versions of these, and more complex approaches such as Generalized Random Tessellation Stratified sampling (GRTS)[19]. However, for all approaches, the goal of a sample is to provide an unbiased estimate of some population measure (proportion of area, often), with the smallest variance possible, given all the other constraints we face in the process[20]. These constraints can take the form of budgetary restrictions, narrow windows of time in which to perform the needed work, a lack of workers with the necessary skill set, the need to be able to explain the work to stakeholders, availability of high quality imagery, limited capacity to perform necessary field work for ground-truthing, and many other possibilities. All of these factors should be considered when thinking about the sample design.

Two of the most important factors when designing your sampling systems are to ensure that you are arriving at a *probability sample design* and that the sample is *geographically balanced*.

Using a *probability sample* design means that every sample location should have some probability of being sampled, and ideally this probability should be well-defined and understood.

The sample design being *geographically balanced* means that all regions within the area being studied are properly represented (none are excluded from the data). Geographic balance is a more nebulous and flexible concept than probability sampling, as we may sometimes have good reason to place more samples in some areas than others (areas of high heterogeneity might need more samples to be represented, for instance).

These two properties of your sample are closely related, and many of the barriers to ensuring they are true are also shared. The vast majority of the tools we use to measure accuracy or estimate population metrics from samples depend on all samples having some probability of being included in the sample

[18] Stehman, S. V. (2000). Practical implications of design-based sampling inference for thematic map accuracy assessment. Remote Sensing of Environment, 72(1), 35–45. https://doi.org/10.1016/S0034-4257(99)00090-5
[19] https://sites.google.com/site/r4naturalresources/natural-resource-topics/grt
[20] Stehman, S. V. (1992). Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing, 58(9), 1343–1350.

design[21]. It can often be difficult to apply these principles to field-based ground-truthing, as some areas may be composed of rough terrain or may be inaccessible for various other reasons (private property, etc.). The greatly increased public availability of VHR imagery has made using a true probability sample for many kinds of projects much more practical.

A sample design should[22]:

- Be as simple as possible, while still meeting the needs of the project;
- Take into account the resources available to the project;
- Be a probability sample;
- Have geographic balance;
- Produce a small variance in estimated parameters;
- Allow for relatively simple estimation of parameters; and
- Accommodate changes in sample size.

The Sample Unit

A first consideration when constructing a sample is to ask, "What is being sampled?" In most cases the sample unit will be the same as the unit of assessment (discussed in the response design section). So, if your project is using a plot (as is typical) as the unit of assessment, then plots will likely be your sampling unit. CEO makes it possible to determine the properties of your sample unit using sub-samples (points in a plot). Each point may be considered to represent a certain percentage of the plot area, and the more points you add to the plot, the more completely and accurately described it will be. As the points per plot increases, the variance of the estimates of the plot proportions (covers within the plot) will decrease; there are large decreases in the variance of estimates of plot proportions as the number of points per plot goes to 20, and a slower decrease there-after.

[21] Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: Fundamental principles. Remote Sensing of Environment, 64(3), 331–344. http://doi.org/10.1016/S0034-4257(98)00010-8

[22] Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. International Journal of Remote Sensing, 30(20), 5243–5272. http://doi.org/10.1080/01431160903131000

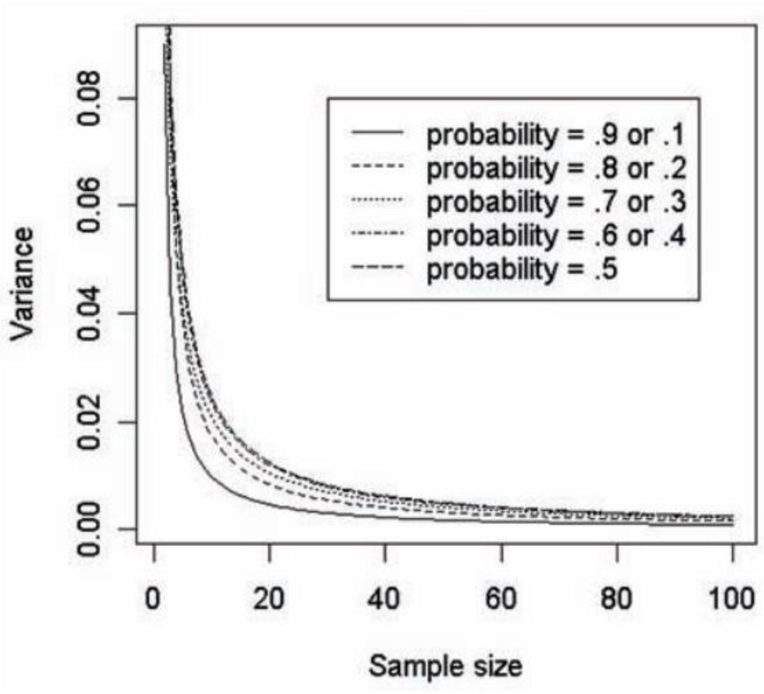

Figure 59. Decreasing variance in estimates of photo plot proportions as a function of the number of sample points. From: Frescino, T. S., Moisen, G. G., Megown, K. A., Nelson, V. J., Freeman, E. A., Patterson, P. L., … Menlove, J. (2009). Nevada Photo-Based Inventory Pilot (NPIP) photo sampling procedures (General Technical Report No. RMRS-GTR-222). Fort Collins, CO. https://doi.org/10.2737/RMRS-GTR-222

Constructing a Sample

There are two basic methods for *selecting* samples: random sampling and systematic sampling. In addition to selecting the sample, one must consider how to *organize* the sample, which is typically done either through clustering or stratification. These four tools can be combined into multiple layers of sample design. A typical CEO project for validating a land cover data product might use stratification to divide the study region according to classes, randomly locate plots (corresponding to a pixel in that data product) within those classes, and then systematically sample a regular grid of points within the plots[23].

Simple Random Sampling

In simple random sampling (SRS), all points/pixels (or whatever unit of analysis is selected) have an equal probability of being selected. A sample size is decided, and then samples are drawn randomly from the overall population. This approach has the advantage of being extremely simple and producing unbiased parameters that are straightforward to calculate. However, geographic balance is not certain with smaller sample sizes, and rare classes are unlikely to receive sufficient coverage unless the sample size is excessively large. SRS can be extremely useful when paired with other sampling methods to ensure geographic balance and coverage of rare classes.

[23] The astute reader may note that this could also be described as a stratified cluster sampling approach. However, this would only be true if the *points* were the sampling unit, and not the *plot*.

Systematic Sampling

Systematic sampling involves placing a grid (often uniform), mesh, or other network on the landscape, and using it to generate a sample. Systematic sampling has the advantage of providing excellent geographic balance, potentially producing better precision of estimates than random sampling, and being easy to implement in the field. However, it is not possible to calculate a truly unbiased estimate of the variance of population metrics[24] when using systematic sampling, and often they will tend to be too large. If there is substantial spatial autocorrelation present in the region of the study, the accuracy of estimates can also be reduced if the sample grid and correlation distances are similar. Like SRS, systematic sampling can be very useful when combined with other methods.

Clusters

Clustering is simply placing samples near to one another using a set of rules. These rules can be simple, for instance a block of 5 by 5 pixels could be selected, or they could be more complex and multi-stage rules that draw on random or systematic sampling methods. The resulting distribution of samples across the landscape is therefore clumpy and unlikely to be geographically balanced. It can also suffer from spatial autocorrelation issues, which can increase the standard error of parameter estimates. However, clustering allows for thorough sampling of neighboring locations, and can be very convenient and extremely cost-effective to use when collecting field data. Generally there is little reason to use cluster-based sampling approaches in CEO unless field data will be collected in tandem with photo interpretation work in CEO.

Stratification

Stratification is the division of the study area into groups, or *strata*, from which samples can be selected. Strata must be designed such that they are exhaustive (they include the entire study area) and exclusive (a sample unit cannot be part of more than one strata). Put simply, your strata should cover your entire map and they shouldn't overlap. Strata can be divisions based on geographic, political, or ecological properties of the landscape. Typically when working with LULC data, the strata would correspond to the classes of cover or use. Using strata ensures that rare classes on the landscape will receive a sufficient number of samples to make inferences about. This means that while stratification approaches are probability sampling methods, they are *unequal* probability methods at the highest level, though the probability of being selected inside a strata is usually equal. Depending on how it is employed, stratification may or may not produce a geographically balanced sample. Samples inside strata can be selected at random or systematically.

An excellent and more in-depth discussion of the relative strengths and weaknesses of each of these sample design strategies, as well as their combinations, can be found in the paper *Sampling designs for accuracy assessment of land cover*, by Stephen V. Stehman[25]. In the case of conducting an inventory or other point-based estimate, the use of simple random or systematic samples is common. For an accuracy assessment project, a stratified random sample will be an extremely good candidate approach, and they

---

[24] Stehman, S. V. (1992). Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing, 58(9), 1343–1350.

[25] Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. International Journal of Remote Sensing, 30(20), 5243–5272. https://doi.org/10.1080/01431160903131000

are very common to encounter in the research literature. If no rare classes exist on the landscape, stratification may not be necessary, and other concerns may drive the selection of the sampling design.

Exercise 5.2: Select Sampling Strategy

Introduction

Sampling and analysis are complex topics and in-depth coverage is beyond the scope of this manual. However, understanding the basics will help you to ensure that the data you collect in CEO is useful and can be used to answer the analytical questions and address your objectives. In this exercise, we will review sampling strategies used in the example projects. We will explain the rationale including information needs and constraints that guided these project. Understanding the benefits and drawbacks of different sampling approaches will inform your decisions regarding your own CEO projects.

Objectives

- Review example projects and sampling strategies
- Understand various options for sample strategies and how to choose one for you project

Part 4: Review sampling strategies in example projects

A. El Salvador Land use inventory – Sample design strategy and key points

Recall that the objectives of this project were to characterize historical trends in land use from 2000-2018 and potentially expand and continue the project into the future to monitor trends. Geographic evenness, simplicity, and repeatability were high priorities. To meet these needs, this project used a systematic grid of 2,662 0.49 ha plots distributed across the study area (see figure below). Each plot contained 25 points (sub-samples, each representing 4% of the plot). Recall that systematic sample designs don't allow for the unbiased estimation of variance, and the variance will typically be too large compared to the real value. This was deemed a reasonable trade-off to meet other objectives of the project.

1. Sample strategy - Systematic grid

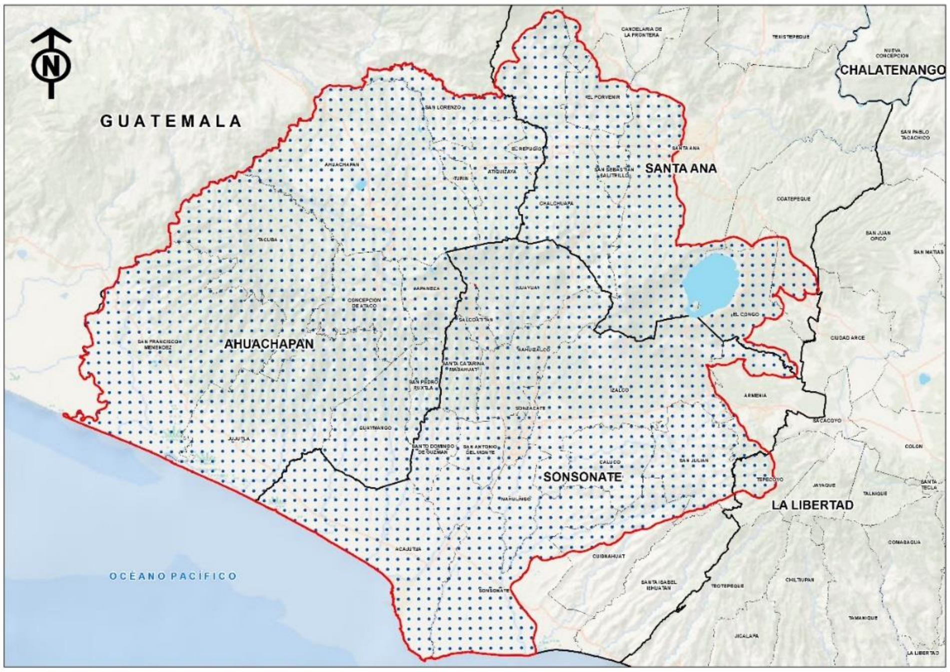

Figure 60. A systematic distribution of 2,662 plots was overlaid on the study area in El Salvador.

B. Ecuador LULC mapping– Sample design strategy and key points

Recall that the objective of this project was to produce a LULC change map for two years, and CEO would be used to perform an accuracy assessment of that change map. This means that the sampling approach would need to be able to sample rare classes with a small geographic extent with sufficient density to be able to provide a reasonable estimate of their area and the accuracy of their classification. To some extent, this means that geographic evenness will be a lower priority than maintaining a probability sample.

1. Sample Strategy – A stratified random sample was developed for the stable and changes classes.

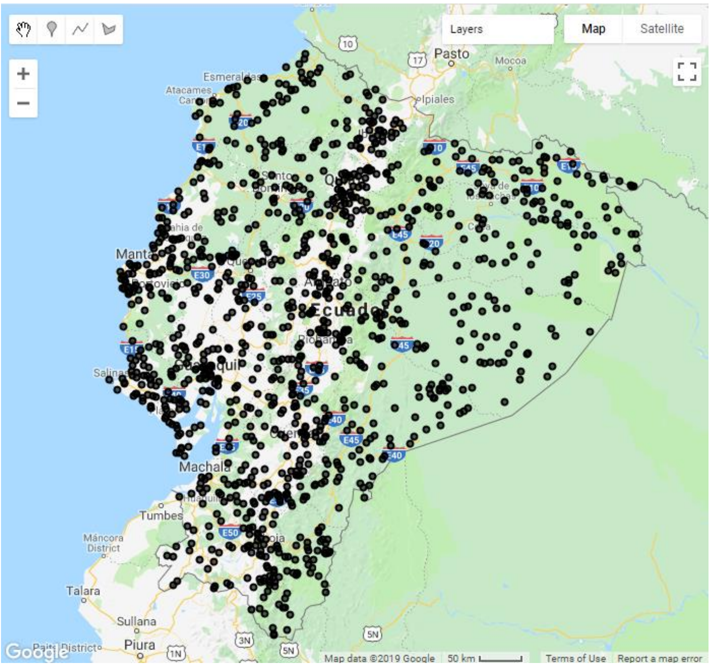

Figure 61. A stratified random sample of roughly 8,000 points was placed in the area of continental Ecuador.

Part 5: Check your knowledge

A. Review theoretical canopy cover examples below

1. Canopy cover inventory project:
   i. Objective – Use CEO to obtain a single inventory estimate of average tree canopy cover in a single forest. Recall that this approach won't produce a map, rather a statistical estimate of the tree canopy cover.
2. Canopy cover map validation project:
   i. Objective – Use CEO to obtain a reference dataset that will be used to validate a continuous pixel canopy cover map. This map covers the entire study area, and uses a percentile-based classification, with eleven classes: 0% canopy, 1-10% canopy, 11-20% canopy, etc.

B. Answer the questions related to sampling strategy

1. In the project seeking to estimate (inventory) canopy cover, what might be the simplest logical sampling strategy and why?
2. For the continuous canopy cover map validation project, what would be the simplest logical sampling strategy and why?

C. Review answers

1. For the canopy cover inventory project, a simple random sampling approach is likely to suffice. If other considerations are a factor that would favor the use of a systematic grid, that would also be an appropriate option.

2. For the validation project, a stratified random sampling approach would probably be best to ensure that the sample produces a sufficient number of samples in each class. In this case, the continuous canopy cover values could be stratified into classes (e.g., low, medium and high canopy cover).

Part 6: Plan the sampling strategy for your CEO project

A. Review project objectives

1. Is your project an inventory or mapping project?
2. How important is it for your project to produce well-defined estimates of uncertainty?
3. Will your sampling and measurement be conducted entirely in CEO or will plots also be visited in the field? If so, what resources are available for field sampling?

B. Consider your study area and phenomena of interest

1. Does your project include rare classes which are likely to be missed using simple random (SR) or systematic sampling? If so, do have existing data that you could use to stratify your sample?
2. Is your project area highly heterogeneous requiring additional measures to ensure geographic evenness?

C. Evaluate potential strategies

1. What are the advantages and disadvantages of the different potential sampling strategies?

5.4 Sample Size

This section contains information on setting the overall sample size (*N*) for accuracy assessment using a number of different approaches.

Approaches to setting sample size

There are two broad approaches to setting sample size that are presented here: comparing against a target accuracy and using a confidence interval. Within these approaches, several possible variations exist. Confidence interval approaches may be useful in cases of resource constraints that make collecting large samples difficult. They may also be useful if it is beneficial to think about accuracy in terms of the level of uncertainty associated with the result, rather than a comparative testing method's binary result (i.e., does or does not meet the target accuracy)[26].

For comparing against a target accuracy, the problem of sample size can be framed as setting the *minimum* "meaningful difference" in map accuracy and calculating sample size based on that. In other words, asking the question "what is the smallest difference between the target outcome and the result of the work that you wish to be able to detect?" However, the minimum detectable difference may not be an intuitive concept. Suppose that you've got a map, and at the end of the map validation process, it appears to have an accuracy of 0.87, and the goal for the project was 0.85. Given those values can you actually say that it is better or different than the goal? If your sample size is too small, you may not be

[26] Foody, G. M. (2009). Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing, 30(20), 5273–5291. https://doi.org/10.1080/01431160903130937

able to say that you've done better than your goal accuracy, because the minimum difference a small sample can detect isn't very large, maybe not even 5%. In that instance, you can't be statistically confident that the real accuracy of the map is equal to or better than 85%. It might even be lower. However, if you have a sample large enough to detect a 1% difference, you now have enough statistical power to say with confidence that the accuracy is better than the target, because you can confidently tell numbers apart when they differ by 1% or more.

It also may not make intuitive sense to talk about "being able to tell numbers apart", because we can look at them and see that they are different. But the reality is that the composition of your sample will determine the accuracy numbers that you get as a result, and you can't be sure that you'd get the same number if you had selected a slightly different set of points. In fact, with a small sample, you almost certainly would not get the same number, and it might vary substantially. Having a larger sample reduces that uncertainty and allows you to "tell numbers apart" more effectively. This is what the "minimum detectable difference" means in practice.

In comparison, when using a confidence interval, the approach may be framed as considering the maximum allowable range of deviation for the result; or, restated, what is the window around the target you'd like the value to fall inside? Here we are thinking more about how big the acceptable window of uncertainty for our results should be. Are you okay with a result where you can only say that it is ± 5%? Or do you need to be able to say you are confident to within 1%?

Both of these approaches are available for use when thinking about accuracy assessment. However, when conducting point-based inventories it unlikely that a target proportion will exist to compare against, and thus a confidence interval based approach (which will focus on the uncertainty of the estimates) is the more appropriate choice.

Factors influencing Sample Size, Comparing against a Target

When determining the size of a sample for assessing the accuracy of a land cover product, there are four factors that need to be considered:

- Expected accuracy of the product (P_0)
- Precision of detecting differences from this accuracy (minimum detectable difference, δ)
- Tolerance of Type I error (alpha, α)
- Tolerance of Type II error (beta, ß)

They are related to sample size through this equation[27]:

[27] Foody, G. M. (2009). Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing, 30(20), 5273–5291. http://doi.org/10.1080/01431160903130937

$$n' = \left[\frac{Z_{\alpha}\sqrt{P_0(1-P_0)} + Z_{\beta}\sqrt{P_1(1-P_1)}}{\delta}\right]^2$$

Equation 1. The sample size determination equation.

The following equation can then be applied:

$$n = \frac{n'}{4}\left(1 + \frac{2}{n'\delta}\right)^2$$

Equation 2.The continuity correction.

Equation 2 provides a continuity correction to the previous estimate of N. We apply Equation 2, because Equation 1 treats *n* as it if it were distributed continuously (as if samples existed like decimal numbers), but it is really discrete (we can only select whole samples), and this leads to a slight underestimation of the number of samples needed.

Each of the parameters influences sample size:

- As expected accuracy decreases, sample size goes up;
- As minimum detectable difference decreases, sample size goes up;
- As tolerance for error decreases, sample size goes up.

Stated in reverse:

- As expected accuracy increases, sample size goes down;
- As minimum detectable difference increases, sample size goes down;
- As tolerance for error increases, sample size goes down.

Expected accuracy (P_0) is simply the overall accuracy of the product, based on past experience. This can also be thought of as a target accuracy. The threshold of acceptable accuracy when making LULC maps is often set as 0.85, or 85%. It may not make immediate sense why altering this number would change the sample size, however the sample size equations contain the quantity $P_0(1-P_0)$, and the largest value for that quantity occurs at $P_0 = 0.5$. It also makes some intuitive sense that when the expected accuracy is close to 1 or 0, it will be easier to tell if actual results differ.

The minimum detectable difference (δ) is the size of difference from P_0 that we wish to be able to reliably detect using statistical measures, and it also sets a threshold that prevents the sample size from being set unnecessarily large. Very large sample sizes often allow for arbitrarily small differences to be declared "statistically significant" when they are of no practical significance. It's important to remember that when we perform an accuracy assessment, we are making an *estimate* of the accuracy, and the *true* accuracy may be some distance away from the estimate. Thus setting δ establishes how big the gap between the estimate of the accuracy and the target or expected accuracy we wish to be able to talk about with certainty should be.

Tolerance for error (α, ß) describes how willing we are to either declare false results true, or miss true results. It is related to the concept of statistical power (Power = 1 - ß), the ability to reliably find the truth.

A Type I error occurs when we state something is true (the map product exceeds 0.85 overall accuracy) when it is in fact false. It is a false positive. Our tolerance for Type I errors is indicated by alpha, α. It is typical that α is set to 0.05, meaning that we expect to declare false results true 5% of the time. However, do not rely on convention when setting your parameters, unless it is necessary to do so.

A Type II error occurs when we state something is false (the map product did not exceed the 0.85 overall accuracy threshold) when it is in fact true. It is a false negative. Our tolerance for Type II errors is indicated by beta, ß. It is much less common to explicitly set ß, but a value of 0.20 is generally considered acceptable. This is due to the fact that false negatives in many experiments are considered less problematic than false positives. However, when it comes to LULC mapping and accuracy assessment, the amount of resources expended means that false negatives have the potential to be very costly, and so ß should be set accordingly.

To determine Z_α, or Z_β first set α or β, and then look up the closest value in a z-score table. There are many resources available online to do this. Searching for a "z score calculator" will find an appropriate reference. Most statistical textbooks will also contain a table of z-scores, typically in the back of the book.

Table 5. A small selection of commonly used alpha and beta values, with the related z score. Note that in many equations you will need to divide alpha or beta by two to find the right z score to use.

α, β	α/2, β/2	*z*
0.2	0.1	1.28
0.1	0.05	1.64
0.05	0.025	1.96

The table below contains some examples of sample sizes based on different parameter settings, calculated using the equations above (plus the continuity correction factor). Note that changing the minimum detectable difference often has a very strong impact on sample size, much more so than reducing the acceptability of Type II errors. This is because δ is a decimal number in the denominator of the equation.

Table 6.Example sample sizes, calculated using equations 1 and 2. ***P_0*** *is the target accuracy value,* ***δ*** *is the desired detectable difference,* ***α*** *represents our tolerance for Type I error,* ***ß*** *represents our tolerance for Type II error, and* ***n*** *is the number of samples determined by the calculation.*

P_0	δ	α	ß	*n*
0.85	0.05	0.05	0.20	490
0.85	0.05	0.05	0.05	683
0.85	0.025	0.05	0.05	2948
0.90	0.025	0.05	0.05	1991
0.90	0.01	0.05	0.05	13300
0.95	0.01	0.05	0.05	6681

Factors influencing Sample Size, Setting a Confidence Interval

When setting size using confidence interval, you will need to know the:

- estimated proportion of cover, or expected accuracy (*p*);
- acceptable margin of error (*e*); and
- size of confidence interval ($Z_{\alpha/2}$).

$$n = \frac{Z^2_{\alpha/2} \times \ p(1-p)}{e^2}$$

Equation 3. The equation for setting a sample size using the confidence interval information.

The estimated proportion of the cover (*p*) is your best guess, based on previous research or other data, of the amount of the cover of interest you are attempting to measure. Note, this seems to suggest a binary classification. If you have a particular class that you are more interested in than anything else, then you can set it to *p* and all other cover then becomes 1 – *p*. If you are interested in multiple LULC classes equally, you may wish to set the sample size you'd need for each class, and then add them together to get your total sample size. If you have no idea what percentage of cover to expect, set *p* to 0.5, as this is the most conservative approach you can take, and will set a larger sample size.

The acceptable margin of error (*e*) can be thought of as the maximum acceptable distance you'd like the true population value to be from your estimate. It is commonly represented as *Result ± margin of error*. Larger sample sizes lead to smaller margins of error. Or, put in reverse, smaller margins of error require larger sample sizes.

The size of the confidence interval ($Z_{\alpha/2}$) is how certain you want to be that the true value for the population lies within the margin of error of your results. This is a slightly difficult concept, but put simply, a 95% confidence interval means you can assume that if you drew 100 different samples, 95 of them would produce a cover estimate that was within your margin of error of the true value. The notation here is related to Type I error, as discussed above, with the confidence interval being equal to 1 – α. To determine $Z_{\alpha/2}$, determine α, divide it by two, and then look up the closet value in a z-score table. A confidence interval of 95% produces a $Z_{\alpha/2}$ of 1.96. Narrowing your confidence interval (confidence interval = 99%) will produce a larger $Z_{\alpha/2}$, and thus require a larger sample size.

*Table 7. Example sample sizes, calculated using equation 3. In this table **p** is the target accuracy value, **CI** is the confidence interval**,** $Z_{\alpha/2}$ is the z score that results from our confidence interval, **e** is our margin of error, and **n** is the number of samples determined by the calculation.). Note that changing the minimum detectable difference often has a very strong impact on sample size, much more so than reducing the acceptability of Type II errors. This is because δ is a decimal number in the denominator of the equation.*

p	CI	$Z_{\alpha/2}$	*E*	*n*
0.70	0.95	1.96	0.01	8067
0.70	0.95	1.96	0.03	896
0.70	0.95	1.96	0.05	323
0.70	0.90	1.64	0.01	5648
0.70	0.99	2.56	0.05	551
0.50	0.95	1.96	0.05	384

Similar to in the other approach, here the margin of error has a very large impact on sample size, because it is also the denominator of the sample size equation. The margin of error and minimum detectable difference are closely related statistical concepts, and thus influence the sample size in similar ways.

Setting Sample Size

Setting sample size is an exercise in trade-offs between desired levels of accuracy, statistical power, and available resources. Being able to detect small changes in outcome (such as improvements between two different methods of producing a map, or if a rare class changed in area) with high certainty requires large sample sizes. The sample size may also need to be increased if it is expected that a certain amount of the samples collected will need to be discarded (perhaps due to lack of imagery in the area) or are otherwise not reliable.

Dividing *n* into classes for use in a sample design other than Simple Random Sampling (using stratification, for instance) requires taking other considerations into account, such as number of classes, rarity of classes, classes of interest, etc. If one class in a stratification is of particular interest (or if each class has its own accuracy target), the above equations can be used to set the class size. In instances where there is only a small number of classes, there are also optimization methods for selecting class sample sizes, based on minimizing the variance of parameter estimates[28].

It is often the case that even well-funded projects are not able to collect as many samples as they would like for the purposes of accuracy assessment[29]. When setting sample size, it is important to consider the primary goals of the project and the resources available for data collection. Adopting a flexible sampling design that allows for the easy addition of more data (often a sample design that makes use of random sampling will allow this) is often wise, in that it allows for increased accuracy if more resources become available or the work goes more quickly than expected. It is also often a good idea to set up a series of small pilot projects in order to allow for an accurate estimation of how many samples your team can accurately collect in a given time.

Exercise 5.3: Select a Sample Size

Introduction

Now that you have planned your response and sample designs, you need to determine the required sample size to meet the project objectives. Approaches to determining the required sample size will depend on the application as well as multiple parameters related to objectives. Depending on how the results of your CEO project will be used, the required level of certainty in the final estimates or map may

[28] Wagner, J. E., & Stehman, S. V. (2015). Optimizing sample size allocation to strata for estimating area and map accuracy. Remote Sensing of Environment, 168, 126–133. http://doi.org/10.1016/j.rse.2015.06.027

[29] Wickham, J., Stehman, S. V., Gass, L., Dewitz, J. A., Sorenson, D. G., Granneman, B. J., … Baer, L. A. (2017). Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD). Remote Sensing of Environment, 191, 328–341. http://doi.org/10.1016/j.rse.2016.12.026

differ. Likewise, tolerance for different kinds of errors in your product are determined by information needs. In this exercise, you will practice methods for determining the required sample size.

Objectives

- Understand the importance of the planning the sample size;
- Know how to apply two different strategies for determining sample size based on objectives;
- Practice estimating samples sizes for 1) an inventory sample and 2) a map validation sample;
- Understand how to plan your own sampling.

Part 1: Review two strategies for planning sample size

A. Comparing against a target accuracy

1. When comparing the results of a map validation against a target accuracy, you need to determine several things:
   i. What you expect or would like the accuracy of the product to be (P_0); a value often set for this parameter is 75-85%.
   ii. The precision in detecting differences from this accuracy, the minimum detectable or meaningful difference (δ); conventional values for this parameter might be 3-5%.
   iii. How tolerant you are of making Type I errors, or false positives (alpha, α); alpha is typically set at 5% or 0.05. In the map example, a Type 1 error in this case would mean that you conclude the map meets the target accuracy when in reality, it does not.
   iv. How tolerant you are of making Type II errors, or false negatives (beta, ß); beta is often set at 20% or 0.2, but it may be wise in a map validation context to have a lower tolerance for Type II error. In the map example, a Type II error in this case would mean that you conclude the map *does not* meet the target accuracy when in reality it does.

B. Using the confidence interval

1. When determining the size of your sample using a confidence interval approach, as for the estimate of a canopy cover inventory, you need to decide the following details:
   i. What you expect the proportion of cover to be for your target class (*p*); if you have no idea what it might realistically be, set *p* to 0.5.
   ii. What you think is an acceptable margin of error (*e*) for your results; conventional values here are often around 3%.
   iii. What size of confidence interval (this determines $Z_{\alpha/2}$) you'd like to set; a typical value here is 95%.

Part 2: Check your knowledge

A. Point-inventory canopy cover estimation project example

1. Which of the two methods would you use for determining the sample size in this case and why?

B. Canopy cover map validation project example

1. Which of the two methods would you use for determining the sample size in this case and why?

C. Interpret parameters

1. In the example of validating a final canopy cover map based on target accuracy, explain the implications of the following set of parameters:

   i. Target accuracy (P_0) = 0.8

   ii. Minimum detectable difference (δ) = 0.1

   iii. Tolerance of Type I error (α) = 0.05

   iv. Tolerance of Type II error (β) = 0.05

D. Target accuracy methods – Interpretation

1. In the case of this application, which explanation of the minimum detectable difference is correct?

   i. The minimum detectable difference is the smallest gap between two numbers your sample size allows you to tell apart.

   ii. The minimum detectable difference is the size of the difference between your map and the target accuracy that you expect to encounter.

2. Which is the correct explanation of the practical implications of committing a Type I error?

   i. The map is uninterpretable.

   ii. The map is deemed to be sufficiently accurate when in reality it is not.

   iii. The map result is not within margin of error.

3. Which is the correct explanation of the practical implications of committing a Type II error?

   i. The map is deemed to be *inaccurate* and gets thrown out when in reality it is not.

   ii. The map gets published.

   iii. The map accuracy is not within the minimum detectable difference.

E. Knowledge check – Answers

- A1 – Answer: The confidence interval approach is most appropriate here, as we want to know our estimates with a specific level of certainty. Sample size based on targets is not relevant.
- B1 – Answer: The target accuracy approach is most appropriate for setting sample size to validate a map.
- C1 i-iv– Answer: The target overall accuracy of the map is 80% and the smallest difference worth detecting is 10%. The probability tolerance for falsely accepting the map is 5% while the probability tolerance for falsely rejecting the map is also 5%.
- D1 – Answer: i. is correct. The minimum detectable difference sets the lower limit on your ability to tell two accuracies or proportions apart.
- D2 – Answer: ii. is correct. A Type I error would result in the map being deemed accurate when it’s not.
- D3 – Answer: i. is correct. A Type II error would result in the map being rejected when in fact it is sufficiently accurate.

Part 3: Practice – Sample size based on target accuracy

A. Calculate sample size – Numerical example

1. Parameters
   i. Target accuracy ***P_0 = 0.8***
   ii. Minimum detectable difference (δ) = 0.1 which corresponds to ***P_1 = 0.9***
   iii. Tolerance of Type I error (α) = 0.05 which corresponds to a z-score of 1.96 for a two-sided test. ***z_α = 1.96***
   iv. Tolerance of Type II error (ß) = 0.05 which corresponds to a z-score of 1.96 for a two-sided test. ***z_β = 1.96***
   v. Step 1. Plug the parameters above in the equation below and solve for *n'* to derive an initial estimate of the minimum require sample size

$$n' = \left[\frac{Z_\alpha\sqrt{P_0(1-P_0)} + Z_\beta\sqrt{P_1(1-P_1)}}{\delta}\right]^2$$

2. Plug the values and your estimated minimum sample size (***n'***) into the following equation to apply the continuity correction and get a final estimated sample size.

$$n = \frac{n'}{4}\left(1 + \frac{2}{n'\delta}\right)^2$$

B. Numerical example – Answer

1. Use the steps in the table below to check your work

Equation 1 Plug in values	$$n' = \left[\frac{1.96\sqrt{0.8(1-0.8)} + 1.96\sqrt{0.9(1-0.9)}}{0.1}\right]^2$$
Equation 1 Solve	$$n' = \left[\frac{1.372}{0.1}\right]^2$$
Equation 1 Estimate of minimum required sample size	$$n' = 188.24$$
Equation 2 Plug in values	$$n = \frac{188.24'}{4}\left(1 + \sqrt{1 + \frac{2}{188.24 \times 0.1}}\right)^2$$
Equation 2 Solve	*n* = 198.11 or 198

C. Calculate sample size to assess accuracy of a canopy cover map

1. Use the equations above and the same method as before to calculate the sample size required to assess a binary canopy cover map – Use the values below.
   - i. Target accuracy $\boldsymbol{P_0 = 0.8}$
   - ii. Minimum detectable difference (δ) = 0.1, which corresponds to $\boldsymbol{P_1 = 0.9}$
   - iii. Tolerance of Type I error (α) = 0.10 which corresponds to a z-score of 1.64 for a two-sided test. $\boldsymbol{z_\alpha = 1.64}$
   - iv. Tolerance of Type II error (ß) = 0.10 which corresponds to a z-score of 1.64 for a two-sided test. $\boldsymbol{z_\beta = 1.64}$

D. Canopy cover example – Answer

1. Use the steps in the table below to check your work
   - i. What parameters changed? How did that affect the required sample size and why? Does this make sense?

Equation 1

Plug in values

$$n' = \left[\frac{1.64\sqrt{0.8(1-0.8)} + 1.64\sqrt{0.9(1-0.9)}}{0.1}\right]^2$$

Equation 1

Estimate of minimum required sample size

$$n' = 131.79$$

Equation 2

Plug in values

$$n = \frac{131.79}{4}\left(1 + \sqrt{1 + \frac{2}{131.79 \times 0.1}}\right)^2$$

Equation 2

Solve

n = 141.61 or 142

Part 4: Practice – Using the Confidence Interval

A. Calculate sample size – Numerical example

1. Parameters
   i. Size of confidence Interval – 95% confidence interval corresponds to ***$Z_{\alpha/2}$ = 1.96***
   ii. Estimated proportion of class cover ***p = 0.5***
   iii. Acceptable margin of error ***e = 0.05***
   iv. Plug the parameters above in the equation below and solve for n to derive an estimate of the minimum require sample size

$$n = \frac{Z_{\alpha/2}^2 \times\ p(1-p)}{e^2}$$

B. Numerical example – Answer

1. Use the steps in the table below to check your work

Equation 3

Plug in values

$$n = \frac{1.96^2 \times 0.5(1-0.5)}{0.05^2}$$

Equation 3

$$n = 384.16 \text{ or } 384$$

Solve for n

C. Calculate sample size – Estimate canopy cover in a dense tropical forest

1. Use the same equation to estimate the minimum required sample size to estimate canopy cover in a region of dense tropical forest. In this area, canopy cover is expected to be quite high, 70%. For this calculation use 95% confidence intervals and a 5% margin of error.
   i. Note that the only parameter that has changed from the previous example is the expected cover (***p***) therefore, how do you expect the sample size to change?

$$n = \frac{Z_{\alpha/2}^2 \times p(1-p)}{e^2}$$

D. Tropical canopy cover example – Answer

1. Use the steps in the table below to check your work
   i. How was the estimated sample size affected by the change in ***p*** and does this make sense?

Equation 3

Plug in values

$$n = \frac{1.96^2 \times 0.7(1-0.7)}{0.05^2}$$

Equation 3

$$n = 322.69 \text{ or } 322$$

Solve for n

Part 5: Plan your sample size

A. Guided questions

1. Which sample size estimation approach suits the nature of your project and why?
2. What level of accuracy or cover do you expect to have at the end of your project? You may need to look at previous projects in your area to make an estimate.
3. How tolerant are you of error? This will inform setting alpha and beta, or the confidence interval and margin of error.

4. What sample size do these parameters suggest you use? Is this a realistic sample to collect? Will you need to make it smaller to meet budgetary concerns, or bigger to account for missing or low quality imagery?

6. Data Collection and Quality Assurance and Control

Overview

When you collect data, you want to ensure that the data you end up with is as high quality as possible, free from biases and errors. Part of how this is achieved is through the use of Quality Assurance and Quality Control (QA/QC). These are defined by the European Joint Research Center[30]:

> *Quality assurance (QA) is a set of approaches which is consciously applied and, when taken together, tends to lead to a satisfactory outcome for a particular process….*
>
> *A quality control (or check) is a clearly specified task that scrutinizes all, or a sample, of the items issuing better during, or even at the end of, the… process in order to ensure that the final product is of satisfactory quality. The scrutiny involves review, inspection or quantitative measurement, against well-defined pass/fail criteria….*

In other words, *quality assurance* is a system of practices that should be employed through the entire process of planning a project and collecting data to ensure a good result, and part of that process should be *quality control* or *checks*, that actively verify the status of data as good or bad.

Because photo-interpreted data is often treated as "truth", and used as the gold standard against which other data are compared, PI data need to be consistently and rigorously checked for errors to prevent the creation of erroneous and misleading results. It is often the case that good QA/QC work is what separates a good result from one that is unusable.

Training as QA/QC

Photo interpretation data should be consistent both *across* and *within* interpreters. This means: 1) given the same image, two different interpreters will return similar results and 2) given similar images over time, the same interpreter will return similar results. This can be difficult to achieve, but there are methods to ensure good results. In particular, a robust training process for photo interpreters is a very important component of the QA/QC process.

Training should begin with detailed discussion and review of the classification system, response design, and photo interpretation key. The questions raised by interpreters when first working with the key may indicate areas where the key or classification system need revision.

Following this discussion, introduction training data sets should be created, analyzed in CEO, compiled, and then analyzed to assess agreement between interpreters. This can be conducted through very simple methods, such as using Microsoft Excel and manually determining agreement, or it can be done through more sophisticated statistical methods that measure concordance, such as iota (ι), a multivariate measure

---

[30] Kapnias, D., Milenov, P., & Kay, S. (2008). *Guidelines for Best Practice and Quality Checking of Ortho Imagery* (No. EUR 23638 EN-2008). Ispra, Italy. https://doi.org/10.2788/36028

of agreement between raters or interpreters (essentially a generalized form of kappa)[31,32] or interclass correlation coefficients (ICC)[33].

Thereafter, training sessions can begin with focus placed on the plots with the highest rates of disagreement or difficulty in interpreting classes. After each session's data is processed and metrics calculated, the data should be reviewed to find plots that show the most disagreement. A slide can then be created for each difficult plot, showing an image of the plot and how it was interpreted by each team member or group, along with a set of brief discussion points and tips on how to improve interpretation of similar plots.

It is also a good idea to document these issues. Consider creating a brief summary detailing the overall performance of the photo interpretation team, what types of errors seem to be most common, and how those errors might be addressed going forward.

This kind of collaborative training process should be followed up with regular communication about problem plots and other difficulties among the photo interpretation team.

QA/QC Tools

Self-reviewing Results

At the ends and beginnings of sessions, participating interpreters can and should rapidly review their past work. This can allow them to look at everything (or a percentage) of a given class they have labelled and ensure that they are being consistent in their labelling. This also give the interpreters an opportunity to spot simple mis-click errors. In addition to finding unintentional errors, repeated exposure to many examples of the same LULC classes or landscape elements helps to build the interpreter's "internal key" and helps them become quicker and more consistent in the work.

Duplication of Interpretation

If sufficient resources are available, images can be interpreted by multiple individuals or teams. This can allow for the calculation of concordance of interpretations, as well as indicate if there are particular landscape types that consistently produce variable interpretations.

However, duplication of effort does reduce the total sample size that can be collected, and a system must be in place to deal with conflicting interpretations, especially if no clear "winner" exists.

Cross Validation

The cross validation process tests an interpreters results against another set of data that is considered accurate and reliable. This allows for an external standard that interpreters can rely on to maintain a consistent approach.

---

[31] A. J. Conger, Integration and generalisation of kappas for multiple raters., Psychological Bulletin 88 (1980) 322–328

[32] H. Janson, U. Olsson, A measure of agreement for interval or nominal multivariate observations., Educational and Psychological Measurement37861 (2001) 277–289

[33] J. J. Bartko, The intraclass correlation coefficient as a measure of reliability., Psychological Reports 19 (1966) 3–11.

At the beginning and end of an interpretation session, teams can work with sets of imagery that have already been classified or interpreted by an expert (with ground-truthing). These examples might have been part of another related project, created as part of key development, or made specifically for cross validation.

Each interpreter can classify some of these images and report their reasoning. If there are any discrepancies with the expert or ground-truthed interpretation, they can then be resolved. Throughout the session, interpreters can share thoughts with each other and compare results on images as they work.

Formal Review of Results

The primary objective of this type of plot review is to provide ongoing calibration of the interpreter's labeling throughout the photo interpretation process. There are two types of plot reviews that can be used, "hot checks" and "cold checks".

Hot Checks

In CEO, it is possible to mark plots that should be reviewed due to low confidence, unusual circumstances, or other situations that might warrant review. These marked plots should be further reviewed by other interpreters, the supervising analyst, and discussed during review meetings. Once reviewed, labels that are deemed to be incorrect on these plots should be adjusted by the interpreter.

A good way to handle "hot checks" is to include them in a PowerPoint presentation, each with their own slide. These slides should include three main elements: 1) the interpreter's notes explaining the question or why the plot was flagged for review; 2) the response prepared by the supervising analyst; and 3) the necessary imagery (e.g., NAIP, Google Earth, Landsat, Digital Globe, etc.) to facilitate discussion among the group as a whole.

Cold Checks

These are plots that are randomly selected from the data produced by interpreters. The calls made by the interpreters are reviewed by a supervising analyst or group of interpreters meeting together. In most cases, errors found in the cold checks should not be adjusted by the interpreter. Only under rare circumstances will the interpreter be asked to revisit and correct issues found. This is generally only required if the review uncovers a systemic error that had previously escaped attention.

The frequency of cold checks can typically decrease as interpreters gain familiarity with the project and specific area of interpretative work. The following numbers are recommendations used on sampling intervals at various stages of the project. Note: these may be adjusted for each interpreter, but this serves as a starting point.

Table 8. Interval at which to duplicate samples for cold checks, for various sizes of sample sets.

Plots	Sampling Interval
1 to 50	5
51 to 200	10
201 to 400	20
401 and on	30

Cold Check plots should be reviewed by supervising analysts or the convened team of interpreters and information about them is recorded. If a plot is found to contain an error, information about it is included in the PowerPoint with the Hot Checks and is discussed during the review meetings. However, if the error needs to be corrected immediately or if it is specific for that interpreter, then the interpreter is contacted directly.

Repeatability

The goal of these checks is to provide a way to check for repeatability and consistency among and between interpreters. These plots are completed after the interpreter finishes up their main set of plots.

Self-Checks

A randomly selected subset of an interpreters' dataset is reviewed and interpreted again by the same interpreter. These will typically be plots from previously completed projects and are considered blind checks. Approximately 2.5% of plots can be used for self-checks.

Cross-Checks

A randomly selected subset of an interpreters' dataset is reviewed and interpreted by a different interpreter. These will typically be assigned concurrently to several interpreters and are considered blind checks. Approximately 2.5% of plots can be used for cross-checks; generally these should be different from those used for self-checks.

Upon completion of a project, the supervising analyst compares the Self Checks and Cross Checks to the main sets of plots and creates a summary of the results. Information from this summary should be included with the reporting of the results, to provide some estimate of the uncertainty associated with the photo interpretation.

Exercise 6.1: Quality Assurance and Quality Control

Introduction

The final CEO project planning stage covered in this manual is the development of a QA/QC process to ensure the quality of the data you collect. In this exercise you will review methods and concepts, check your knowledge and begin the planning process of developing a QA/QC approach for your project.

Part 1: Review Quality Assurance and Quality Control

A. What are QA and QC?

1. Quality Assurance – This is the process that is established at the beginning of a project and followed throughout to ensure data quality.
   i. Example – *In the Ecuador LULC mapping project, QA began with the development of the photo interpretation key. Examples and descriptions were included in the key as well as additional expert knowledge to help train interpreters to identify the features of interest.*

Additionally, multiple training projects were completed prior to collecting data for the main project. The use of point sub-labels to characterize plot labels and clear definitions of plot-classes were also important components of QA as illustrated in Figure 62 *and* Figure 63*, below.*

2. Quality Control – These are the discrete tasks that are executed to ensure and maintain data quality.
   i. Example - *In the Ecuador mapping and validation project, a sample of plots were interpreted multiple times independently by two different interpreters and the resulting labels were compared statistically. Through this process, problematic classes were identified and additional training and practice was done to improve interpreter skill. Disagreements were resolved by a supervising interpreter who assisted in training interpreters.*

B. Training as QA/QC

1. Image interpretation training and practice is important part of the QA/QC process. It is important that the different interpreters working on a project are consistent in labelling points and plots. Training projects should involve the multiple interpreters independently interpreting the same of plots. Agreement should then be analyzed and disagreements resolved. This is important in helping interpreters to become more accurate and consistent.
2. Example - *In the Ecuador LULC mapping example project, several iterations of training projects were completed at the start of the project. After each, agreement was analyzed and disagreements were identified and reviewed.*
   i. In some cases, disagreement among interpreters didn't change the final plot label. In the example below, the two interpreters disagreed on the extent of cover between crops (yellow) and roads and lots (brown). The threshold definition for crops is 50% in this system so these disagreements did not affect the plot label.

Figure 62. Example plots showing apparent disagreement in labelling. The left set of plot labels has a much higher amount of impervious cover than the right labelling (36% vs 16%). However, both plots cross the necessary threshold for this plot to be labelled as "crops" (64% and 84%, respectively), and so there is agreement on the final class label.

ii. In other cases, disagreement resulted in the final plot label differing. In the example below, two interpreters disagreed on the shrub (brown) and tree (dark green) cover. In this classification system, forest plots were defined as > 30% tree cover over the 5 meter height threshold. This plot was encountered during the training process, flagged as a disagreement and revisited by interpreters to potentially resolve future, similar issues.

Interpreter 1

Plot label - Forest

Tree – 32%

Shrub – 52%

Herbaceous – 16%

Interpreter 2

Plot label – Shrubland

Tree – 24%

Shrub – 60%

Herbaceous -16%

Figure 63. Example plots showing disagreement in labelling. The left plot has enough tree cover labelled to cross the threshold and be labelled as forest (32%), which is a dominant class over shrubland. However, the set of labels on the right has less than 30% tree cover, leading the plot to be labelled as shrubland, rather than forest. This leads to a disagreement in final class.

C. Additional QA options

1. Self-review - At the end and/or beginning of each interpretation work sessions, interpreters review their past work.
2. Duplication - Images can be interpreted by multiple individuals or teams. This requires more resources as well as a plan to settle 'disagreements'.
3. Cross validation – Using a reference dataset considered to be accurate, interpreters' results are compared and evaluated.

D. Formal Checks

1. Hot and Cold Checks – Hot checks involve interpreters 'flagging' low confidence plots for review by a supervising interpreter. Cold checks involve a random selection of interpreted plots being pulled and reviewed by a supervisor.
2. Cross and Self Checks – A self-check involves randomly selecting a subset of interpreted data and having the same interpreter review and interpret again to check consistency. Cross-checks involve randomly selecting a subset of an interpreted dataset and having a *different* interpreter review and interpret them.

Part 2: Check Knowledge of QA/QC

A. Mark the statements below as True or False

1. Developing a detailed and comprehensive photo-interpretation key is not an important component of QA.
2. Using a field-validated reference dataset to evaluate CEO interpretations is an example of cross-validation.
3. Duplication requires additional interpreter resources and reduces the sample size that can be interpreted.
4. When completing training projects for QA, the interpretation key is the primary piece of training material.

B. Answers

- A1- False
- A2 - True
- A3 - True
- A4 - True

Part 3: Planning Your QA/QC

A. Resources

1. Do you have access to an expert on the landscapes of your project? Do you need one? Or are your classes sufficiently easy to identify that one is not necessary?
2. Do you have access to ground-truthed reference data for cross-validation?

B. Image interpretation key

i. Does an interpretation key already exist and if not, what resources can you use to help develop one?

C. Interpretation team

i. How many people do you have for your image interpretation team?

ii. For how long?

D. Interpreter training and evaluation

1. What resources do you have for completing and evaluating training projects? What is the timespan available?
2. How will you facilitate collaboration and working together to achieve agreement?
3. For evaluating and measuring agreement, are simple methods sufficient or are more complex statistical comparisons required?

E. Formal Checks

1. What can you implement and how will you do so?

7. Summary and conclusion

After having worked through this manual, you should now have a sense of what questions you need to ask and what resources you need to gather to begin planning your project. You will also likely have had moments where you revised your thoughts about previous steps as you encountered new information. Recall the diagram of the planning process, and how each new building block relates back to the project goals, needs, and constraints. While the project planning process creates many opportunities to get lost in small details, the primary goals of the project should remain foremost in your mind. Additional complexity can sometimes be appealing, especially if it seems to make your process look more "professional" or "scientific". However, the simplest and most straightforward process that meets your goals is very often the best approach to take.

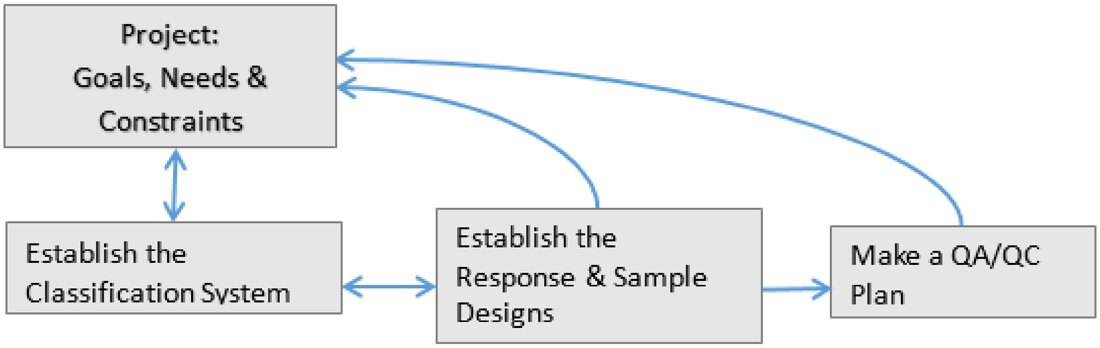

Figure 64. Iterative Collect Earth Online project planning. At each stage, decisions are made and evaluated in terms of goals, needs and constraints. In some cases, goals may need to be revised.

This process is inherently iterative and can require a significant amount of careful thought and research to produce an internally consistent, well-planned project. Most of the skills presented in this text are relevant to nearly any project or process focused on land cover and land use. The final section of this text presents a brief summary of the essential points of each chapter. In the appendix that follows it, you will find a worksheet to help guide you through the planning process, and which will provide a centralized location for information related to your project and its needs, goals, and resources.

7.1: Uses for Collect Earth Online

CEO can be used for a variety of projects related to remote sensing, map making, and resource inventories. The most likely uses for CEO are:

- Creating an inventory of land cover, land use (LULC) and LULC change, either through point-based sampling, or via modeling.
- Producing reference data for map validation
- Any other data that relies on photo interpretation

7.2: Project Planning

CEO is a powerful and flexible tool, but using it effectively requires significant forethought and planning. Project planning is essential to success when working with CEO, as in most other contexts. The first step in that process is defining the project's goal, needs, resources, and constraints.

- What are the key objectives of the project? What products will it produce? What questions will they provide answers to?
- What resources are necessary to achieve these objectives? Does the imagery need to be from a particular window of time? What sorts of other data might be useful?
- Of the needed data, which are already available, and which will need to be obtained?
- How much staff might be required, with what level of expertise?
- What budget is available?

7.3: Land Use and Land Cover

After defining your project's objectives, you will need to create clear definitions for the landscape elements that will need to be identified to meet your goals. You'll also need to determine if you are interested in land cover, land use, or both. Recall that 'land cover' refers to the biophysical make-up of the landscape; rocks, tress, plants, water, and structures. In contrast, 'land use' refers to how people organize and make use of those biophysical elements; farms cities, roads, and so on. Any given land cover element may occur in many types of land uses but being in different land uses does not change what it is. That it is a tree (a land cover element) could occur in a natural forest, in a city park, or on a farm, all of which we would label as different land uses, but we would always label it as a tree.

Recall that a LULC system needs to be both exhaustive and exclusive; every given object or area on the landscape must fit into one and only class. LULC classification systems can be hierarchical, however, with simple broad categories containing more narrowly defined sub-categories. Forest Land may contain Deciduous Forest, Mangroves, and Evergreen Forests. The more detailed and specific your class definitions are, the better.

Remember also that there will be differences between classes intended for points and those intended for plots or other large areas; a single point can't be a forest, but an area can. Typically, you will want your point classes to allow you to create plot classes. If, for instance, you want to differentiate between types of forest—Deciduous, Evergreen, or Mixed—then you will need to have point classes that allow you label trees as Deciduous or Evergreen, and your LULC system must have rules for how to decide when an area is one type of forest or another.

7.4: Image Interpretation

Interpreting images is the core skill required for making use of CEO. While machine learning algorithms are increasing useful for identifying objects in images, the large scale of the data involved, technical complexity, and black box nature of such techniques mean they have yet to be widely adopted in the context of remote sensing. Humans are still considered the gold standard for identifying the "truth" of landscape composition. Image interpretation relies on making use of the following image attributes:

- Tone/color
- Shape
- Size
- Association
- Shadow

- Pattern
- Texture

Each of these attributes provides a different type of information to the interpreter and will be most useful in different contexts. Often the interactions between them will be extremely important for determining land use. The accurate interpretation of images is a complex skill that requires familiarity with the landscape being interpreted and practice. It should be supported with a well developed interpretation key containing examples of each class of object or area to be interpreted and types of ancillary information that may be useful for correctly labelling points.

7.5: Sample and Response Design

Most CEO projects will make use of a design-based inference, where a sample of a real population of interest is taken in order to estimate some parameters of interest for that population. To do so, you will need to create a response design and a sample design. The response design describes what will be sampled (unit of assessment), what surrounding context will be considered (spatial support unit), and how values will be assigned (rules of agreement and labeling protocol). The sample design determines where the sample points will be, how they will be collected, and how many of them will be collected.

The unit of assessment might be a pixel, an object that a point has fallen on, blocks of pixel, or polygons, and should be assigned to align with the goal of the project. The spatial support unit is the area around the unit of assessment that will be considered when interpreting and labeling it. The spatial support unit is often essential when assigning land uses. The labeling protocol defines how points will be labelled in CEO, and can include things like what to do when a point occurs on the edge of an object, a radius to consider around the point, what constitutes change, *etc*.

Once you have created a response design you will need to create a sample design. Your sample will need to be both a proper probability sample and be geographically balanced. This means that every part of the area being sampled has some probability of being selected, preferably a well-known probability, and that all parts of the area being studied are properly represented. Sample-design strategies should also generally be as simple as possible, consider the resources of the project, and allow for simple estimation of parameters, and accommodate changes in sample size.

There are many strategies for constructing samples, but for most CEO projects simple random sampling and stratified random sampling are likely to be the most useful. When it comes to setting a sample size, the most important considerations are the resources available to the project, and the required accuracy of the outputs. When thinking about the accuracy of estimates, we can frame this either through the lens of meeting a target number or as falling within a range of a certain size (confidence interval). The higher we want the accuracy to be, and the closer to the target we can say that it is, the larger the sample size that will be required. Similarly, the narrower the window of deviation we want to have, the larger the sample size. Stated very generally, the less tolerance for error you have the larger the sample will need to be for your project.

7.6: Data Collection and QA/QC

It is essential to maintain good quality data collection and quality control for your project. Using low quality, inconsistent data makes the outputs for your project equally questionable, even if you have designed an otherwise perfect project plan. The first step in this process is making sure that your interpreters are consistent with each other and with themselves across the duration of the project. This should be an integral part of the training process, and is another place where a well-developed interpretation key, clear classification system, and a robust set of labeling rules are very useful.

As the interpretation process moves forward, it’s important to include various checks to make sure that interpreters are remaining consistent. These can be self-checks, where the same point is reviewed at different times, or cross-checks between team members. It is also good practice for a system of checks on plots or points, either via reviewing data interpreters labelled indicated they were uncertain about, randomly selecting plots, or both.

Appendix 1. CEO Project Planning Worksheet

Before getting started on the technical work in Collect Earth Online (CEO), it is critical to gather the information needed to guide and plan your CEO image interpretation project. Review and answer the questions below and refer to the manual content for more information. If after review, you are unable to answer any of the questions below, you are likely not ready to start your CEO project and need to gather more information.

1) Objectives

a) What are the primary objectives of the project?

b) What are the specific objectives that will be me by using CEO?

c) What is the desired end product?

d) Is this a mapping (model-based estimation) or an inventory project (sample-based estimation)?

2) Classification system

a) What are the land cover or land use elements of interest?

b) What classes of interest can the land cover or use elements be grouped into?

c) What non-target classes and subclasses also need to be identified in order to have an exhaustive classification?

d) Do all classes have clear and exclusive definitions?

e) For multi-temporal analyses, do you need to define change classes?

3) Imagery and image interpretation

a) What are the spatial resolution, spectral resolution, and temporal resolution requirements for imagery to support the project?

i) Spatial requirements (what do you need to be able to see?):

ii) Spectral requirements (in what ways do you need to be able to see?):

iii) Temporal requirements (how often do you need to be able to see?):

b) Is the imagery readily available in CEO suitable for the needs listed above?

c) If readily available imagery is not sufficient what additional imagery exists and/or what resources are available to obtain additional imagery?

d) Can your classes be reliably identified in the available imagery? If not, return to review the project objectives and your classification system.

4) Response design

a) What unit of assessment best supports your project goals?

b) What is your plot size?

c) How many points are you using per plot, and how are they distributed?

d) Do your point classes need to be distinct from your plot classes (do they differ from your primary classification system)?

e) How will you define your spatial support unit?

f) What rules should be used to break ties and label points or plots?

5) Sample design: Sampling method and sample size

a) If you are working on an estimation or inventory project, answer the following questions:

   i) What sampling strategy will you use?

ii) Keeping in mind that larger sample sizes produce smaller standard errors, what are your accuracy needs? This could be expressed in various ways: Tolerance for Type I or Type II error, confidence interval width, or a target accuracy.

iii) Do you need this level of accuracy overall, or for individual classes or regions?

iv) Using the appropriate equation described in section 5.4, calculate your sample size. Is this a realistic sample to collect given your resources? If not, what can be adjusted?

b) If yours is a mapping project, answer the following questions:

i) For model training data collection:

(1) How will you target your sampling areas?

(2) Do you need to collect a purposive sample, or will random points suffice?

(3) How many points does your modeling method require for training?

ii) Map validation:

(1) Will you be validating a single or multiple map products?

(2) What are your accuracy needs? This could be expressed in any of the ways described above: tolerance for Type I or Type II error, confidence interval width, or a target accuracy.

(3) Using the appropriate equation described in section 5.4, calculate your sample size. Is this a realistic sample to collect given your resources? If not, what can be adjusted?

6) QA/QC

a) Do you have sufficient expertise about the landscapes being studied in your project to train others? Are your classes sufficiently easy to identify that little training is necessary (tree vs. not tree)?

b) Does an interpretation key already exist and if not, what resources can you use to help develop one?

c) What staffing resources do you have for your photo interpretation team?

d) What resources do you have for completing and evaluating training projects? What is the timespan available?

e) How will you facilitate collaboration and working together to achieve agreement?

f) How will you evaluate agreement?

g) What types of formal checks can you implement and how will you do so?

This manual was published with the support of the

Food and Agriculture Organization of the United Nations.