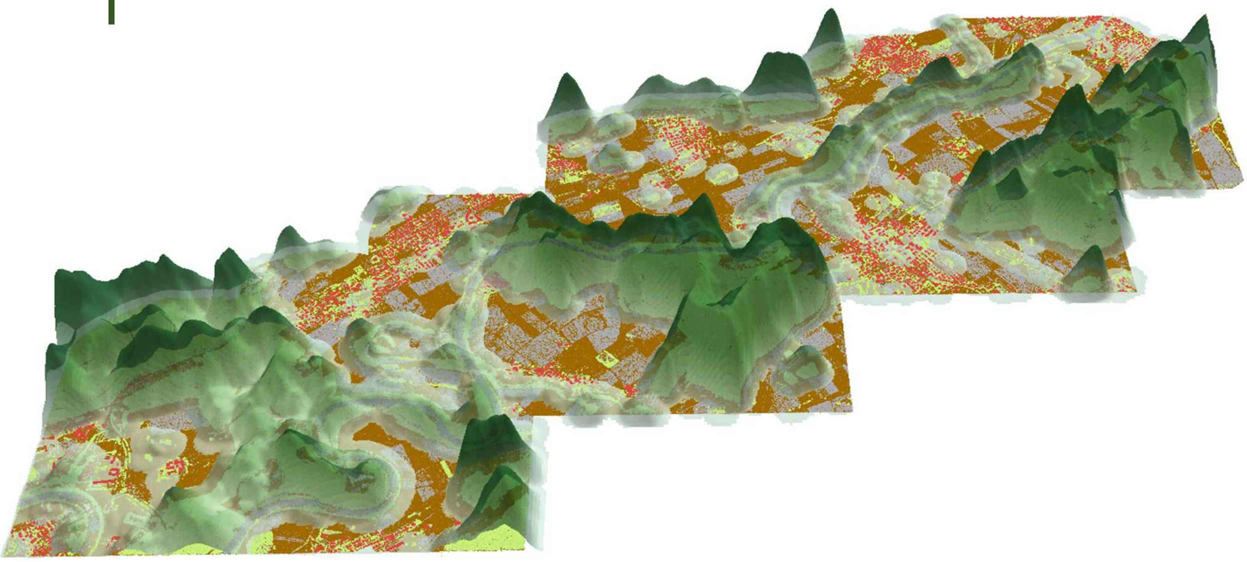
Swiss Federal Institute for Forest, Snow and Landscape Research WSL

Vertical Vegetation Structure Classifier (VVSC) : A novel tool for ArcGIS

The user manual



Pierre Cothereau

4/18/2016

Supervisors

Janine Bolliger

Christian Ginzler

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

WSL

ABSTRACT

The importance of 3D structure in spatial ecological assessments is continuously increasing and its influence on animal ecology is established, but most of the landscape ecological analyses are still considering landscape as a planar surface without any 3D elements. Advancements in Light Detection and Ranging (LiDAR) remote sensing offer new opportunities to quantify and visualize ecosystem structures in 3D at large spatial scales. To make LiDAR technology accessible to a broad range of users, we developed the Vertical Vegetation Structure Classifier (VVSC). This easy-to-use classifier extracts vegetation heights in user-defined 3D classes within user-defined regions (buffers, corridors). First, VVSC differentiates between vegetation and non-vegetation by classifying the LiDAR points according to the ASPRS (American Society for Photogrammetry and Remote Sensing) using LAStools from rapidlasso. VVSC then sorts the vertical structure of vegetation into user-defined height classes. Two input data sets are required: the LiDAR data in LAS or LAZ format, and the shapefile in the multipoint format containing the center points of the future user-defined buffers in the area of interest. Four output files for each buffer size and vegetation height are delivered: a shapefile with the distribution of the vegetation area, a text file with the coordinates, the intensity and the ASPRS class of each point, and two raster files to visualize the point density and the maximum point density.

TABLE OF CONTENTS

1. PRODUCT INFORMATION

1.1 About the Authors

VVSC was developed by Pierre Cothereau during his Master thesis at the Swiss Federal Institute for Forest, Snow and Landscape Research WSL under the supervision of Janine Bolliger and Christian Ginzler. In April 2016, VVSC received the Switzerland ESRI Young Scholar Award 2016.

1.2 Download

This manual, as well as the model and the material necessary to the completion of the practical example, are freely available on the WSL website:
http://www.wsl.ch/fe/landschaftsdynamik/projekte/Vertical_Vegetation_Structure_Classifier_VVSC/index_EN

1.3 System requirements

VVSC simply requires a standard computer with a minimum RAM of 2GB and at least 2.4 GB of disk space running under Windows XP, Windows Vista, Windows 7 or Windows 10. An ArcGIS license with the extension Spatial Analyst and 3D Analyst and the version 10.3 of ArcGIS are required, as well as a license for the LAStools from rapidlasso (http://rapidlasso.com/) [Lastools].

1.4 Technical setup

The first section of this guide describes how to set up your LAStools toolbox within ArcGIS to be able to use VVSC without errors or problems. This setup is only necessary the first time you use VVSC.

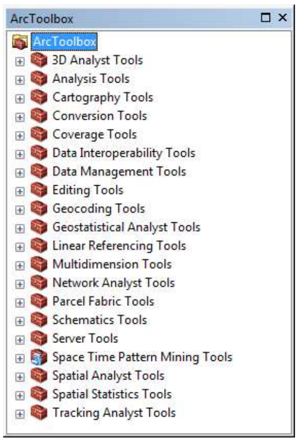
First of all, install LAStools in this exact location “C:\Prog\”. If the “Prog” folder does not exist on your local disk (C:), you have to create it. Long paths that include too many characters will often lead to errors in ArcGIS. But as the VVSC toolbox is a custom toolbox, it will only recognize the tools from the LAStools toolbox if they are located at the location having this absolute path “C:\Prog\LAStools\ArcGIS_toolbox\LAStools.tbx”. If you already saved the LAStools in any other location on your computer, please copy the LAStools in order to also have them stored as indicated above.

Concerning the VVSC toolbox, its only precondition is that it should be saved on the local disk (C:). The name of the folder in which you save it is not relevant, as long as the VVSC toolbox is on the local disk (C :).

VVSC was built using ArcGIS ModelBuilder, and some parameters of the LAStools toolbox need to be modified. The modifications are necessary to allow ArcGIS to recognize the tools within VVSC as the tools from your LAStools toolbox. These modifications concern the data type of the LAStools functions. In ArcGIS, the input and output data type of the LAStools functions always appear as a data type called "File". However, to link these LAStools functions with ArcGIS Modelbuilder, you need to adapt the necessary tools with the next few steps:

Step 1: Add the LAStools toolbox to the ArcToolboxes:

- Open ArcMap or ArcScene.
- Under Geoprocessing, click on ArcToolbox . A window with the default ArcGIS toolbox will now be displayed (see Picture 1).
- Right-click on ArcToolbox → Add Toolbox and go in C:\prog\LAStools\ArcGIS_toolbox and select the LAStools.tbx



Picture 1: Default ArcToolbox

Step 2: Create copies of three tools of the LAStools:

- In the LAStools toolbox, right-click on the tools Lasclip → "copy". Then right-click on the LAStools toolbox and click "paste"
- Repeat this operation for Lasboundary and Las2shp.

They will now appear in your LAStools toolbox as Lasboundary (2), Lasclip (2), and Las2shp (2).

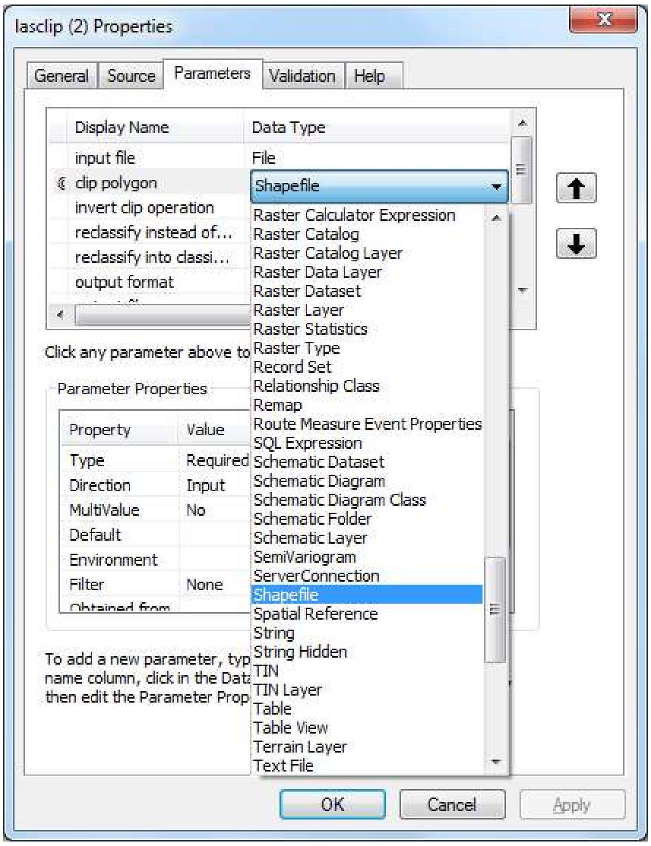
Step 2: Modify the data type of these three copied tools:

- Right-click on each one of them → Properties → Parameters and perform the following modifications:

For Lasclip (2), change the data type of "clip polygon" from "File" to "Shapefile".

For Lasboundary (2), change the data type of "output file" from "File" to "Shapefile"

For Las2shp (2), change the data type of "output file" from "File" to "Shapefile"



Display Name	Data Type
input file	File
clip polygon	Shapefile
invert clip operation	
reclassify instead of...	
reclassify into classi...	
output format	

Property	Value
Type	Required
Direction	Input
MultiValue	No
Default	
Environment	
Filter	None
Obtained from	

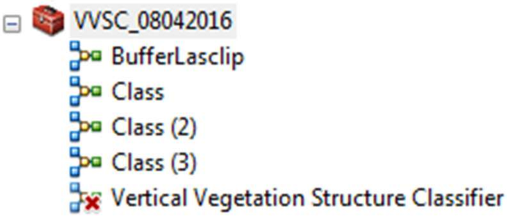
Picture 2: Setup of the correct data type for the Lasclip (2) in the Lasclip (2) Properties window

Step 3: Add the VVSC toolbox to the ArcGIS Toolboxes:

- In the window called ArcToolbox[1], right-click on ArcToolbox at the top of the window and click on "add toolbox".
- Then search for the folder where the VVSC toolbox is saved, and select it.

Step 4: Set up VVSC:

- Within the VVSC toolbox, five models appear: BufferLasclip, Class, Class (2), Class (3) and VVSC.[2]


VVSC_08042016
- BufferLasclip
- Class
- Class (2)
- Class (3)
- Vertical Vegetation Structure Classifier

Picture 3: The five submodels from the VVSC toolbox

---

[1] If this window is not open, the ArcToolbox window can be open by clicking on ArcToolbox in the standard toolbar.

[2] If only two models are present, you need to update your version of ArcGIS.

- If you correctly performed the three previous steps, four of these five models will appear ready to be used and one (VVSC) will have a red cross, meaning it is not ready to be used (if the five submodels appear ready to be used, then ignore the next part, your tool is ready). Right-click on VVSC, and click on "Edit".

A window will open, showing you all the tools included in the model VVSC. You will quickly notice that four functions appear with a red cross; these are models, included within the main model (which are referred as submodels). These four submodels are the models included in the VVSC toolbox called BufferLasclip, Class, Class (2), and Class (3). The problem is simply that ArcGIS does not link these four submodels with the ones from the VVSC toolbox. So, the user has to make this connection manually:

- Double click on the BufferLasclip, search for the VVSC toolbox and click on the model BufferLasclip. In this way, ArcGIS will know that the submodel BufferLasclip is the model with the same name from the VVSC toolbox.
- Repeat the same operation for Class, Class (2) and Class (3).
- Save the VVSC window and close it.

Congratulations, the model VVSC should now be ready to run. However, in order to have VVSC all set for your next ArcMap or ArcScene sessions, I highly recommend to now save your ArcToolbox as Default:

- Right-clicking on ArcToolbox in the ArcToolbox window, and click Save Settings as Default.

For the next steps on how to use VVSC, consult Chapter 6.2 "step-by-step tutorial".

2. INTRODUCTION

2.1 Introduction/Context

3D information on ecosystems has long been an essential topic in ecological research and spatial ecology. Already in 1961, MacArthur assessed the importance of the vertical vegetation structure on bird diversity and proposed the Foliage Height Diversity (FHD) index to quantify the relationship between foliage height and bird diversity [MacArthur, 1961]. Currently, the importance of the 3D structure in spatial ecological assessments is continuously increasing and its influence on animal ecology is now well established [Davies *et al*, 2014]. However, most of the landscape ecological and spatial connectivity analyses are still applying the "patch-corridor-matrix" method proposed by Forman in 1995 [Forman, 1995]. This method perceives the landscape as a planar surface and does not take into account any elements of the third dimension, neither the relief, nor the vertical vegetation structure [Hoechstetter *et al*, 2008].

Recent technical advancements in Light Detection and Ranging (LiDAR) remote sensing offer new opportunities and challenges to science and ecological management. The possibility to measure, quantify and visualize the ecosystem structure in 3D at large spatial scales and at relatively low costs has allowed for new insights in our environment's structure.

The main benefit of the active LiDAR sensor is that the beam of the scanner passes through the vegetation, and therefore gives information on the canopy height as well as on the understory. The LiDAR data are represented usually by points, resulting in very large data sets, making the extraction of ecosystem 3D structure difficult for users unfamiliar with remote sensing techniques. In order to make LiDAR, the most recent 3D technology, accessible to a broad range of users - including those without an extensive knowledge in GIS and remote sensing - I developed the Vertical Vegetation Structure Classifier (VVSC). This classifier extracts vegetation heights in user-defined 3D classes within user-defined regions (i.e. buffers and corridors) which can then be combined with additional information on 2D ecosystem properties (land use, climate, etc.).

2.2 What is the Vertical Vegetation Structure Classifier (VVSC)

This novel tool, the Vertical Vegetation Structure Classifier (VVSC), is an ArcGIS model that aggregates LiDAR data into different classes of user-defined vegetation heights (i.e. woody biomass) and provides the user with multiple information on these classes such as the area covered per height class, the point density and the spatial distribution of each vegetation-height class across the landscape. The aggregation of the height classes is performed in specific user-defined regions. This allows for a broad range of ecological applications. Among many, the main goal of VVSC is to offer the user a visualization of the connectivity of the 3D vegetation structure.

The number of vegetation height categories is up to the user, as its relevance and significance depend on the user's interest. Thereafter, VVSC will first differentiate between vegetation and non-vegetation by classifying the LiDAR points according to the ASPRS (American Society for Photogrammetry and Remote Sensing) using LAStools from rapidlasso [LAStools]. Secondly, VVSC extracts and sorts the vertical structure of vegetation into the chosen height classes, up to a maximum of six. Again, the user chooses the ASPRS classes of points which will be included into the analysis and classification. Since buildings may represent a potential habitat for certain species (e.g. bats), it might be necessary to include them into the process.

VVSC requires only two input data sets; the first is the LiDAR data of the area of interest in the LAS or LAZ format, and the second is a shapefile in the multipoint format containing the center points of the user-defined areas of interest. The first step in the model is the creation of buffers around points or areas of interests (shapefile format). The centers of these buffers are the coordinates of the points from the input shapefile, whereas the size of the radii of the different buffers are user-defined and depend on the research question. In a second step, the buffers are used to clip the LiDAR data within the buffers. Thirdly, VVSC will sort each of the LiDAR points according to the categories set by the user and create the different classes of vegetation heights. Finally, VVSC will provide the user with four output files for each buffer size and vegetation height; a shapefile with the distribution of the vegetation area, a text file with additional information of each LiDAR point, such as the intensity and the ASPRS class of each point, and two raster files to visualize the point density and the maximum point density. By using ArcScene, these two raster files allow the user to visualize the result of the point densities in 3D, i.e. the areas of high density of a certain vegetation-height class.

2.3 Possible applications of VVSC

VVSC has a large range of possible applications in the domain of spatial analysis as well as in ecological connectivity, forestry and agroforestry. These may include the assessment of habitat suitability [Zimble *et al*, 2003] and determination of potential corridors, monitoring of understory regrowth and canopy growth, connectivity analysis etc. [Merrick *et al*, 2013]. Indeed, every field of research or management that requires information about the vertical structure of vegetation can make the most of VVSC.

In landscape ecology, the main benefit of VVSC is the possibility to study the natural corridors that appear depending on the vegetation height. VVSC allows the user to spot which layers of vegetation would most fit the needs of a certain species in term of connectivity, vegetation cover and environment. One of the main idea behind VVSC was to determine the height at which a barrier (such as a road, a river, a building, etc.) would become a possible pathway for wildlife. In fact, many landscape connectivity analyses were based on two-dimensional maps that simplified the reality by bringing out elementary elements such as crops, forest, road, river, etc., and by looking at the vegetation connectivity between these different elements, the most probable wildlife pathways could be identified. With the vertical vegetation structure information, we can look more subtly at more hidden pathways such as the paths created by trees crossing a river or canopies connecting high above a road. Depending on the species of interest, such hidden corridors can have a significant-impact on the estimated connectivity of the area.

Concerning the applications in forestry, VVSC provides all the information about the vegetation height and canopy cover, as well as the data necessary to identify the seral stage of a forest. Indeed, by knowing the canopy and understory cover, it becomes relatively easy to notice gaps and areas of regrowth inside a forest. Furthermore, if the user possesses LiDAR data of the same area at different periods of time, an important amount of information about the forest's management and development can be extracted, such as the area of deforestation, the growth rate, the recovery time after a disturbance and potential expansion or diminution of the forest’s boundary.

In a more general way, VVSC has the means to provide many useful answers about wildlife and plant conservation, their protection and management. From the vertical organization of the vegetation, the user can deduce the vertical arrangement of fruits, flowers and foliage, which can lead to an estimation of the species breeding, nesting and perching sites [Hunter, 1999]. Therefore, VVSC has great potential to provide access for both, researchers and managers, to a more effective comprehension of the recent remote sensing tools and the new possibilities they offer.

2.4 Functions and methods of VVSC

The goal of this part is to provide some technical details about the functions employed within the model framework, as well as the reasons behind these functions’ choices, in order to enable the more experienced users to adapt and modify VVSC to fulfill more accurately their needs.

The initial objective behind the development of VVSC was to design a model to study the ecological connectivity of species based on the vertical structure of vegetation. As a consequence, the choices made during the development phase of VVSC were carried out in this direction, even if it appears that VVSC has the potential to address additional challenges.

Before starting to detail the functions used within VVSC, it is important to clarify the notion of “class” used along this tutorial. Indeed, there are two different types of classes, and thus, two types of classifications, performed by VVSC: The ASPRS classes and the vegetation-height classes. The ASPRS, American Society for Photogrammetry and Remote Sensing, developed standard LiDAR point classes as shown in Table 1. Each class is attributed a number called classification value with a specific meaning. For instance, points with a classification value of six are buildings and the ones with a classification value of two are the bare ground. This ASPRS classification plays a major role in VVSC, as the points classified as “Ground” serve as the reference to compute the height aboveground of all the other LiDAR points. Basically, the ASPRS classification has two essential purposes: Identification of the points that represent the bare ground to obtain the height aboveground of all the other LiDAR points and the differentiation between vegetation and non-vegetation (buildings, roads, rails, water etc.) for each LiDAR points. It is this height aboveground obtained during the ASPRS classification that is used to perform the second type of classification: The vegetation-height classification. The vegetation-height classes are the user-defined categories of vegetation heights obtained as the main outputs of VVSC. For example, as default in VVSC, the first class is composed of all vegetation with a height between 0.5 to 3 meters, the second class with a height of 3 to 6 meters, and so on.

Classification value	Meaning
0	Never classified
1	Unassigned
2	Ground
3	Low Vegetation
4	Medium Vegetation
5	High Vegetation
6	Building
7	Noise
8	Model Key/Reserved
9	Water
10	Rail
11	Road surface
12	Overlap/Reserved
13	Wire - Guard
14	Wire - Conductor
15	Transmission Tower
16	Wire - Connector
17	Bridge Deck
18	High Noise
19 - 63	Reserved

Table 1: ASPRS classes as used in VVSC

To start with VVSC, it is important to notice that VVSC uses four submodels, three of them being identical, as they operate the same functions on the three different buffers (Fig. 2). These four submodels are BufferLasclip, Class, Class (2) and Class (3) and their main functions are summarized in Fig. 2. The first submodel is the BufferLasclip that, as its name suggests, creates the buffers around points of interest and clips the LiDAR data located within the buffered area. This is done by a combination of the classical Buffer function offered by ArcGIS and the Lasclip tool provided by the LAStools. This solution is a really good example of the very efficient completion of using specific LAStools tools able to handle important quantity of data, with the ArcGIS interface and numerous functions.

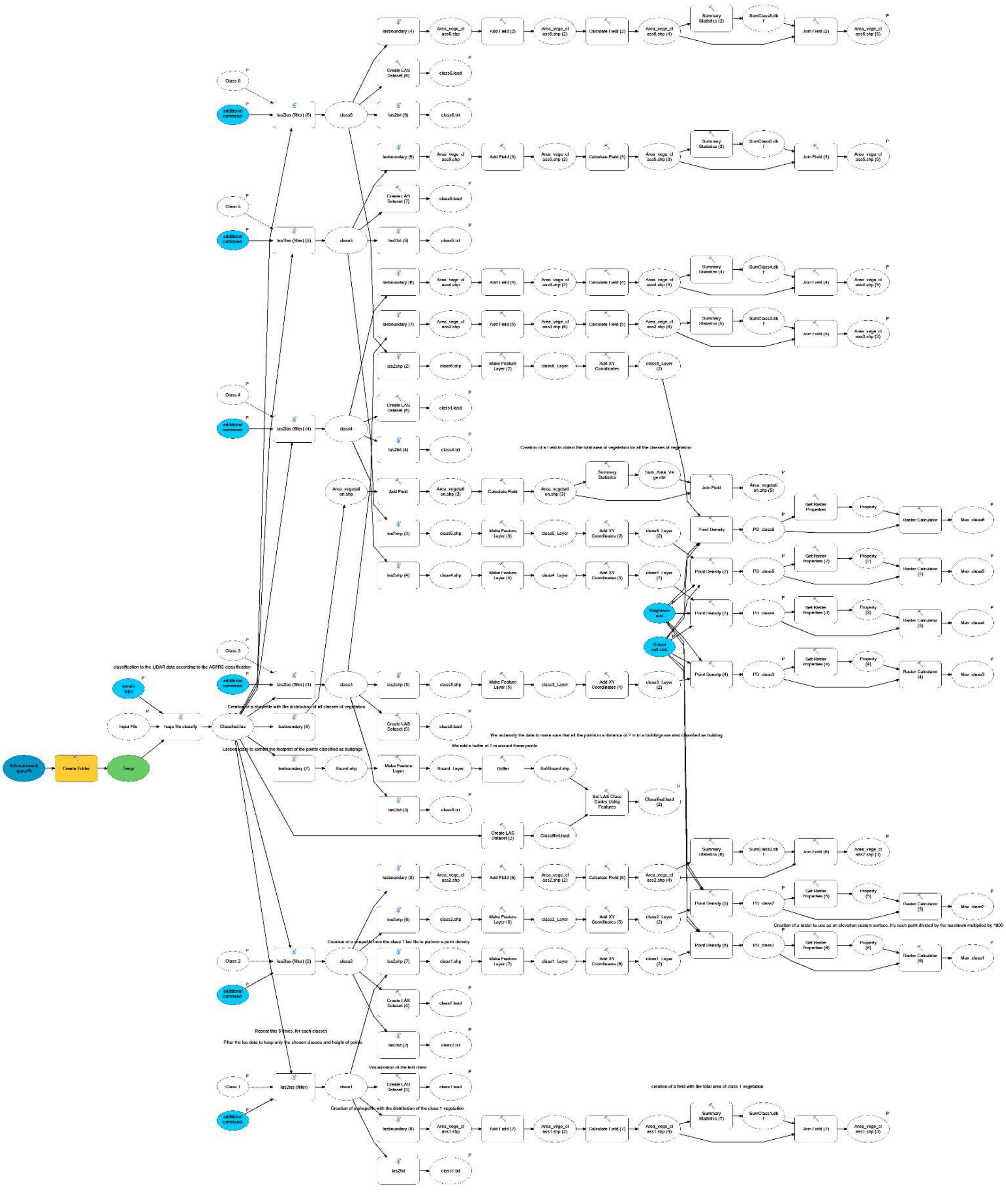


Figure 1: Visualization of the tools forming the submodel Class

Once the submodel BufferLasclip is completed, the data is moved to the next submodel called Class (Fig. 1). Class handles the ASPRS classification of the LAS points. Indeed, the classification of the LiDAR data is performed with a succession of Lasground, to identify the LAS points as ground or non-ground, followed by Lasheight, which assigns the height of each LiDAR point above the newly-identified ground. Lastly, Lasclassify distinguishes between vegetation height and buildings [Lastools]. This succession of tools was efficient to correctly classify the LiDAR points into the correct ASPRS class. Nevertheless, the

ASPRS classification's efficiency is further discussed in the Chapter 7 "Use and limitations of VVSC". Yet, this combination of tools failed to process large datasets, requiring the pipeline "huge file classify" from the LAStools. A LAStools pipeline is a combination of diverse LAStools tools put together in a model to form a new tool. This "huge file classify" pipeline divides the large LiDAR input into numerous tiles. On each one of them an ASPRS classification is performed through the succession of the three tools Lasground-Lasheight-Lasclassify, allocating the LiDAR points into their appropriate ASPRS class. Finally, the now classified tiles are reassembled together to reform the initial LiDAR dataset. The only drawback of using a pipeline is the reduced freedom in the quantity of parameters that can be modified and optimized. Indeed, without using the pipelines, there are around five parameters for each tool that can be defined by the user whereas when the pipeline is used, only five parameters can be chosen for the whole combination of tools. For instance, when using the tool Lasground individually, the user has the choice to let the Lasground algorithm identify the points that represent the bare ground, or to provide a digital terrain model (DTM), which allows for a very efficient identification of the points that form the ground. This possibility of using a DTM is lost when using the pipeline "huge file classify". However, despite this drawback, choosing "huge file classify" fits with the concept of VVSC being a tool also usable by people not familiar with remote sensing software, as the optimization of the different parameters for a more efficient ASPRS classification would have made the model much more complex. In the submodel "Class", the user notices two tools, called Lasboundary (2) and Lasboundary (3), used directly after Classfied.las (Fig. 1). The Lasboundary function constructs a polygon shapefile of the boundaries of LiDAR points by projecting the LiDAR points from the third dimension to the second dimension. Lasboundary (3) creates the output "area_vegetation.shp" that presents the footprint of the points classified as "Vegetation" for all LiDAR points inside the buffer. Lasboundary (2) is an additional classification performed by VVSC which assigns a two meters buffer to all points classified as buildings. Thus, all the small objects in an area of 2m around a building (ASPRS class 6) are reclassified as buildings. This ASPRS reclassification of the already classified LAS dataset is performed by the function called "Set LAS Class Codes Using Features".

As mentioned previously, the first step of the Class submodels is to classify the LiDAR points and to prepare and create the final outputs. On Fig. 1, after the pipeline "huge file classify", there are six groups of tools, all identical and all beginning with the tool Las2las (filter), which represent the creation of the outputs for the six vegetation height classes. Las2las (filter) is the first tool that executes the preferences of the user, by filtering and storing in a new file, all the LAS points that match the user-defined height classes. The tool Las2txt creates the text file containing information about all the points selected by the user-defined criteria. The Lasboundary constructs a polygon shapefile of the LiDAR points for each vegetation-height class. Basically, it projects the points from the third dimension to the second dimension and extract polygons with the area of vegetation. Subsequently, the area of each polygon is calculated and an additional field is added in the attribute table indicating the sum of each polygon's area and hence, giving the total area of vegetation for each vegetation-height class. To sum up, Lasboundary provides the area of each polygon of vegetation (e.g. to obtain the area of one specific forest in the area of interest) and the total area of vegetation in the whole vegetation height-class (e.g. to know the total area of vegetation that corresponded to the user-defined criteria in the buffer). One important parameter of the Lasboundary function is the concavity number. Concavity numbers indicate a trade-off between the boundary accuracy and storage space on the disk. In the case of VVSC, the concavity is 1 allowing for a good accuracy of the boundaries, but requiring more storage space. By increasing this number, the accuracy is reduced but saves space on the disk. In contrast, decreasing this number results in a more accurate delimitation of the area of vegetation but requires supplementary storage capacity. Yet, 1 is already a rather small concavity choice. For more information, I highly recommend to always look at the Help provided within the tools. The tool Create Las Dataset simply does what it says and generates the LAS dataset which can be displayed on the ArcScene main window. The user can then visualize the LAS dataset for each vegetation-height class. The last tool used from the classfied.las is the las2shp, which converts the LAS file into a shapefile in order to perform the point density. The two next tools from ArcGIS "Make Feature Layer" and "Add XY Coordinates" format the shapefile to fit the conditions required to make a point density. The point density is a smoothing function that will allow for an easier and more understandable representation of the points. Its role is mainly for visualization purposes, even if it also allows for a qualitative assessment of the repartition of the vegetation inside the areas of interest. The user should not pay too much attention to the values obtained, but rather focus on the colors (darker colors signifying a higher presence of vegetation matching the user-defined criteria). As a kernel density weights the points near the centers higher than the points at the boundaries, the point density was chosen. This method weights all points equally.

All main operations are performed by the submodels Class, Class (2), Class (3). The main model called VVSC only serves as a platform to regroup the four submodels within one main model and manage the order of operations of each tool as well as the precise storage of the outputs. It creates a folder for each buffer and controls that each outputs is placed within the correct folder.

To conclude, VVSC is and remains a tool, and as such, it can be modified, adapted and optimized by every user based on its own understanding and knowledge of remote sensing tools and software. It can also be simplified by deleting the submodels Class (2) and Class (3) to function only with one buffer, or by removing the BufferLasclip submodel to analyze the whole dataset instead of using buffers (See 12.2 "Annex" for further information).

2.5 Summary diagram of VVSC

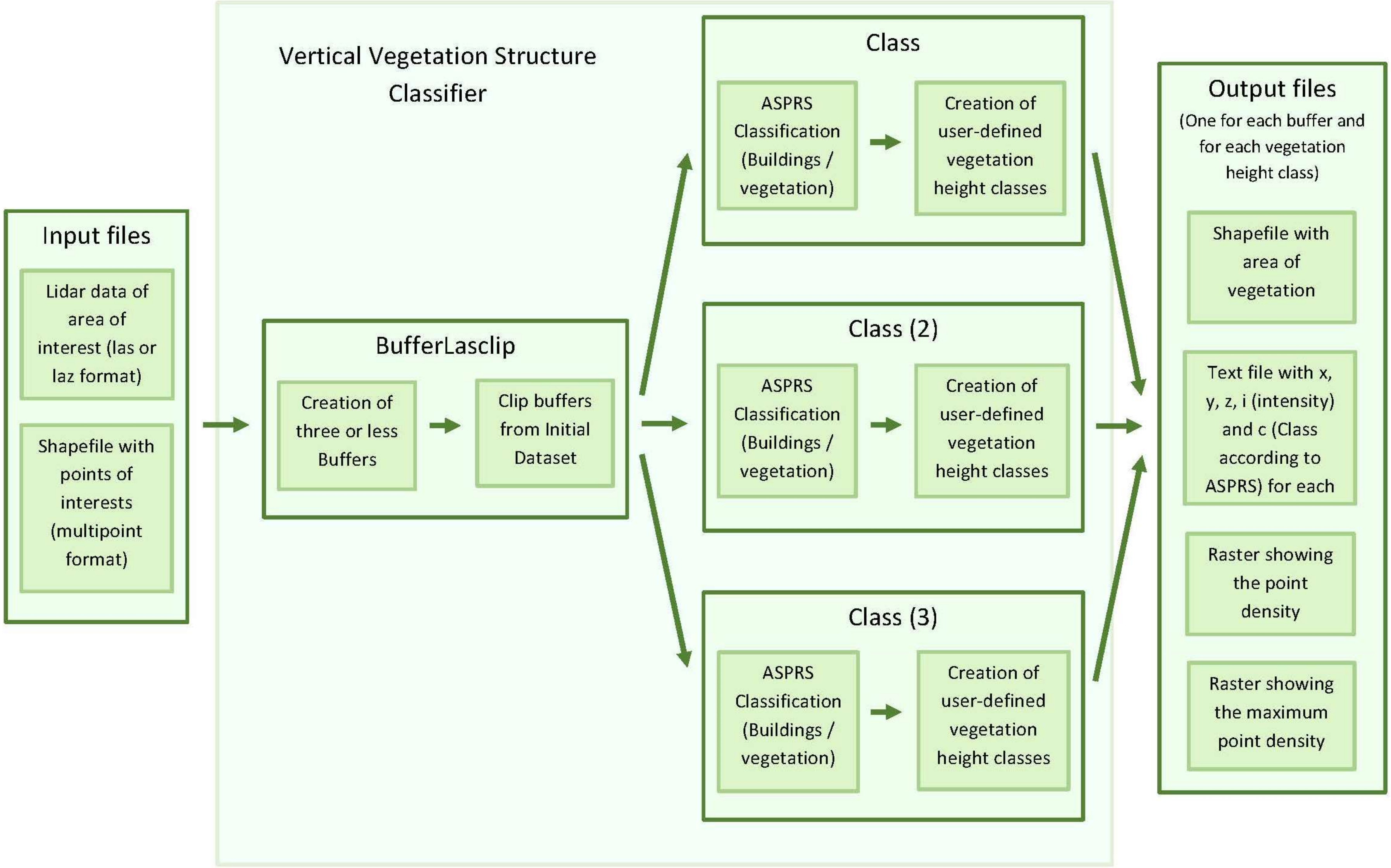
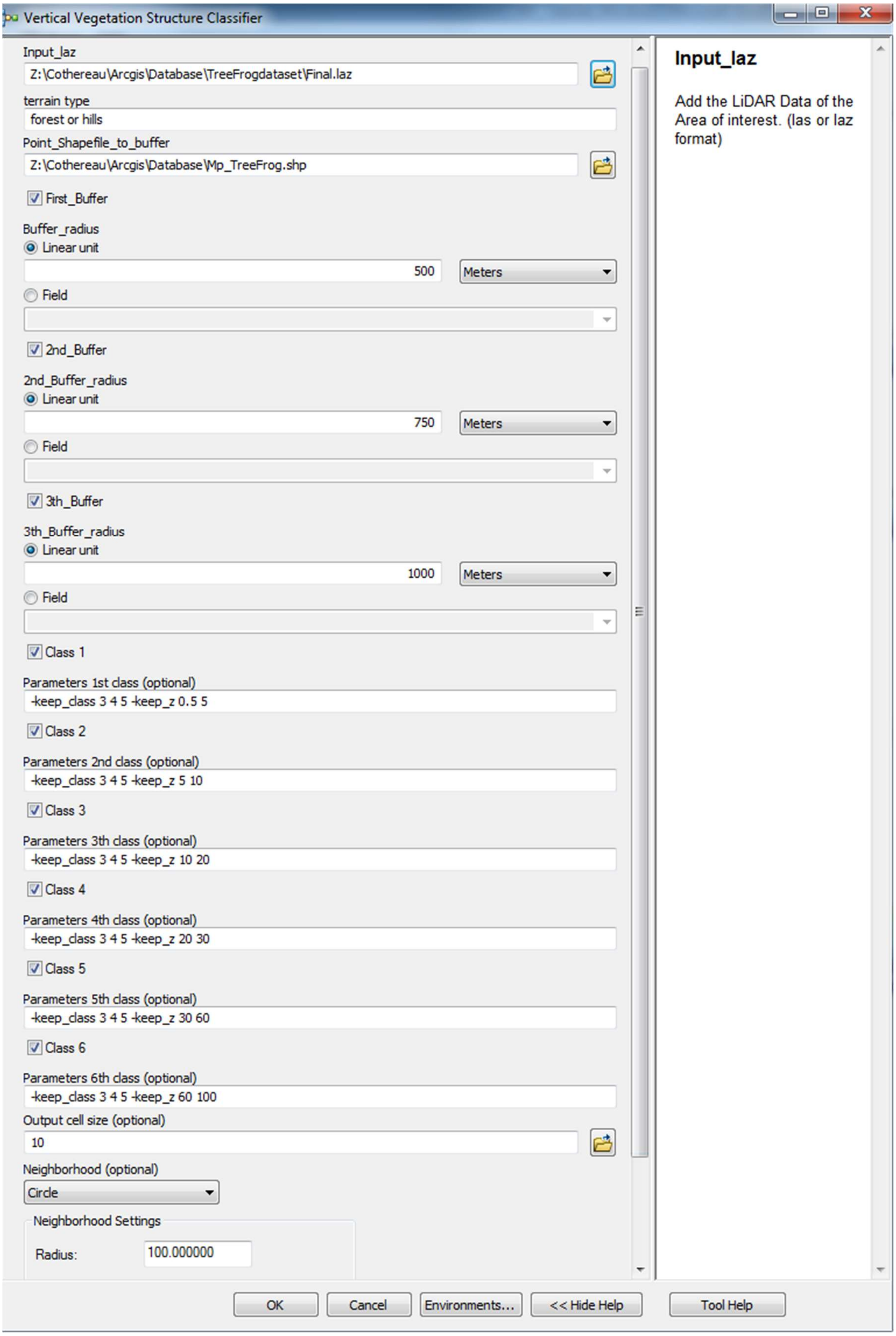


Figure 2: Summary diagram of the main process of each of VVSC model. BufferLasclip, Class, Class (2) and Class (3) are the four submodels of VVSC

3. USER INTERFACE



Vertical Vegetation Structure Classifier

Input_laz
Z:\Cothereau\Arcgis\Database\TreeFrogdataset\Final.laz

terrain type
forest or hills

Point_Shapefile_to_buffer
Z:\Cothereau\Arcgis\Database\Mp_TreeFrog.shp

☑ First_Buffer

Buffer_radius
● Linear unit
500 Meters
○ Field

☑ 2nd_Buffer

2nd_Buffer_radius
● Linear unit
750 Meters
○ Field

☑ 3th_Buffer

3th_Buffer_radius
● Linear unit
1000 Meters
○ Field

☑ Class 1

Parameters 1st class (optional)
-keep_class 3 4 5 -keep_z 0.5 5

☑ Class 2

Parameters 2nd class (optional)
-keep_class 3 4 5 -keep_z 5 10

☑ Class 3

Parameters 3th class (optional)
-keep_class 3 4 5 -keep_z 10 20

☑ Class 4

Parameters 4th class (optional)
-keep_class 3 4 5 -keep_z 20 30

☑ Class 5

Parameters 5th class (optional)
-keep_class 3 4 5 -keep_z 30 60

☑ Class 6

Parameters 6th class (optional)
-keep_class 3 4 5 -keep_z 60 100

Output cell size (optional)
10

Neighborhood (optional)
Circle

Neighborhood Settings
Radius: 100.000000

Input_laz

Add the LiDAR Data of the Area of interest. (las or laz format)

Figure 3: User interface of VVSC within ArcGIS (each parameter is described in the Tool help)

4. VERTICAL STRUCTURE OF VEGETATION

When describing vegetation, there are two main elements that need to be taken into account; the taxon (floristics) and the physiognomy (structure). The vegetation's physiognomy can be divided into two components, the vertical and the horizontal structure [Bergen *et al*, 2009; Dunning *et al*, 1992]. The vertical structure of vegetation is defined as the "bottom to top configuration of aboveground vegetation [Brokaw and Lent, 1999]. But in our case, a more recent definition more adapted to this remote sensing context proposed by Davies *et al* in 2014 will be preferred. They refer to the vertical structure of vegetation as "the arrangement of canopies and tissues in the z dimension", whereas the horizontal component is composed of these arrangements within the x and y dimension, such as the canopy cover and extent [Davies *et al*, 2014].

The history of vertical structure of vegetation dates back to over 100 years ago, when ecologists first noticed that different animals lived in different strata of vegetation. However, it was really in 1961, with the FHD (Foliage Height Diversity) index developed by MacArthur to quantify the relation between foliage height and bird diversity, that a real step forward was made [MacArthur, 1961]. Actually, since MacArthur, most of the ecologists agreed that the more complex the vertical structure of a forest is, the higher its biodiversity and species richness. Indeed, as vertical complexity increases, the possibilities of foliage arrangements are multiplied, offering more diverse habitats and niches and thus leading to an enhanced biodiversity [Hunter, 1999].

There are direct and indirect effects of vertical structure on species richness. The most evident direct effects are the vertical arrangements of fruits, flowers and plants. Since these are the main sources of food for wild animals, their distribution has a major impact on the species repartition within the forest. In the same concept, the arrangement of branches and stems within the forest has straightforward consequences on the location of nesting, breeding and resting sites of most species [Hunter, 1999].

However, most of the effects of vertical complexity on biodiversity are indirect and difficult to measure, as they impact the forest's microclimate and the distribution of animals. For instance, the quantity of light and water reaching the sublayers of vegetation and the ground will alter the habitat conditions considerably and affect the local temperatures, which will decide which species, both animals and plants, can establish [Davies *et al*, 2014].

Birds are without any doubt the most studied group when speaking of vertical structure of vegetation, even though recent studies also focus on amphibians, primates and reptiles [Bergen *et al*, 2009]. Despite the growing interest in the subject, the potential for further discoveries and a better comprehension of our environment is immense and tools like VVSC possess the means to help scientists improve the common knowledge about the relations between species richness and the vertical structure of vegetation. This is particularly important for ecosystem assessments at large spatial scales, where the spatial distribution and configuration of 3D vegetation structure may impact connectivity. Tools like VVSC that improve our knowledge on the physical properties of the landscape lead to a development of the structural connectivity metrics. These same connectivity metrics are the means used to measure the landscape connectivity. Hence, VVSC may result in more accurate estimations of the habitat accessibility and species movements and prevent further fragmentation of the landscape. In the long run, this enhanced understanding of ecosystem structures could result in a more adequate management of our environment and help choosing the most efficient strategies for species conservation.

5. LIDAR (LIGHT DETECTION AND RANGING)

In the previous chapter, the importance of the vertical structure of vegetation was discussed, and we now focus on the most qualified remote sensing instrument to obtain this information: the Light Detection and Ranging scanners called LiDAR.

LiDAR is an active remote sensing technique emitting non-visible electromagnetic waves in the near-infrared. A laser pulse is emitted from a source and reflects on the target surface back to a sensor's receiver. Knowing the exact location of the source and the travel time of the laser pulse, the distance to the target can be calculated by multiplying the travel-time by the speed of light and dividing it by two (round trip). As the light pulse progresses toward the ground, it will be intercepted by objects including human infrastructure, vegetation and eventually the ground. This return signal will allow to calculate the height (and position) of each object encountered. This high capacity of LiDAR to calculate object heights (as well as canopy height) explains why it is also known as laser altimetry. Furthermore, LiDAR is famous for its capacity to provide direct information about the canopy height, the understory topography and the distribution of all objects encountered during the signal's progression [Dubayah and Drake, 2000]. Indeed, as the laser pulse sinks into the vegetation, most of it will be reflected, but some will pass through the foliage to reach the ground. In addition, the laser pulse will also pass through any small gaps or holes in the vegetation cover, providing an entire overview of the vertical structure of the vegetation [Davies *et al*, 2014]. The capacity of the light pulse to cut through the vegetation depends on the wavelength chosen, as well as its capacity to pass through water [Lefsky *et al*, 2002].

There are two types of LiDAR systems: Discrete-return and "continuous" waveform devices. Discrete-return devices have a very high spatial resolution and a very small footprint (less than 1 cm of diameter) in opposition to the waveform systems having a larger footprint (5 to 7 meters of diameter) [Vierling *et al*, 2008]. However, most of commercial ALS systems possess a small footprint of approximately 30 cm [Dubayah and Drake, 2000]. For forestry purposes, the second type of LiDAR system is recommended as their larger footprint allow the signal to constantly reach the ground and provide a better visualization of the 3D structure of the vegetation [Dubayah and Drake, 2000; Lefsky *et al*, 2002]. In the case of VVSC, it is this second type of LiDAR system that possess the most potential due to its enhanced ability at vertical profiling.

Nevertheless, there are some limitations of using LiDAR. There are many factors that could create confounding signals, such as the presence of clouds when the airborne LiDAR scanner is used, the ground roughness, the presence of human infrastructures or objects (cars, fences, cranes, powerlines etc.) and the seasonal variations [Bergen *et al*, 2006]. Furthermore, LiDAR cannot directly provide some essential variables such as “stem densities (for forestry applications), tree species and densities of woody debris on the forest floor” [Bergen *et al*, 2009]. Yet, some of these variables can be estimated based on the tree’s height acquired with the LiDAR scanners, or, more simply, by combining the information obtained with the LiDAR scanners with data coming from different sources or additional satellite measurements.

Another important limitation is the fact that part of the laser pulse does not pass through the vegetation (depending on the wavelength), which often results in an underestimation of the density of the understory [Richardson *et al*, 2011]. When using VVSC, it is essential to remember this underestimation of the understory, and also that, as the LiDAR used in VVSC are from airborne origin, the quantity of points near the ground is lower than the quantity of points at higher altitude. It is simply due to the progression of the light signal that goes from the sky to the ground, so most of the signal will be reflected in higher heights. These two points will be discussed more in depth in the Chapter 7 “Use and limitations of VVSC”.

6. PRACTICAL EXAMPLE

6.1 Study area and dataset

In this part, a step-by-step example on how to use VVSC is provided. The data used comes from the Canton of Aargau along the river Reuss and concerns tree-frog breeding ponds. The point's coordinates are available in the file "Mp_TreeFrog.shp", and will serve as the buffer centers in this example. The LiDAR dataset was obtained with an airborne laser scanner LMS-Q680i for the Canton of Aargau by numerous flights at 700 m above the ground between March and July 2014.

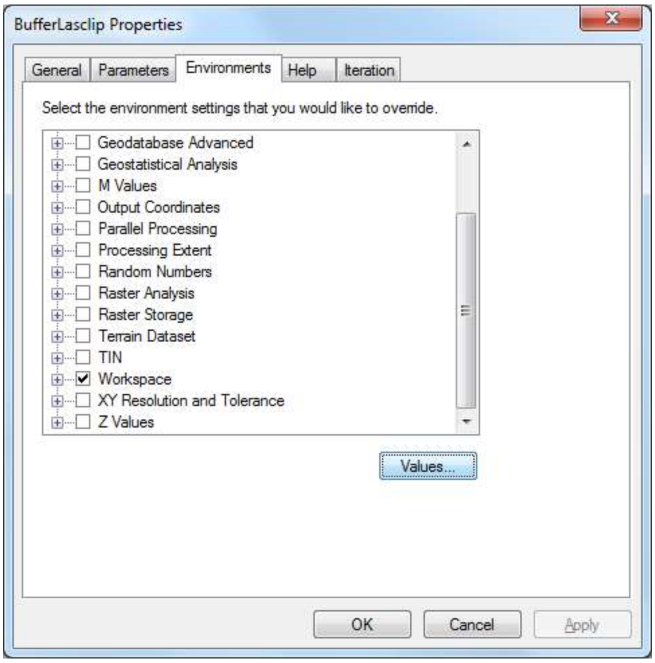
6.2 Step-by-step tutorial

To be able to perform the next steps, you should already have downloaded the VVSC toolbox as described in Chapter 1.2 "Download" and followed the Chapter 1.4 "Technical set-up".

Before starting the installation, please choose the folder in which all the results will be saved. We will call it "Results" for this example. You must have authorization to access this folder and you need to be able to save a lot of data. Depending on the LAS dataset you will use, it will require more or less space, but you need at least 10 times more available space than the size of the initial LAS dataset (so if your dataset contains 2 GB of information, you need at least 20 GB of available space).
Inside the "Results" folder, create two additional folders called "Current" and "Scratch". The Current folder will contain all the main outputs of VVSC, and the Scratch folder all the intermediate and additional/non-used outputs.

1st step: Defining the current and scratch workspace folders

- Click right on the first model and under Properties → Environments: check the box in front of Workspace. Then click on Values (see Picture 5). The environment settings window will open with one word: Workspace. After clicking on Workspace, you will be able to choose the location of the Current Workspace and Scratch Workspace. Choose the location of the Current folder you created at the really beginning of the tutorial as Current Workspace and the location of the Scratch folder as Scratch Workspace.
- Repeat this operation on each of the five models inside the VVSC Toolbox.



Select the environment settings that you would like to override.

- ☐ Geodatabase Advanced
- ☐ Geostatistical Analysis
- ☐ M Values
- ☐ Output Coordinates
- ☐ Parallel Processing
- ☐ Processing Extent
- ☐ Random Numbers
- ☐ Raster Analysis
- ☐ Raster Storage
- ☐ Terrain Dataset
- ☐ TIN
- ☑ Workspace
- ☐ XY Resolution and Tolerance
- ☐ Z Values

Picture 4: Setting up the Current and Scratch workspace within the properties of the model BufferLasclip

2^{nd} step: Starting VVSC

- Double click on the model called VVSC. The window can take some seconds to open. You should then see exactly the Fig. 3 described in the Chapter 3 “User interface”.
- Enter each of the parameters. The right side of the windows is called the Tool Help and helps you to choose your parameters wisely.
- After entering all the parameters simply click “OK” and wait. The duration of the model is highly dependent on the Input LAS dataset. The larger it is, the longer the model will take to run. If the model encounters any error, please consult the Chapter 10 “Known errors” for possible solutions.

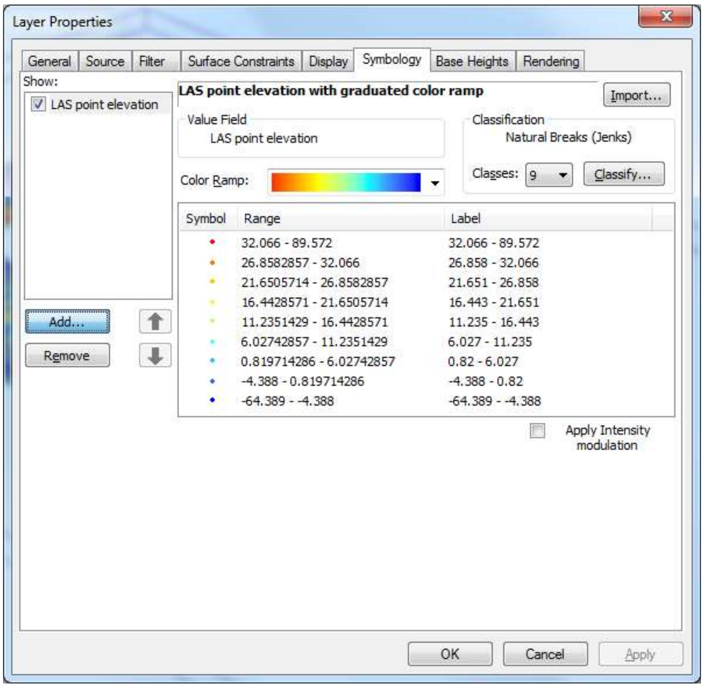
3^{rd} step: Analyzing and visualizing the outputs

- In the Catalog window, under Connect to Folder , choose the folder “Results” in which you created the folders Current and Scratch.
- Under Folder Connections, click on the path that leads to the “Results” folder, and then click on the folder Current. Inside you have three folders, one for each buffer. In each of these folders, there are 32 files.

- To display each of these files, drag them from the Catalog to the Table of Contents, or simply to the main window. To read the txt files, drag them into the Table of Contents and right-click on the text file and select "open".
- For the shapefiles, I highly recommend you to visualize them in ArcMap instead of ArcScene, in order to be able to place these shapefiles on top of an orthophoto of the area. You can display the area of each polygon by right-clicking on the shapefile and select Display Attribute table. The SUM_Area is simply the total area of this specific shapefile, so the sum of the areas of each polygon.

4th step: display the ASPRS classification of the LAS dataset

- Display the LAS dataset (.lasd) of your choice into the main window of ArcScene (by dragging it from the Catalog to the main window or to the Table of Content). It will appear with a symbology based on the LAS point elevation.
- Right-click on the lasd, properties and then Symbology (see Picture 6). On the left side of the window, click on Add → LAS attribute group with unique symbol. In LAS attribute, you can choose to display the file according to different attribute, but here we are interested in the "Classification" (You can remove all the values after 19 with the button "Remove values").
- Click on OK and the LAS dataset will be displayed according to the ASPRS classification (with the classification made by VVSC, only the classes 0, 1, 2, 5, 6, 7 are used).



☑ LAS point elevation

LAS point elevation with graduated color ramp

Value Field: LAS point elevation

Classification: Natural Breaks (Jenks)

Classes: 9

Color Ramp:

Symbol	Range	Label
●	32.066 - 89.572	32.066 - 89.572
●	26.8582857 - 32.066	26.858 - 32.066
●	21.6505714 - 26.8582857	21.651 - 26.858
●	16.4428571 - 21.6505714	16.443 - 21.651
●	11.2351429 - 16.4428571	11.235 - 16.443
●	6.02742857 - 11.2351429	6.027 - 11.235
●	0.819714286 - 6.02742857	0.82 - 6.027
●	-4.388 - 0.819714286	-4.388 - 0.82
●	-64.389 - -4.388	-64.389 - -4.388

☐ Apply Intensity modulation

Picture 5: Changing the symbology of a LAS dataset in its Properties

5th step: displaying the point density raster in 3D

- Drag the point density raster of your interest (named as pd_class#, # being the class number) into the Table of Contents.
- Then right-click on it and go into properties.
- In the label “Display”, set the transparency around 30% (to be able to see other layers).
- In the label “Symbology”, choose the color ramp of your choice. I also recommend you to check the “Display Background value 0 as no color”.
- In the label “Base Height”, under “Elevation from surfaces”, select “floating on a custom surface” and choose the max_class# (with the same class number # as the pd_class#). Under “Elevation from features”, you can change the “Factor to convert layer elevation values to scene units” to change the heights of the cones.
- In the label “Rendering”, under “Effects” check the box called “shade areal features relative to the scene’s light position”

Now that this short tutorial is completed, you should be able to extract and obtain all the information you need from the VVSC outputs. I highly recommend you to display and accumulate multiple layers into your main windows of ArcScene, in order to be able to compare the different outputs with each other.

6.3 Results

The main goal of this part is to have a look at the different outputs provided by VVSC and to study their level of accuracy and precision based on observations and comparisons with orthophotos. Each output will be presented and described as well as their qualitative relevance and reliability. Does VVSC really manage to provide so much information about our environment structure and organization? After analyzing the results, everyone should have a clear answer to this question.

6.3.1 Visualization of the LAS dataset based on the points’ heights

One of the most essential output is also the simplest: the visualization of the LAS dataset based on the points’ heights above ground (Fig. 4). The output called “Classified.lasd” is the complete LAS dataset that was used as an input in VVSC, displayed according to the point’s heights. It gives the user a proper first idea of the organization of the vegetation within the area of interest as well as its maximum height and the areas of high vegetation.

Another important role of this first visualization of the dataset is to look at the area included within the user-defined buffer. Indeed, the limits of the buffers are very clearly displayed and allow the user to check if the radius chosen was large enough to fulfill all the intended purposes.

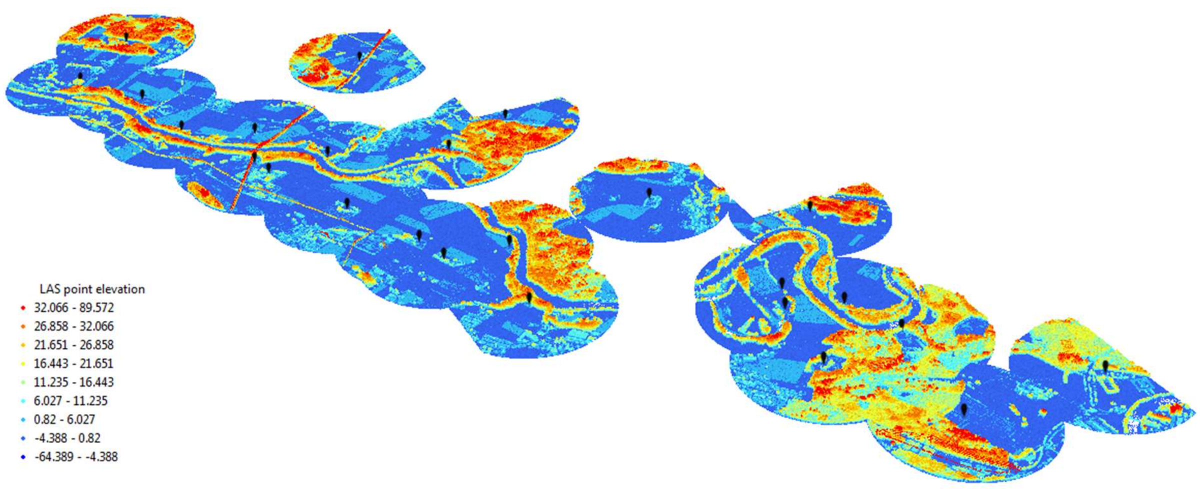


Figure 4: 3D Visualization of the input LAS dataset based on the point's elevation (meters above ground) with a user-defined buffer in ArcScene (buffer size 500 m)

Fig. 5 presents the LAS dataset called "Class6.lasd", which is a dataset composed of all the LiDAR points that fitted the user-defined parameters for ASPRS class 6 (buildings, Table 1). In this example, the parameters chosen for the sixth class selected all the points with a height ≥ 15 meters and ≤ 50 meters, with an ASPRS classification value of 3, 4 or 5 (Low, Medium or High Vegetation). It allows the user to obtain a very clear idea of the repartition of this height class of vegetation within the area of interest. For species with specific needs in terms of vegetation structure, this display of information, enhanced with additional data on the vegetation type and species, can provide a useful estimation of possible habitat.

A relatively high vegetation, between 15 and 50 meters of height, is consistently present along the river (Fig. 5). Despite not being displayed, the river is easily recognizable due to its meanders and to the trees that grow alongside. After checking with different classes of vegetation heights, the presence of vegetation along the river has always been detected, which highlights the fact that LiDAR successfully provides information about the vegetation from the canopy to the ground.

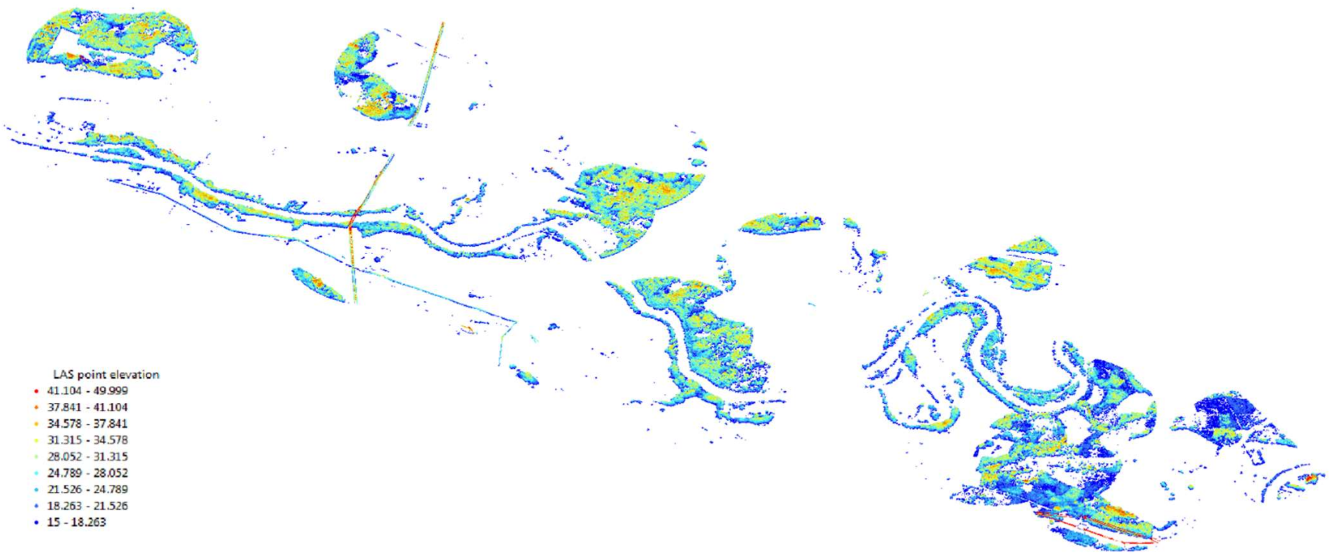


Figure 5: 3D Visualization of the input dataset's vegetation with an elevation between 15 to 50 meters with a user-defined buffer in ArcScene (buffer size 500 m)

6.3.2 Visualization of the LAS dataset based on the ASPRS classes

Fig. 6 shows the same LiDAR points as Fig. 4, but this time displayed according to the ASPRS classification. In the case of VVSC, the main advantage of this visualization is to distinguish the buildings from the vegetation allowing the user to extract more specific information on the landscape structure. In most of the studies, the buildings need to be excluded from the analysis as they represent potential sources of errors concerning the vegetation height and density. By being able to choose which ASPRS classes the user wants to use during the analysis, the buildings can be removed. To check this classification is essential in order to detect points that may be wrongly classified. In fact, this classification process is very useful as assigning each of these points manually would take an infinite time. The classification accuracy varies highly. However, some constructions such as powerlines, bridges and cranes always are an important source of misclassification (Figs. 5, 7). By checking the classification visually, the user can identify and delete all suspicious and misclassified points.

In Fig. 6, most of the crop fields are classified as ground points or as "Unassigned", which means that the model did not manage to choose one specific class for these points. The river is also classified as "unassigned" as the automatic way of classifying used in VVSC do not include the identification of the water points as well as the road points. The classification method will be discussed in Chapter 7 "Use and limitations of VVSC". Nonetheless, most of the buildings were very well classified, as well as all the areas of forest and dense vegetation. We can see that the most important features are correctly classified, but the user still needs to pay attention to incoherence or doubtful results.

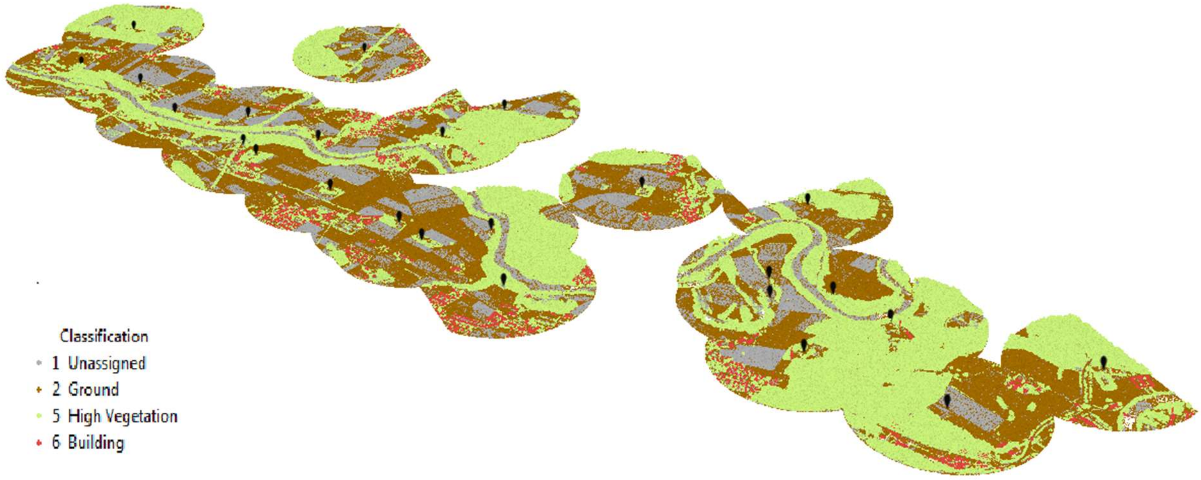


Figure 6: ASPRS Classification of the input LAS dataset in ArcScene

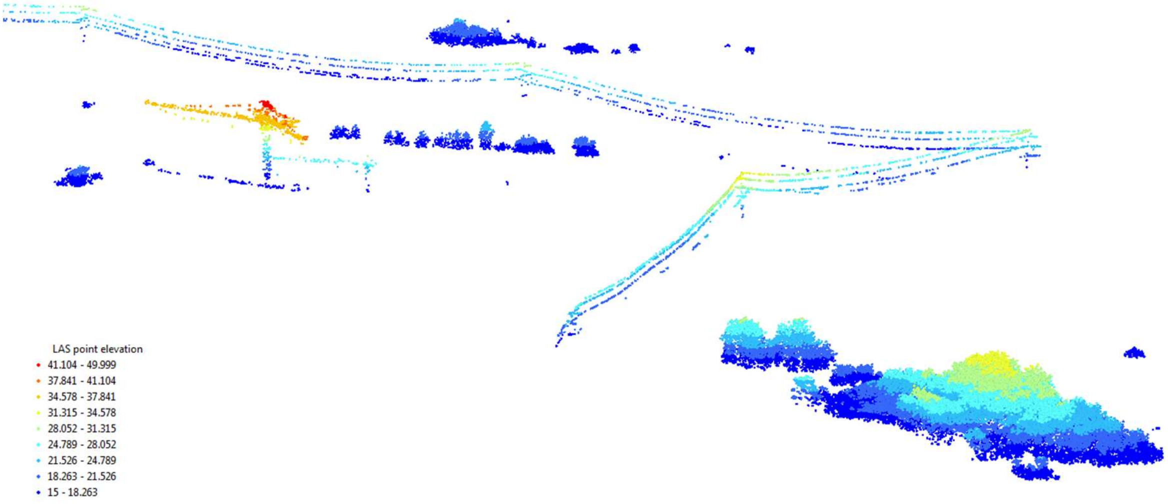


Figure 7: A powerline and a crane misclassified by VVSC clearly stand out of this visualization of the LiDAR points classified as vegetation with a height between 15 to 50 meters aboveground.

6.3.3 3D visualization of the point density raster

The raster named “pd_class” presents the results of the point density applied on each of the different chosen classes. They reveal which areas have the highest density of points within the area of interest for the given class of vegetation height (Fig. 8). Depending on the efficiency of the separation between vegetation and anthropogenic structures, a high density of points in an area means a high density of vegetation with the chosen height parameters.

The objective of the point density is to help the user identify which zones, within the area of interest, possess the user-defined parameters, and to avoid misinterpretation based only on the 3D visualization.

Indeed, small areas of high density could be missed, as well as the pattern that could explain the presence of a high density vegetation zone at a certain place. For instance, by superposing the LAS dataset and the raster of the point density, it becomes clear that most of the areas without an important presence of high vegetation are crops fields or inhabited areas. With additional information such as maps of protected areas or forested areas, further explanations and applications of the output density can be obtained, such as the growth or development of these areas.

Yet, one of the main advantage of these point densities is the possibility to superpose and compare them with each vegetation-height classes. To allow these comparisons, the results of the point density had to be transformed from absolute values to relative values. Therefore the absolute values of the point density were divided by the maximal density and multiplied by a factor of 750. This factor can be changed and adapted for each point density, as long as it provides a clear visualization of the point density in 3D. This calculation facilitates the final visualization and avoids the large variations obtained at the end of the point density due to the differences in number of points within each class.

Here, the base height is a function of the maximum, which means that the maximum value is the highest, and the relief displayed in the ArcScene window is calculated according to it. So the highest density in the user-defined class 2 reaches the same height on the ArcScene window than the maximum density of the user-defined class 6, despite their number of points per unit of area being totally different. We use the raster “max_class” as floating surface for the base height as it is the same raster as “pd_class” but with the units relative to the maximum.

However, the units displayed for the “pd_class” are still points per map unit square (in this case per m^2).

As an example, Fig. 9 presents the point density obtained for a vegetation height between 3 to 6 meters in light green and a vegetation height between 15 to 50 meters in dark green. As expected, the areas where the vegetation is between 3 to 6 meters are spread around the landscape and also appear at high density in the same locations than the higher vegetation. It is most likely the understory of the forested areas and the bushes that also grow along the river. Nonetheless, its presence near the inhabited area is also important and might reflect the small bushes, hedges and gardens of the inhabitations. Hence, all points classified as low vegetation are highly uncertain and debatable (see Chapter 7 “Use and limitations of VVSC”).

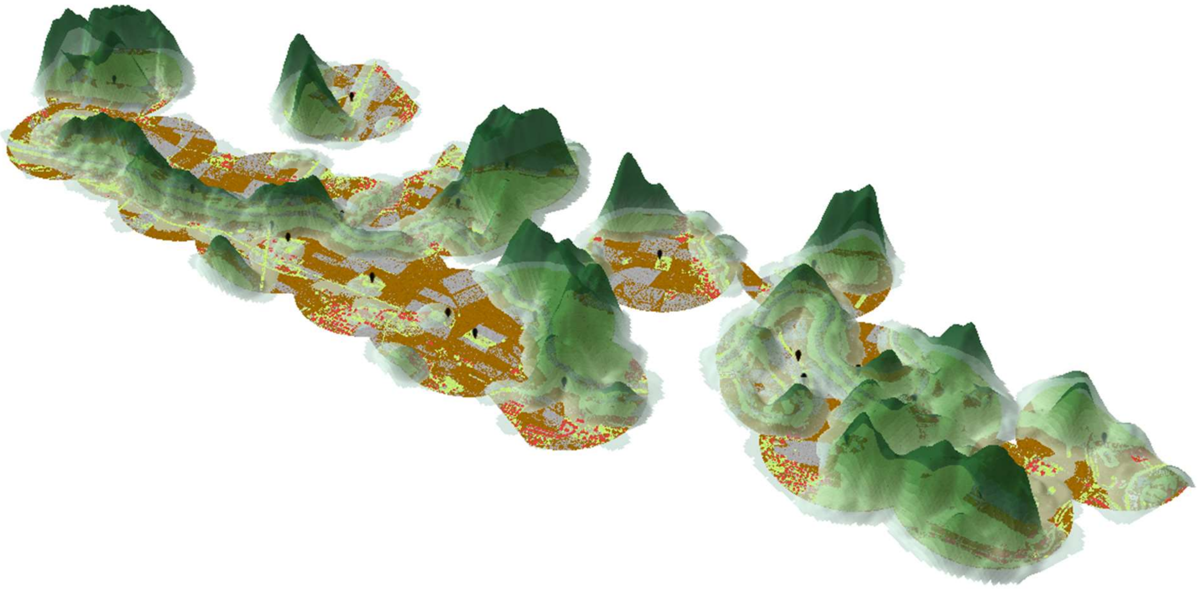


Figure 8: 3D visualization in dark green of the point density performed on the LAS dataset for a VVSC vegetation class between 15 to 50 meters of height in ArcScene (buffer size 500 m).

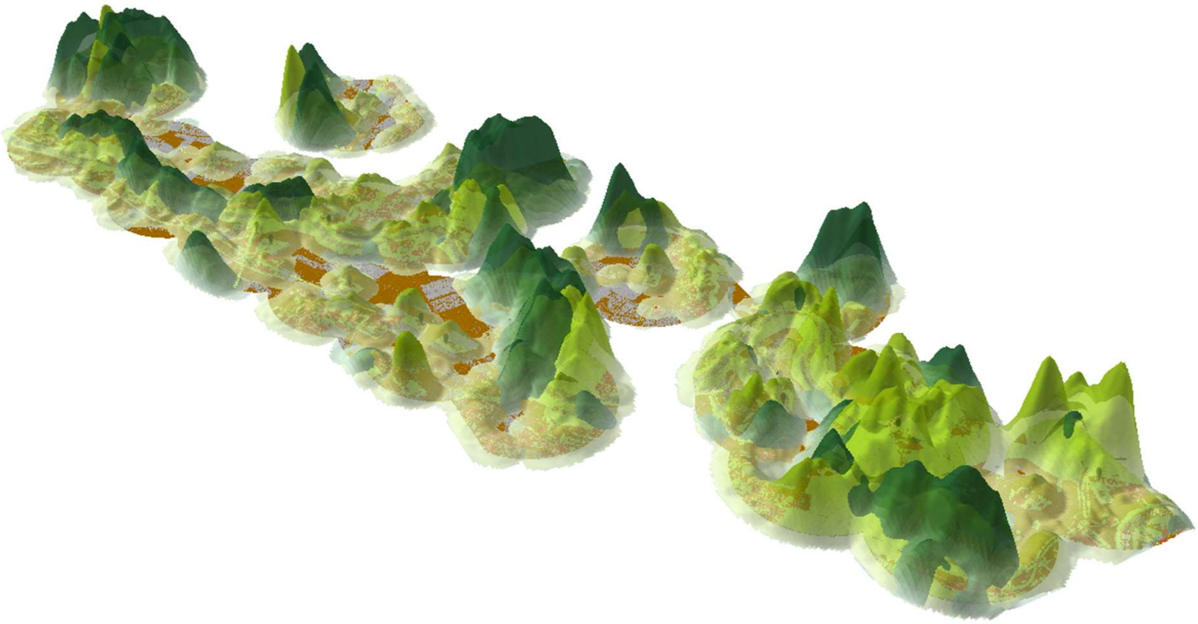


Figure 9: 3D visualization of the point density performed on the LAS dataset for a VVSC vegetation class between 3 to 6 meters of height (light green) and for a vegetation class between 15 to 50 meters of height (dark green) in ArcScene (buffer size 500 m).

6.3.4 Shapefiles of the area of vegetation

The two last formats of outputs were produced to serve additional purposes such as the 2D visualization of the distribution of the vegetation height classes over the landscape, the calculation of the area of vegetation for each vegetation-height class and the possibility to go on with further details and applications, thanks to the text files.

The shapefiles named "Area_vege_class" are the links between the 3D structure and the 2D representation of the vegetation. In ArcScene, the shapefiles appear as 3D files that represent the footprint of the points classified as vegetation. Yet, these files were created to be used in ArcMap where a chosen VVSC vegetation-height class can be spatially overlaid with an orthophoto for a visual check of accuracy. In summary, these polygon shapefiles are the area of vegetation obtained from three-dimensional information that can be displayed within ArcMap in two dimensions. The spatial overlay of VVSC vegetation-height classes with orthophotos is very useful and gives a significant feedback on the landscape context of the vegetation-height classes provided by VVSC.

For instance, Fig. 10 displays an orthophoto of the area of interest overlaid with the VVSC output "Area_vege_class6.shp". This shapefile represents the area of a vegetation-height class between 15 - 50 meters.

A visual evaluation of the overlay indicates a good agreement between the orthophoto and the VVSC vegetation-height class (Fig. 10). Even small trees or groups of trees are represented by the VVSC output. Differences can be explained by the fact that the LiDAR data was obtained in 2014 and the orthopicture was taken in 2011. Furthermore, some of the trees displayed are higher than 50 meters, or lower than 15 meters and therefore are not represented by the shapefile. There is also an important part of forest that was not identified as vegetation within one of the buffers at the right-bottom of the picture. This is due to the fact that the input LAS dataset did not encompass this area (Fig. 10; also applicable to Figs. 4, 5 and 6).

An advantage of visualizing the vegetation-height classes derived from VVSC in two dimensions is the possibility to calculate the area covered by the vegetation-height class. For example, the total area of the class in Fig. 10 encompasses 2.404 km^2 of vegetation between 15 to 50 meters height. The user can also obtain the area for one specific polygon of vegetation, such as a forest, a hedge etc.

Unfortunately, it remains some obvious misclassifications of LiDAR points into vegetation height-classes, for example roads or powerlines (Fig. 10). Here, it is suggested to filter the roads using additionally available geodatabases on land use (Switzerland, e.g., TLM) as explained in the chapter 12.3 "Annex". For powerlines, it is recommended to use additional software's specialized in the classification of point clouds data.

Figure 10: Spatial overlay of the VVSC vegetation-height class between 15 to 50 meters height with an orthophoto for a user-defined buffer of 500 m

6.3.5 Text files with numerous information about each LiDAR point

The last VVSC outputs refer to text files containing information on each of the LiDAR points included in VVSC vegetation-height class.

As these text files possess information about all LiDAR points included in each vegetation-height class such as the coordinates, the ASPRS class and the intensity (the return strength of the laser pulse), they might help the user to go further into the analysis. Moreover, these text files allow to identify all LiDAR points in the corresponding VVSC class. An R code is available to count the number of points per file and to draw a diagram of the results (See 12.1 “Annex” and Fig. 12).

7. USE AND LIMITATIONS OF VVSC

VVSC is a tool able to perform an initial ASPRS classification of the data and allocate each point within user-defined vegetation height classes. The numerous VVSC outputs provide spatially explicit information on the distribution of user-defined vegetation-height classes across the landscape. This information can be widely applied in solution-oriented natural resource management, e.g., forestry, conservation biology or landscape ecology. This section discusses decisions made while designing VVSC and assesses its use and limitations.

7.1 Software and computer requirements

First, VVSC requires a range of licensed software (LAStools, ArcGIS). It would have been preferred to create a free tool to work with LiDAR data, but the lack of free software that offers the possibilities to filter and visualize LiDAR data did not allow this. In addition, the use of LiDAR in the environmental domain is still recent and many progresses are continuously being made. So, an advantage of licensed software such as ArcGIS and LAStools include continuous updates with the latest technical improvements. The amount of computer memory space and RAM being used by LiDAR data is also very substantial and could be an obstacle to a wider utilization. Nevertheless, despite being mostly available within first world regions, LiDAR data become more and more available to the public and institutions [Bergen *et al*, 2006].

Indeed, due to the large size of the LiDAR data, VVSR uses a LAStools pipeline called “Huge file classify”, which simply classifies the points into ground, buildings and high vegetation. Despite the limited number of categories that this classification process includes, the results appeared appropriate and reasonable. Moreover, further techniques and programs can be employed to improve and correct this automatic classification to include powerlines or roads (see Chapter 12.3 “Annex” and Chapter 7.4 “Misclassifications in VVSC”).

7.2 LiDAR data

Obviously, all the limitations linked with the acquisition of the LiDAR data needs to be taken into account as well as their implications on the results obtained with VVSC. The most important limitation of the LiDAR is the fact that leaves partially absorb the signal and reflect it, which prevents most of the signal to pass directly through them. However, the fraction of the signal reflected depends on the visible wavelength used and LiDAR can still provide information on the understory by passing through small gaps or holes within the canopy cover [Bradbury *et al*, 2005]. This fact means that airborne LiDAR data often underestimate the density and quantity of vegetation within the understory, most particularly when the canopy cover is really dense and opaque [Richardson *et al*, 2011]. In addition, the high number of points observed for the higher vegetation classes is explained by the fact that most of the laser signal will have been reflected when it reaches the ground. Thus, the deeper the signal penetrates the vegetation layers, the fewer points will be extracted. Therefore it is important to notice that the higher vegetation classes obtained with VVSC are more reliable than the ones concerning vegetation classes of lower height. Indeed, measuring heights of higher trees is one of the strengths of LiDAR data [Richardson *et al*, 2011]. It is nonetheless possible that some of the highest points of a tree crown might be missed by the laser. For instance, if the peak of the tree is right between two successive measurements, the height of this tree will be underestimated [Zimble *et al*, 2003].

Another possible source of error can be observed in cases where the density of the understory covers the whole ground surface as the height of the understory will then be considered as the height of the ground [Lefsky *et al*, 2002]. Overall, the most severe underestimation of vegetation heights are observed for the density and height of the understory. Therefore, additional efforts are required to compensate for the fact that LiDAR scanners cannot always pass directly through the vegetation. Terrestrial LiDAR, combined with airborne LiDAR, possess the potential to compensate for this limitation as they provide accurate information of the understory [Richardson *et al*, 2011]. Despite their limited radius of action, they present a possible solution. In any case, the possibility to complete all LiDAR information with auxiliary data is an essential feature that allow to correct and evaluate the relevance and reliability of the results. All other optical remote sensing tools, as well as aerial images, can compensate the limitations of the utilization of LiDAR data and provide answers to other problems such as the ASPRS misclassification.

To sum up, the LiDAR points identified and categorized (without misclassification) by VVSC as low vegetation in the understory, cannot be completely trusted; the highest vegetation-height classes are probably close to reality, and the understory density is likely to be underestimated.

7.3 Selecting buffers versus analyzing the entire dataset in VVSC

Another important decision made while building VVSC was the decision to carry out the analysis within buffers, instead of using the whole area. The reasons behind this choice are various, even if the main argument remains the simplest: LAS data being so large, creating buffers is a way of reducing the size of the initial dataset, and to narrow the data down to the most important areas. Indeed, many tools used within VVSC encounter errors when the dataset is too large, which stops the whole analysis. So, the tool was developed to focus on analyzing specific areas of interest. Buffers are also an efficient way to remove useless features from the initial dataset and to push the user to choose and take decisions about the relevance of the different areas within the initial dataset. In addition, the utilization of coordinates of points where a species was observed and marked as present, fits for most of the purposes of VVSC.

Nevertheless, a guide is provided in the Chapter 12.2 “Annex” to modify VVSC to work across the whole dataset. However, as explained before, depending on the dataset size, errors may appear.

7.4 Misclassifications in VVSC

An important limitation of VVSC is not the limited ASPRS classes that are included in the classification process, but rather the misclassification of LiDAR points. In addition, many points will end up unclassified, which means that the scripts did not manage to sort any of these points into one of the different ASPRS classes. The quantity of points unclassified or misclassified will depend on the complexity and precision of the dataset as well as the type of landscape.

For instance in the Chapter 6.3.3 “3D Visualization of the point density raster”, it was observed that there was a substantial density of small vegetation within inhabited area. This might be bushes, hedges and small trees, but could also be small objects, such as fences, gates, bicycles or cars that were misclassified. These possible misclassifications cannot be avoided by an automatic classification such as the one performed by VVSC. However, it is always possible to reclassify some of the points manually even if it’s a long and laborious work, or to search for additional programs specified in ASPRS classification of LiDAR data.

In any case, in order to prevent any misinterpretations of the results provided by VVSC, it is highly recommended to have a previous knowledge of the area of interest and to go on the site and look for sources of potential errors. If, for any reasons, a trip is not possible, contacting someone with an important knowledge of the concerned landscape or searching for orthophotos, videos or pictures of the area, are both reasonable alternatives.

A visual evaluation of VVSC classes is illustrated in Fig. 11, which shows an area classified by a succession of Lasground– Lasheight–Lasclassify and an orthopicture of the same location[3]. Obvious misclassifications include bridges classified as buildings, buildings classified as vegetation and trucks classified as low vegetation. Apparently, many trucks were parked in the parking when the LiDAR data was generated in contrast to when the orthopicture was taken. On the left side of the left picture, powerlines classified as vegetation can be noticed. Powerlines as well as cranes are problematic items for any classification process (see Chapter 6.3.2 "Visualization of the LAS dataset based on the ASPRS classes"). Overall, Fig. 11 illustrates the importance to check the VVSC output by a human being with knowledge and documentation of the area.

Figure 11: On the left picture, the LiDAR points classified by VVSC as low vegetation (light green), high vegetation (dark green) and buildings (red), and on the right an orthophoto of the area of interest that appears to be a parking, road and buildings.

7.5 Signal intensity: new avenue to reclassify or identify misclassifications

An important feature provided with the LiDAR data refers to the intensity of the signal for each point in each VVSC vegetation-height class. It is the strength of the pulse when it reached the scanner after its reflection. This information offers details which might lead to even finer-scaled classifications or to the identification of misclassifications. For instance, in [Ryan, 2013], LAS points were reclassified as "Water" or "Roads" depending on their signal intensity. They calculated the Normalized Difference

---

[3] This picture does not come from the same dataset as the one of the practical example.

Vegetation Index with the signal intensity data, and based on the NVDI and signal intensity, they improved the ASPRS classification made by the LAStools of rapidlasso (e.g. points with and intensity < 8 or (intensity < 30 and NVDI > 0) were classified as water, and points with an NVDI < -0.5 were classified as roads) [Ryan, 2013].

8. VVSC – A FIRST STEP FOR FURTHER RESEARCH

The application of VVSC will allow to focus on a better understanding and quantification of the relationship between habitat or biodiversity of diverse ecological group and their 3D vegetation structure. Further, it is necessary to gain a deeper understanding of the parameters and configuration of the LiDAR and their implications and consequences on the resulting physical 3D structure of vegetation. As the efficiency of VVSC is highly dependent on the LiDAR dataset, a better comprehension of the LiDAR parameters that produce the best visualization of the vertical vegetation structure is essential. VVSC is a tool that possesses the means to help scientists and researchers to improve our knowledge of the relationship between species and the three-dimensional structure of the vegetation by testing the numerous hypothesis already formulated. Additionally, by providing spatially explicit outputs, VVSC will also improve our understanding of the spatial configuration of 3D patterns in landscapes and ecosystems. However, VVSC is the result of a six months Master Thesis and additional optimization is needed. How could VVSC be improved in the future? What additional options has VVSC to offer? What could be optimized?

First of all, a thorough quantitative assessment of VVSC's efficiency and reliability is required. As discussed in Chapter 7.4 "Misclassifications in VVSC", misclassification remains a major issue. Therefore, careful vegetation-height class evaluation is important. Here, comparisons between the results obtained with VVSC concerning the area of vegetation and the actual vegetation surface might provide an important feedback. To do so, independent geodata sources such as land-use/land-cover maps provided by Swisstopo (Swiss TLM) can be used. Another meaningful evaluation about the understory is required, as the relationship between species richness and vertical structure of vegetation is essential.

VVSC offers a broad range of outputs which can be used to evaluate the vegetation-height classes with independent data. For instance, comparing the point density as output by VVSC with the actual tree density using e.g., data from a National Forest Inventory would offer insights into the accuracy of LiDAR data. Based on the same concept, an estimation of the volume of woody vegetation for each VVSC vegetation-height class would evaluate the 3D structure assessed by VVSC.

Additional functionalities could be developed to obtain information about the key variables of the vertical organization of vegetation such as the canopy texture or the edges between forest and non-forest areas.

Also, the usage of terrestrial LiDAR or any data originating from other optical remote sensing tools, combined with airborne LiDAR will likely improve the accuracy of the 3D information, particularly for lower vegetation heights such as the understory.

For the more experienced users, an optimization of the functions used within VVSC, as well as the parameters of each function, could be made, in order to reduce the computing time, increase the efficiency of the different processes and improve the capacity of VVSC to handle large datasets.

Finally, further updates and developments will be required for VVSC to remain a useful tool for managers, practitioners and researchers in the future.

9. ACKNOWLEDGEMENTS

I would like to thank my thesis advisors, Janine Bolliger and Christian Ginzler, for their very helpful advice, support and availability at every hour of the days and weekends. I learned a lot while working at WSL and that is because of your efforts. I am very thankful and grateful for the great opportunity you offered me.

I would also like to deeply thank Achilleas Psomas for the splendid webpage he designed for me despite the limited amount of time I gave him, and more generally, the entire group of Landscape Ecology for the nice ambiance I found here at WSL.

A great thanks to Martin Isenburg who took the time to answer my questions about the LAStools and who included me into the google group “LAStools - efficient tools for LiDAR processing” where I found many answers to my problems.

Finally, I would like to express my very profound gratitude to Selina Züst for her unfailing support and patience during this long process of research and designing that was my Master Thesis.

VVSC would not have been accomplished without any of you.

10. KNOWN ERRORS

For each error encountered, I recommend to look at the ESRI support for online help and advice. There are two possible errors that can happen when running a model; the errors before the process starts, and the errors that occur during the analysis. The first type of error is due to problems in the setup of VVSC such as missing parameters or inappropriate input files. As soon as you click on “Ok” in the VVSC user-interface, a window called “error” will pop up and shows errors and warnings. It shows many warnings, but these are perfectly normal. You should only look at the first lines of this “error” window, which will tell you the real problem.

The second type of errors appears during the processing of the model. Here are possible reasons and solutions:

```
ERROR 010092: Invalid output extent.
Failed to execute (Point Density).
```

Reason: The problem comes from the extent of the output. In the case of VVSC, it often means that this specific vegetation height class is empty.

Solution: Check the Classified.lasd file, it will give you the maximum height. For instance, if the maximum height in your dataset is 45m, and one of your class regroups the vegetation between 50 to 60 meters of elevation, this error will appear. So simply adapt the vegetation height classes, or uncheck the class that causes the problem.

```
ERROR 000539: Error running expression: rcexec()
Traceback (most recent call last):
  File "<expression>", line 1, in <module>
  File "<string>", line 5, in rcexec
TypeError: float() takes at most 1 argument (2 given)
Failed to execute (Raster Calculator).
```

Reason: Depending on your region, the decimals are separated by a comma or by a point. Or in the coding language python, commas are used to separate the variables. Here, the error comes from the fact that your regional settings use a comma. For instance, the function float(0,6) will read “apply the function float to 0 and 6”, which explains why it is written “float() takes at most 1 argument (2 given)”.

Solution: unfortunately, ArcGIS bases its regional settings on your computer regional settings. So to change the regional settings of ArcGIS, you need to change your computer regional settings. In Control Panel → Region and Language → Additional settings, under Numbers, change Decimal symbol to “.”. For older versions of Windows, it is in Control Panel → Region → Additional settings.

```
ERROR 000732: Input File: Dataset C:\Users\cothereau\Documents\TRY\Scratch\Area750.las does not exist or is not supported
Failed to execute (Class (2)).
```

Reason: the reason is as mentioned, that this Area750.las does not exist or is not supported. However, In the case of VVSC, when one of the buffer is uncheck, thus when the user decided to use only one or two buffers, the program will end up as above. Indeed, all the operations concerning the first buffer are completed, and the program does not find the file to go on with the other buffers. In summary, this error is totally normal when the second or third buffer were unchecked.

Reason 2: Check previously in the code, did any other function failed? Did the Lasclip at the beginning of the process work? Because when the shapefile called “Point_Shapefile_to_Buffer” is a point shapefile instead of a multipoint shapefile, the Lasclip function reports an error but goes on until the final error showed above.

Solution: Convert the point shapefile to a multipoint shapefile. To do so, use the function “Dissolve (Data Management)”. As “input Features”, select your point shapefile; In “Output Feature Class”, select the location and the name of your future multipoint shapefile. Then in “Dissolve_Field(s) (Optional)” click on “Select All”, and finally make sure that the box in front of “Create Multipart Feature (optional)” is checked. Then click OK. Now restart the model with your new multipoint shapefile.

```
ERROR 999999: Error executing function.
("esri.Multipoint") The parameter is (or has an element that is) the wrong kind of geometry.
ERROR 010005: Unable to allocate memory.
ERROR 010067: Error in executing grid expression.
Failed to execute (Point Density (4)).
```

Reason: I got this error when running the model with an initial dataset of 3GB of size in the LAZ format, which becomes 16 GB in the LAS format. So the problem here is most likely the too large initial dataset.

Solution: I propose to thin the initial LAZ dataset with the Lasthin from the LAStools. When using the Lasthin tool, the most important parameter is the grid size, which will defined the quantity of points kept, as well as the parameter “keep”, that define which points should be prioritize. Depending on your purposes, the highest, the lowest or random points are the most suited.

ERROR 000012: C:\Users\cothereau\Documents\TRY\Current\Buffer500\Area_vegetation_500.shp already exists
Failed to execute (Rename).

Reason: As explained, the file already exists, and VVSC cannot replace or overwrite it.

Solution: There are two possibilities; Empty/delete all the files present in the Scratch and Current folders and restart VVSC. Or Change for the five models the Scratch and Current workspace in the properties of the model to new and empty folders and restart VVSC.

11. REFERENCES

Bergen, K. M., Robert G. Knox, and Sassan Saatchi. *Multi-dimensional forested ecosystem structure: Requirements for remote sensing observations*. Vol. 212778. NASA/CP-2005, 2006.

Bergen, K. M., Goetz, S. J., Dubayah, R. O., Henebry, G. M., Hunsaker, C. T., Imhoff, M. L., ... & Radeloff, V. C. (2009). Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences*, *114*(G2).

Bradbury, Richard B., et al. "Modelling relationships between birds and vegetation structure using airborne LiDAR data: a review with case studies from agricultural and woodland environments." *Ibis* 147.3 (2005): 443-452.

Brokaw, N., and R. Lent (1999), Vertical structure, in Maintaining Biodiversity in Forest Ecosystems, edited by M. Hunter, pp. 373–399, Cambridge Univ. Press, Cambridge, U. K.

Davies, Andrew B., and Gregory P. Asner. "Advances in animal ecology from 3D-LiDAR ecosystem mapping." *Trends in ecology & evolution* 29.12 (2014): 681-691.

Dubayah, Ralph O., and Jason B. Drake. "Lidar remote sensing for forestry." *Journal of Forestry* 98.6 (2000): 44-46.

Dunning, John B., Brent J. Danielson, and H. Ronald Pulliam. "Ecological processes that affect populations in complex landscapes." *Oikos* (1992): 169-175.

Forman, Richard TT. *Land mosaics: the ecology of landscapes and regions*. Cambridge University press, 1995.

Hoechstetter, S., et al. "Effects of topography and surface roughness in analyses of landscape structure—a proposal to modify the existing set of landscape metrics." *Landscape Online* 3 (2008): 1-14.

Hunter, Malcolm L. *Maintaining biodiversity in forest ecosystems*. Chapter 11 - Vertical structure pp. 373-399, Cambridge university press, 1999.
Lastools, Isenburg M., “Rapidlasso, fast tools to catch reality”. https://rapidlasso.com/lastools/ (Tuesday 19th of April 2016)

Lefsky, Michael A., et al. "Lidar Remote Sensing for Ecosystem Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists." *BioScience* 52.1 (2002): 19-30.

MacArthur, Robert H., and John W. MacArthur. "On bird species diversity." *Ecology* 42.3 (1961): 594-598.

Merrick, Melissa J., John L. Koprowski, and Craig Wilcox. "Into the third dimension: benefits of incorporating LiDAR data in wildlife habitat models." (2013).

Richardson, Jeffrey J., and L. Monika Moskal. "Strengths and limitations of assessing forest density and spatial configuration with aerial LiDAR." *Remote Sensing of Environment* 115.10 (2011): 2640-2651.

Ryan C., “Using LiDAR survey for land use classification” (2013). http://www.floodplainconference.com/papers2013/Chris%20Ryan%20Full%20Paper.pdf

Vierling, Kerri T., et al. "Lidar: shedding new light on habitat characterization and modeling." *Frontiers in Ecology and the Environment* 6.2 (2008): 90-98.

Zimble, Daniel A., et al. "Characterizing vertical forest structure using small-footprint airborne LiDAR." *Remote sensing of Environment* 87.2 (2003): 171-182.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Vertical Vegetation Structure Classifier (VVSC): A novel tool for ArcGIS.

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s): Cothereau

First name(s): Pierre

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date
Birmensdorf, 18.04.2016

Signature(s)

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

12. ANNEX

12.1 R code to count the number of points per class

This code will count the number of points per vegetation-height class, and plot the results in a bar plot (Fig. 12). Simply copy-paste this code into an R script, and modify the first line of code with the path to the folders containing all the class files as done in the example. You need to replace this Z:/Cothereau/ArcGIS/Database/n647000_268000/current by your path.

```
#Set the working directory, including the folder containing all different class files; Careful use / and not \
setwd("Z:/Cothereau/ArcGIS/Database/n647000_268000/current")
# Actual working directory path
path= getwd()
# Identify all directories in the directories
Directory <- dir(pattern = 'Buffer', ignore.case=TRUE)
#Length of directory
l = length(Directory)
#
par(mfrow=c(1,l))
#Iterate through each Buffer directory, to extract the number of points for each class in each Buffer
for(a in 1:l) {
#Enter in each Buffer directory
setwd(paste0("",Directory[a]))
# Identify all txt files in the folder
Class <- list.files(pattern = '\\.txt$')
# Create a list with the different txt files
#Class_list = lapply(Class, read.table, sep = " ")
Class_list = lapply(Class, function(x) {
  tmp <- try(read.table(x, sep = ''),silent = TRUE)
  if(inherits(tmp, "try-error"))
    return(data.frame(0,0,0,0,0,0,0))
         else
           return(tmp)
})
# Count number of classes
length = length(Class)
# Create the matrix summarizing the number of points per class
Count = matrix (nrow=length, ncol=length)
# Name the different vegetation class
colnames(Count, do.NULL = TRUE)
column_Names = character(length = length)
```

```
for (i in 1:length) {
 column_Names[i] = paste("Veg_class", i, sep = "")
  }
# Assign these names to the matrix columns
colnames(Count) = column_Names
# Count the number of points in each class
replace(Count,is.na(Count),0)
for (i in 1:length) {
  Count[i,i] = nrow(Class_list[[i]])
  if (Count[i,i] ==1) {
    Count[i,i] = 0
  }
}
Count = replace(Count,is.na(Count),0)
# Color of bars
colors = c("red", "yellow", "green", "violet","orange", "blue", "pink", "cyan")
# Plot a barplot
barplot(Count, col=colors, space=0, main = paste0("Number of points per class in ",Directory[a]), xlab = "Vegetation class", ylab = "Number of points", ylim=c(0,(max(Count))), names.arg = column_Names, axes=TRUE, las=1)
# This line display the total count as the top of the bars, change the (10/100) between 1/100 to 20/100 to move the text counting the total number of point higher or lower.
for (i in 1:length) {
text(x= i-0.5, y= Count[i,i]+(10/100)*Count[1,1], labels=as.character(Count[i,i]), xpd=TRUE)
}
setwd(paste0("",path))
}
```

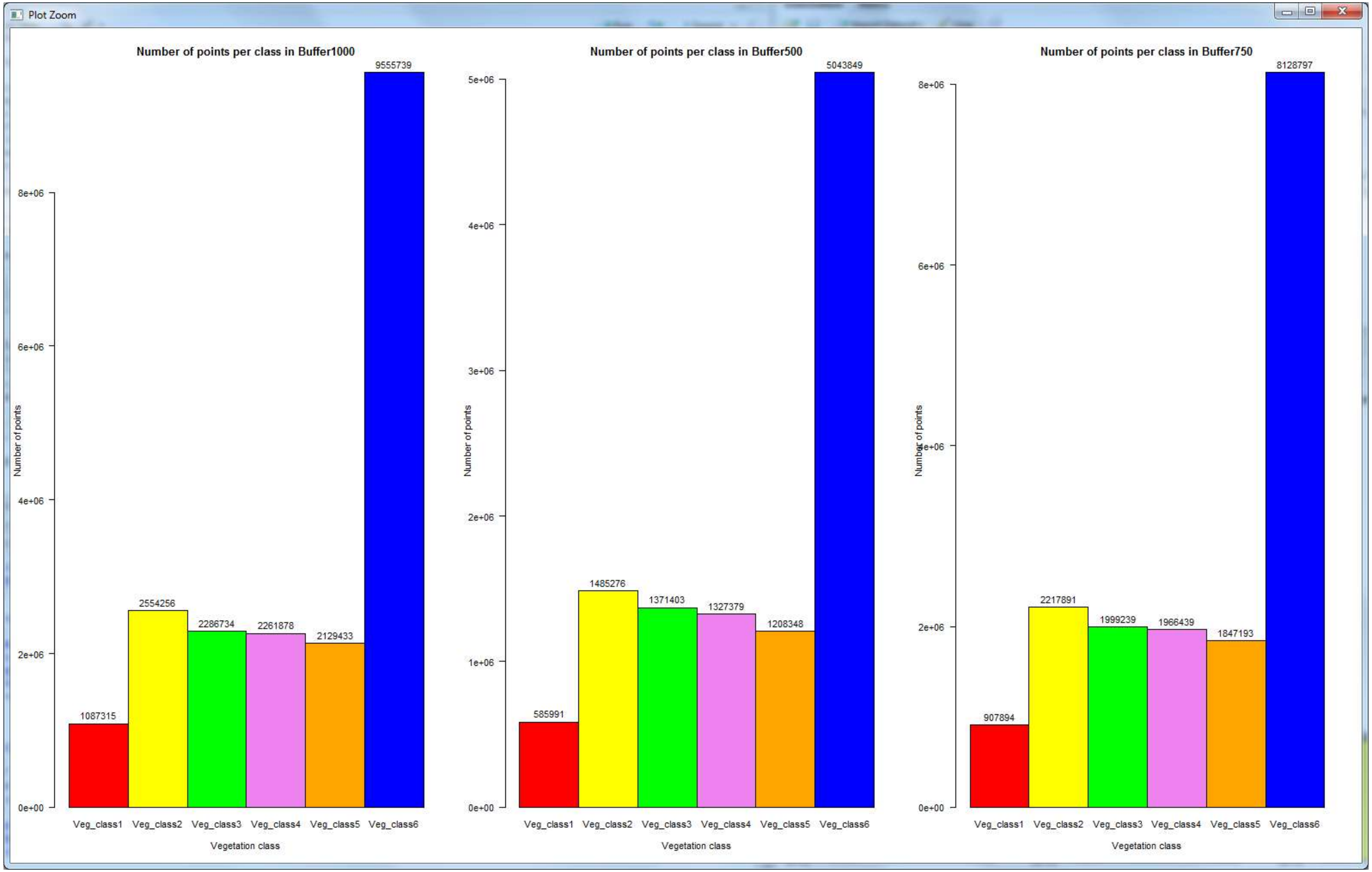


Figure 12: Bar plot obtained at the end of the R code showing the number of points within each vegetation-height class

12.2 How to modify VVSC to apply its functionality over the whole dataset

It consists mostly in deletion of all the parts that create buffers, and store them into the different folders:

- In the VVSC toolbox, right-click on the model called “Vertical Vegetation Structure Classifier”, and select Edit.
- Right-click on the submodel called “BufferLasclip”, and delete it. Repeat this operation for submodel Class (2) and Class (3). Do not forget to delete also the “Rename” function that remains after the deletion of the submodels Class (2) and Class (3)
- Right-click on the blue box called “Point_Shapefile_to_buffer” and delete it. Repeat the same operation to delete the boxes: “3th_Buffer”, “2nd_Buffer”, “First_Buffer” and “3th_Buffer_Radius”, “2nd_Buffer_Radius”, “Buffer_Radius”.
- In the toolbar, click on the “Connect” tool, then left-click on the box “Input_laz” and then connect it to the submodel called “Class” as “Input File”.

Congratulations, VVSC should now work on the whole dataset. You can also delete the functions “Create Folder (2)” and “Create Folder (3)” as they are not necessary anymore as well as the function “Calculate Value” (2) and (3), and the boxes “Workspace” (2) and Workspace (3).

12.3 How to improve the ASPRS classification with a shapefile containing the area of Water/Roads/Buildings

- In the toolbox “3D analyst tools”, under Data Management → LAS Dataset, click on the tool called “Set LAS Class Codes Using Features”.
- As Input LAS dataset, use the “Classified.lasd” and as Input Feature Class, select your shapefile containing the surface of Water/Roads/Buildings.
- Then you can add a buffer in buffer difference if you wish, and simply in the column called “New Class” choose the ASPRS classification value of the corresponding meaning. (Road Surface = 11, Water = 9, Buildings = 6)

If it does not work properly, you might need to “clip” your shapefile with one of the rasters of the LAS dataset, to fit the dimensions of the LAS dataset.