8th ESA ADVANCED TRAINING COURSE ON LAND REMOTE SENSING

Urban Mapping - Practical
Sebastian van der Linden, Akpona Okujeni, Franz Schug
Humboldt-Universität zu Berlin

Instructions for practical

Summary

The Urban Mapping Practical introduces students to the work with remote sensing data from urban areas. Students work with Sentinel-2 data from Berlin, Germany, and detailed reference information on urban impervious cover. After investigating urban Sentinel-2A spectra, the relation of the traditional normalized difference vegetation index (NDVI) and impervious surface cover is explored. Students then generate quantitative maps of impervious surface cover using both the NDVI and the full spectral information in regression approaches. Finally, the influence of spatial scale is discussed along maps of impervious surface fractions at 10 m and 20 m spatial resolution. Students gain deeper insights into the value of Sentinel-2's spectral characteristics for mapping urban areas and into quantitative mapping with regression approaches.

Data sets

The Sentinel-2 data set ***Berlin_S2A.bsq*** covers the Berlin metropolitan region plus surrounding agricultural areas and forests. It was acquired on 4 July 2015. The image was pre-processed using Sen2Cor and consists of bottom-of-atmosphere reflectance data. It is a 20 m raster data set (1700 by 1500 pixels) in WGS-84, UTM projection. Upper left corner is 377115 E, 5831335 N in UTM 33N. The data includes 9 spectral bands at 494 nm, 560 nm, 665 nm, 704 nm, 740 nm, 781 nm, 864 nm, 1612 nm, 2194 nm. Pixel size is 20 m with 10 m bands resampled to 20 m with simple averaging. The 60 m resolution bands and the 10 m nIR band have been removed. Data is stored in BSQ plus header format.

Reference information shows impervious surfaces, low and high vegetation, soils and water (***Berlin_landcover.shp***). The reference information is based on overlaying layers from the municipal urban environmental atlas and cadaster (impervious surface, plus high and low vegetation, and water). Afterwards, soil surfaces were manually digitized from very high resolution ortho-photographs. For the regression analysis, reference information was transferred into 20 m raster values matching the Sentinel-2A data. Each cell includes the fraction of impervious surface cover (***Berlin_S2A_imp_ref.bsq***). A training data set (***Berlin_S2A_imp_train200.bsq***) includes 1400 pixel (i.e. 200 with 0% impervious surface cover, 200 with 100% impervious surface cover, and 200 for each 20-% interval in-between). Data for validation is stratified accordingly and includes 700 pixel, i.e. 100 pixel per interval (***Berlin_S2A_imp_test100.bsq***).

For visual and statistical comparison of results an additional regression map at 10 m spatial resolution is provided. Results are based on a 2016 Sentinel-2A scene and the same reference data and Random Forest regression (***Berlin_10m-S2a_RFR_output.tif***). A matching reference data set with 700 test pixels is also available (***Berlin_10m-S2a_imp_test100.bsq***).

1 - First steps

Analyses are performed in the **EnMAP-Box**, a free and open source software for the analysis of spectral image data that is developed as part of the EnMAP mission preparation activities. The EnMAP-Box 3 is provided as a Python-based plugin for QGIS 3.2 or higher. See http://www.enmap.org/enmapbox.html.

Log on to the VM and **copy** the entire ***Practical*** folder from this session to your desktop.

Start QGIS 3.2 on your VM. <u>Uninstall the present EnMAP-Box Plugin and re-install the latest version by *Install from ZIP*</u>. Restart QGIS. The EnMAP-Box can be started with the button in the QGIS toolbars. The graphical user interface of the EnMAP-Box appears. Enlarge the GUI to full size. Those students with experience in QGIS will discover some similarities but also differences.

Load all data needed in today's exercise by using the "Plus" symbol in the upper left toolbar: Navigate to the data directory of today's exercise and select these eight files

Berlin_S2A.bsq	Berlin_S2A_NDVI.bsq	Berlin_landcover.shp
Berlin_S2A_imp_ref.bsq	Berlin_S2A_imp_train200.bsq	Berlin_S2A_imp_test100.bsc
Berlin_10m_S2A_RFR_output.tif	Berlin_10m_S2A_imp_test100.bsq	

Seven files appear as *Raster Data* in the *Data Sources panel*, one as *Vector Data*.

Display the Sentinel-2 data ***Berlin_S2A.bsq*** in true color: use the context menu (right-click on filename) **Open in new map > true color**. A new map window appears. It is listed as *Map #1* in the *Data Views* panel. The image may require a new data stretch BUT THIS DOES NOT WORK IN LINUX RIGHT NOW.

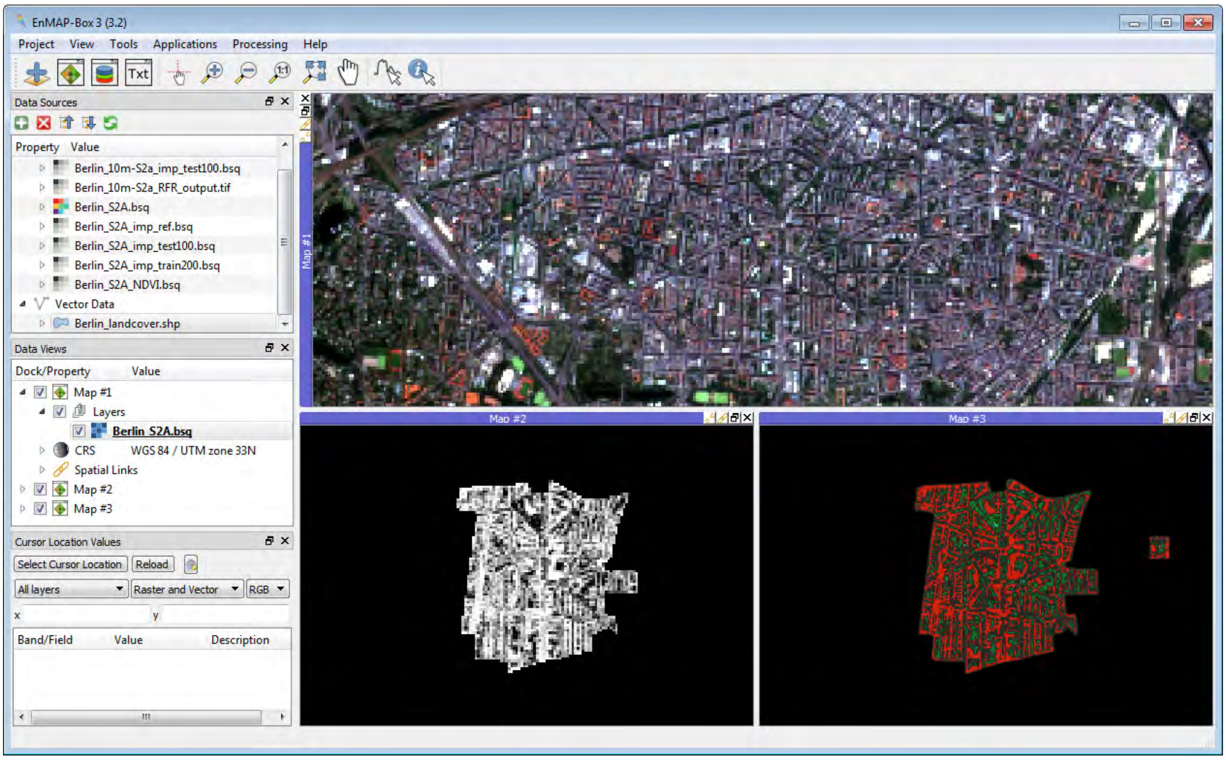
To change the band selection and grey value stretch expand the information for Map #1 in the *Data Views* panel and right-click the raster layer, **Layer Properties > Style**, select bands 4, 3 and 2 as R, G and B. **Apply.** The Berlin-Brandenburg area appears as a true-color composite.

Use the **mouse gestures** (left, middle, right button, wheel) and familiarize yourself with different options for selection etc.

Open a second map view and display the ***Berlin_S2A_imp_ref.bsq*** image (again, right-click: **Open in new map > …**). A second view (*Map #2*) and a second entry in the *Data Views* panel appear. Link the two map views by expanding the either entry in the Date Views panel and right-clicking the icon for linking. Select the option for linking *on center and scale*.

Finally, **open** the vector data ***Berlin_landcover.shp*** to a third map window and link *Map #3* with #1 and #2. Re-arrange the views by dragging the blue title bar of *Map #3* to the right edge of *Map #2*. (Compare figure on next page).

The raster with reference information was derived from the vector layer using the raster outlines of the Sentinel-2 data. High values for impervious are represented in bright-grey or white, areas with high vegetation fractions, water or open soil appear dark-grey or black. Change the tool-tip ("Identify") and the cursor location value panel to find values for individual pixels.



THIS NEXT PART DO NOT WORK TODAY ON THE VM!!! PLEASE PROCEED WITH 4

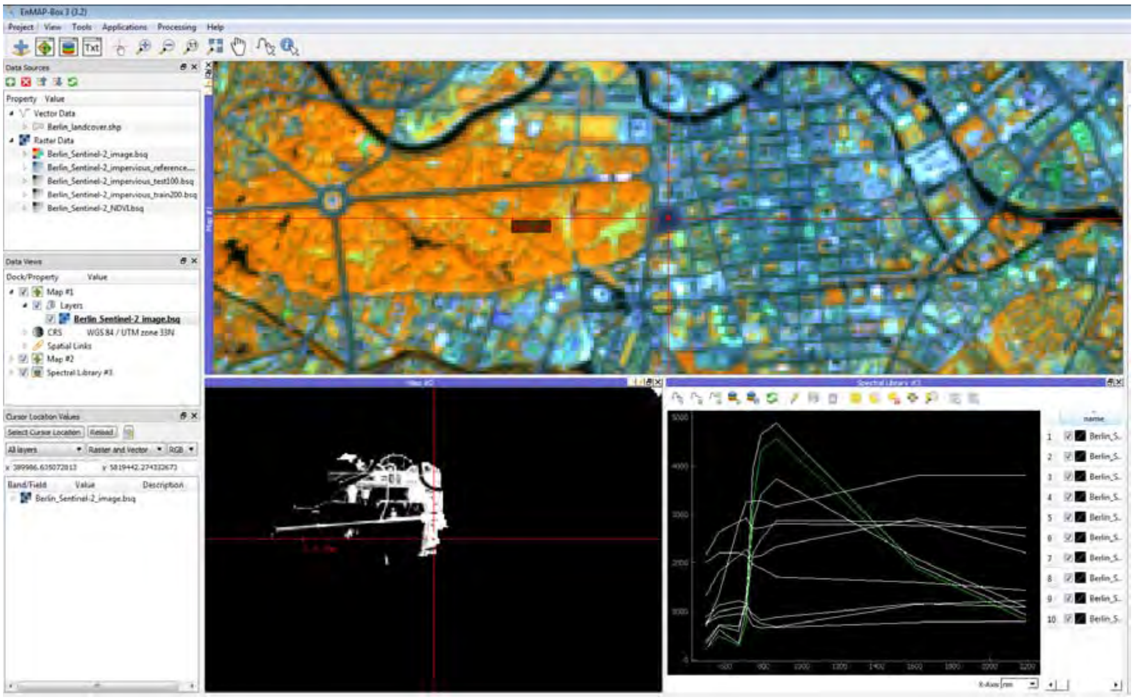
2 - Explore Sentinel-2 spectra

The 9 Sentinel-2 bands represent spectral diversity of urban surfaces well. **Open** a **Spectral Library** view (third button in toolbar). Close *Map #3* and position the new window right of *Map #2*.

Enlarge the area for drawing spectra (compare figure on next page). Now, start selecting different surfaces in the Sentinel-2 image. To do so, select the first icon in the spectral library toolbar and make sure the third icon is de-selected. Use the middle mouse button to zoom (wheel) and pan (click), and the left mouse button to select spectra. Representative spectra may be stored by using the second icon in the spectral library toolbar.

Collect a set of spectra including different variants of vegetation and impervious surface (buildings; non built-up surfaces; water).

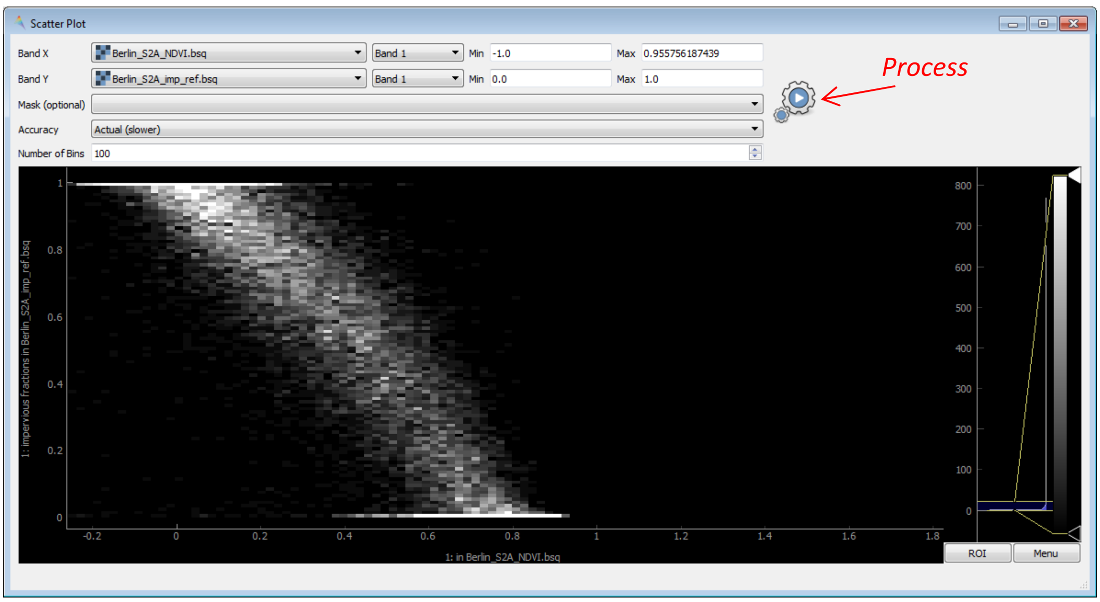
Discussion:
How do vegetated surfaces differ from impervious surfaces? How does water appear in the image data? What NDVI values do you expect for the different surface types?



3 - Exploration of NDVI and impervious surfaces

Close the spectral library window and open the NDVI image in a new map window at the same position. Which cover types appear bright in the NDVI image, which appear dark?

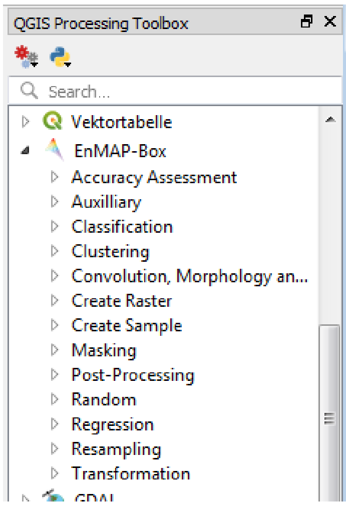
Evaluate the relationship using a scatter plot (**Main menu > Tools > Scatterplot**). Select the NDVI image and reference values. The *Accuracy* should be set to *Actual*. **Process**! Change the stretch on the right scale bar to display the densities. Mouse gestures may be used to zoom/pan to relevant plot areas.



Discussion:
How are the reference information distributed in the 0 to 1 data range?
Which value ranges are well represented by changes in NDVI, which are less good represented?
What is meant by “the NDVI saturates”?

4 - Linear regression with NDVI

The NDVI may be used to approximate impervious surface fraction at pixel level using a linear regression function. To do so, you have to fit a linear function to NDVI values and a set of training pixels, first. Afterwards, the function is used to predict impervious surface fractions for all pixels based on the full NDVI image. These functions are available in the EnMAP-Box algorithms of the QGIS Processing Toolbox.



From the QGIS Processing Toolbox select *EnMAP-Box > Regression > Fit LinearRegression*. Use the NDVI image as *Raster* input and the ***Berlin_S2A_imp_train200.bsq*** as training data (“*Regression”*). Save the regression model to your own working directory with the name ***NDVI_linReg.pkl***.

To create a quantitative map with the model use *EnMAP-Box > Regression > Predict Regression*. Select the NDVI image and the saved model ***NDVI_linReg.pkl***. Save the result as ***NDVI_linReg_output.tif***. Compare the result to the reference information visually. You may have to change the image stretch using *Layer properties > Style > Single band (QGIS)* and select 0 and 1 as *min* and *max*.

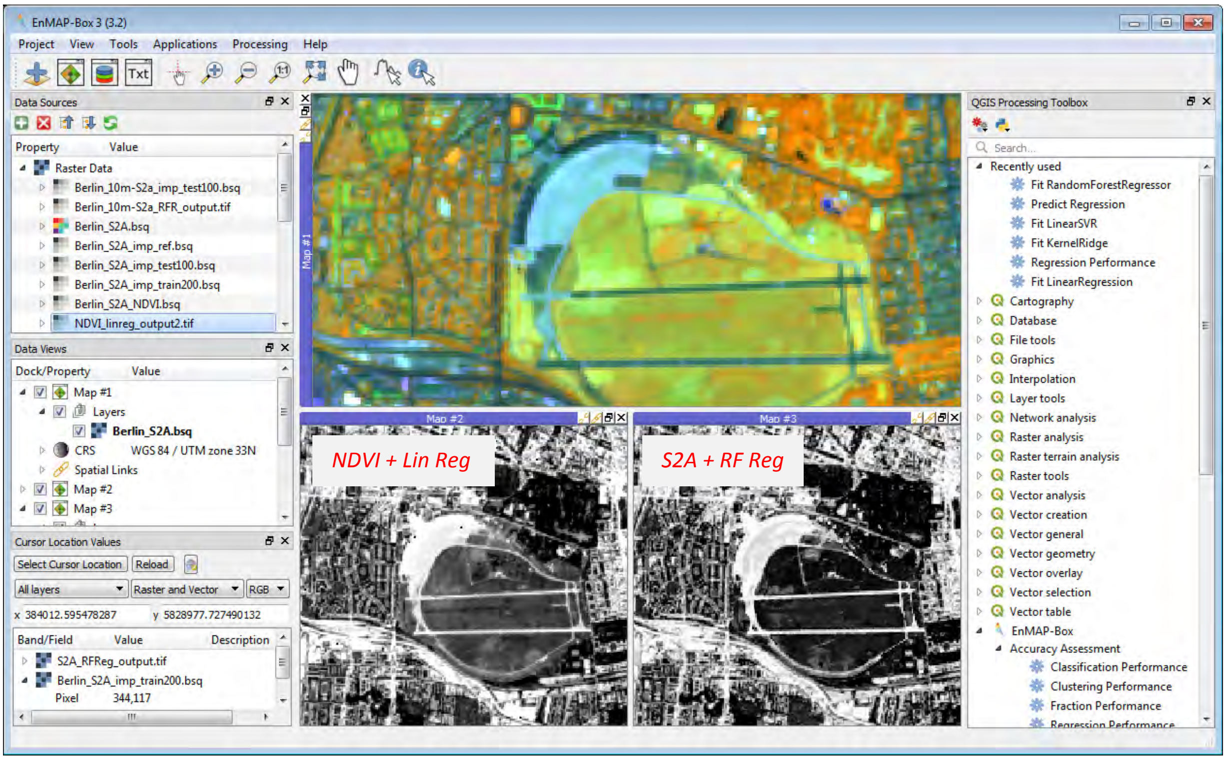
For statistical evaluation perform a quantitative accuracy assessment. Select *EnMAP-Box > Accuracy assessment > Regression Performance* and compare your output to the ***Berlin_S2A_imp_test100.bsq*** data (Note: make sure not to use the Berlin_10m_... data set!). The output for accuracy assessment is displayed in an html-report in the standard browser.

Discussion:
What values and figure does the output show?
What do these measures describe?
How do they compare to accuracy assessment from classification outputs?
How would you rate the results of your linear regression?
Are all value ranges well represented by the NDVI prediction?

5 - Random Forest regression using all spectral bands

To explore the additional value of the Sentinel-2 bands not represented in the NDVI, you will now use a Random Forest regression with all 9 spectral bands. Repeat the steps from the linear regression, but use the *Fit RandomForestRegressor* algorithm. Make sure to use useful filenames (e.g. **S2A_RFReg.pkl** and **S2A_RFReg_output.tif**) to avoid confusion.

Again, display results and perform an accuracy assessment!



You may further improve results, by repeating the model fit with a random forest of size 100. To do so, change the text window for the random forest parameters to:

```
estimator = RandomForestRegressor(n_estimators = 100)
```

Discussion:

Do you have the same results as your neighbors? Why not?
Do you see an improvement compared to the linear regression in the statistical measures?
Has the distribution in the scatter plot of observed and predicted fraction values changed?
Which surface types should be better represented when using all 9 bands?

For fast people

6 – Explore the influence of scale on mapping results

The fraction map ***Berlin_10m-S2a_RFR_output.tif*** was created using the same reference data and a random forest regression. Open the raster file in Map #2 instead of the NDVI based results. Link all maps and explore/compare the results. You will discover that the new result is at 10 m resolution.

Perform an accuracy assessment using the 10 m test pixels ***Berlin_10m-S2a_imp_test100.bsq.***

Discussion:

What is your visual impression comparing the two results?

How do they compare statistically?

What are the additional challenges when working with 10 m data and how does this explain the unexpected increase in errors?

Summary of achievements

During today's practical you have learned how to use the EnMAP-Box for visualizing and handling spectral Sentinel-2A data from an urban area. Based on reference information you have explored the inverse relation of NDVI and impervious cover fraction. In regression approaches you have utilized the spectral information to generate a continuous map of impervious surface fraction. Finally, you have learned how to perform and interpret quantitative accuracy assessments for regression maps and explored the challenge of working with very high resolution data in an urban environment.