# WorldView Biophysical Variables Retrieval over Antwerp and Valencia

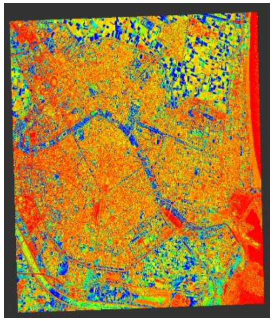
J. Verrelst, L. Alonso, P. Urego S. Belda, P. Morcillo, J.P. Rivera-Caicedo

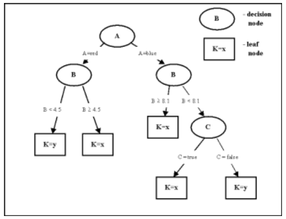
HYPERCITY final meeting

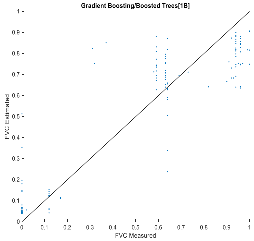
25/03/2019

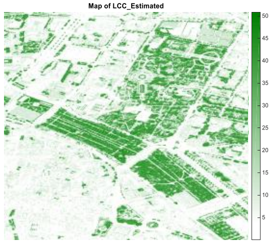
# Overview

- **Rationale**: mapping of biophysical variables from high-resolution WV data
- **Inputs**: empirical training data & NDVI WV images
- **Methods**: machine learning regression algorithms (MLRAs)
- **MLRA results & Maps**: LCC, LAI, FVC
- **Conclusions & recommendations**





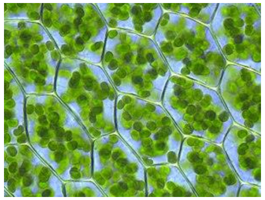



# Rationale

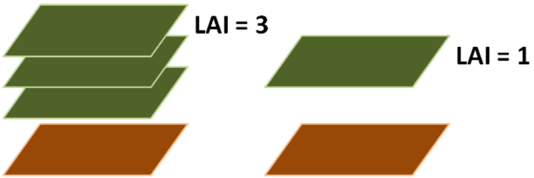
*Can WV images be used for quantitative mapping of biophysical variables of urban vegetation for the cities Antwerp and Valencia?*

**Biophysical variables:**

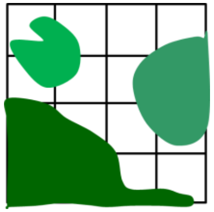
1. Leaf clorophyll content (LCC): µg/cm$^2$



2. Leaf area index (LAI): m$^2$/m$^2$




3. Fractional vegetation cover (FVC): [0-1]



# Inputs

- **Empirical variables+hyperspectral data**: SPARC03
  - ✓Around 120 Field measurements LCC, LAI, FVC
  - ✓Spaceborne CHRIS acquisition (400-1000 nm)
  - ✓To account for structure urban vegetation, additional 7 forest LAI samples and 4 fully covered samples (FVC=1) were added to training data.
  - ✓Additional 30 non-vegetated samples were added (LCC, LAI,FVC=0) to training data.

- **WV images:**

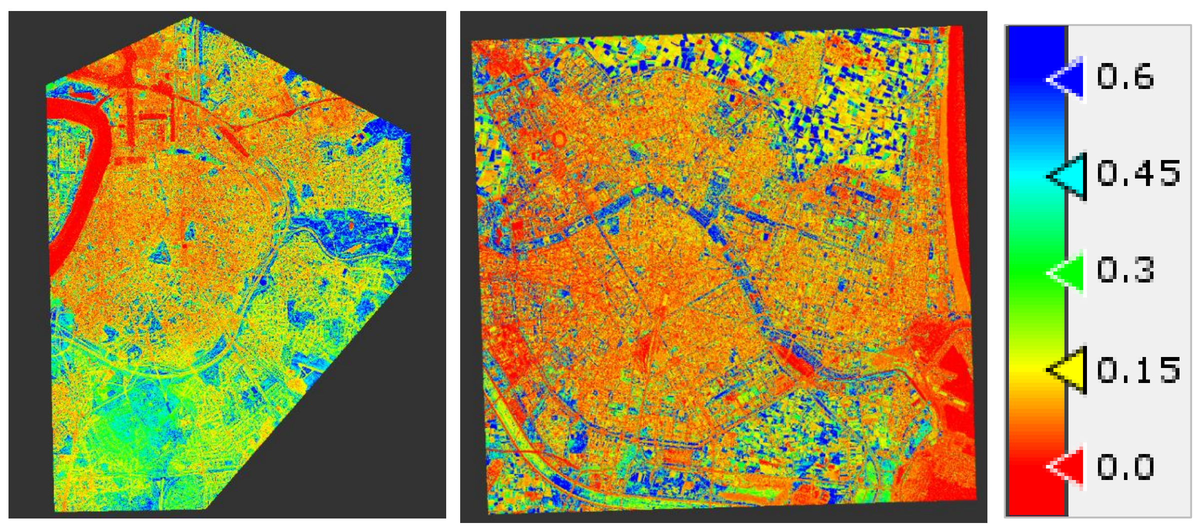
*Table 1 Description of WV images taken over Antwerp.*

| Sat ID | Image Ref. | Image Date | Cloud Cover | View Angle (°) | Time of acquisition |
|---|---|---|---|---|---|
| WV-2 | P005 | 15/04/2015 | 0.00% | 29.10 | 10:52:19.78 |
| WV-3 | P003 | 30/09/2015 | 0.00% | 28.70 | 11:18:23.00 |
| | P002 | 07/05/2016 | 0.00% | 27.20 | 10:55:28.11 |
| | P004 | 10/09/2016 | 0.70% | 29.30 | 11:17:03.00 |
| | P001 | 15/09/2016 | 0.00% | 15.20 | 10:59:41.32 |

*Table 2 Description of WV images taken over Valencia.*

| Sat ID | Image Ref. | Image Date | Cloud cover | View Angle (°) | Time of adquisition |
|---|---|---|---|---|---|
| WV-3 | P005 | 25/03/2016 | 0.00% | 13.30 | 11:17:28.05 |
| | P007 | | 0.00% | 15.70 | 11:17:16.60 |
| | P006 | 09/08/2016 | 0.00% | 24.50 | 11:02:22.96 |
| | P004 | 09/05/2017 | 0.00% | 17.60 | 11:12:54.07 |
| | P003 | 10/06/2017 | 0.00% | 17.0 | 11:27:17.24 |
| | P002 | 10/07/2017 | 0.00% | 17.0 | 11:10:11.91 |
| | P001 | 24/08/2017 | 0.00% | 22.90 | 11:31:37.76 |
| | P008 | | 0.00% | 26.30 | 11:31:55.16 |

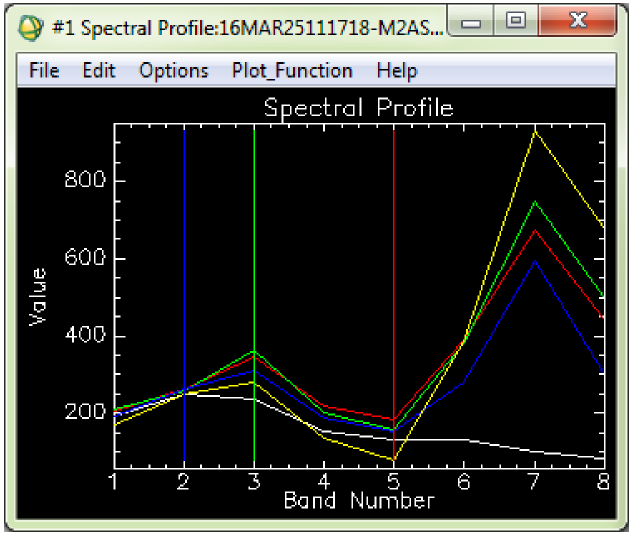
**NDVI WV-3 P004 (10/09/2016)** **NDVI WV-3 P004 (09/05/2017)**



# WV data quality for biophysical variables mapping

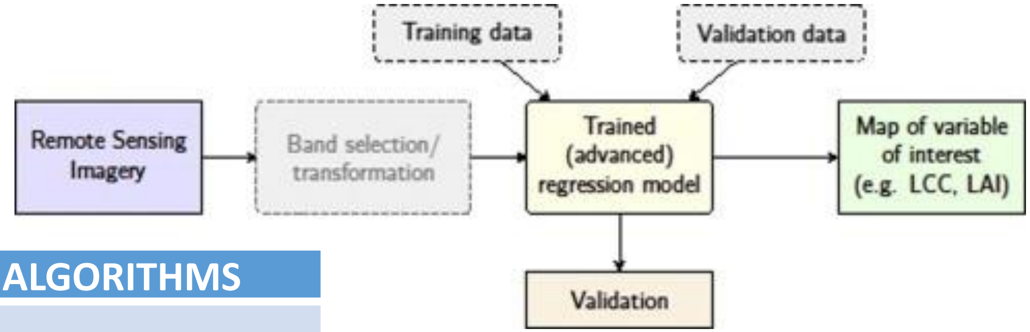
**Zoom-in**


**Some examples spectral data**

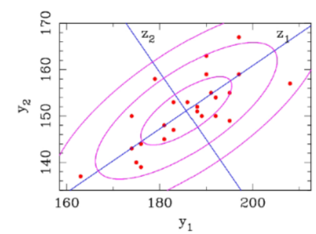


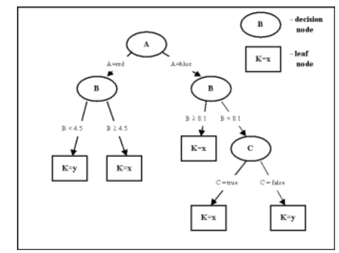

- No atmospheric correction has been applied
- Initial attempts to adjust training data to all spectral bands was only partly succesful
- Using only NDVI appeared to be more successful: spectral data of training data resampled to NDVI
- All NDVI images were processed

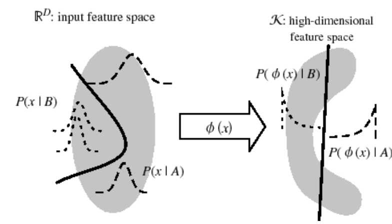
# Machine learning regression algorithms (MLRAs)

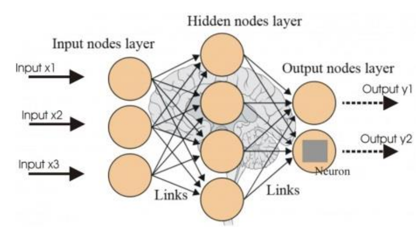


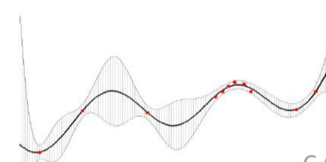

| | MACHINE LEARNING REGRESSION ALGORITHMS |
|---|---|
| 1 | Least squares linear regression |
| 2 | Principal components regression |
| 3 | Twin Gaussian process |
| 4 | Gradient Boosting/Boosted Trees |
| 5 | Regression tree |
| 6 | Bagging trees |
| 7 | Boosting trees |
| 8 | Canonical Correlation Forests |
| 9 | Relevance vector Machine |
| 10 | Extreme Learning Machine |
| 11 | Support Vector Regression |
| 12 | Kernel ridge Regression |
| 13 | Neural Network |
| 14 | K-nearest neighbors regression |
| 15 | Weighted k-nearest neighbors regression |
| 16 | Gaussian Processes Regression |
| 17 | VH. Gaussian Processes Regression |



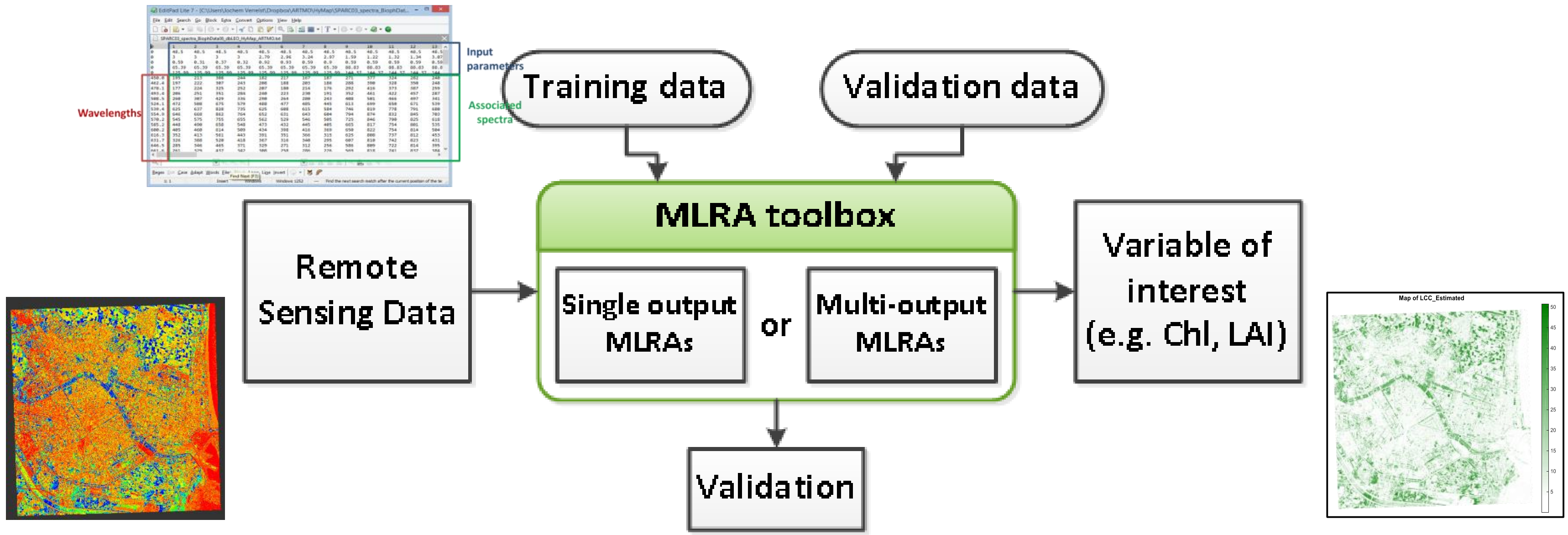





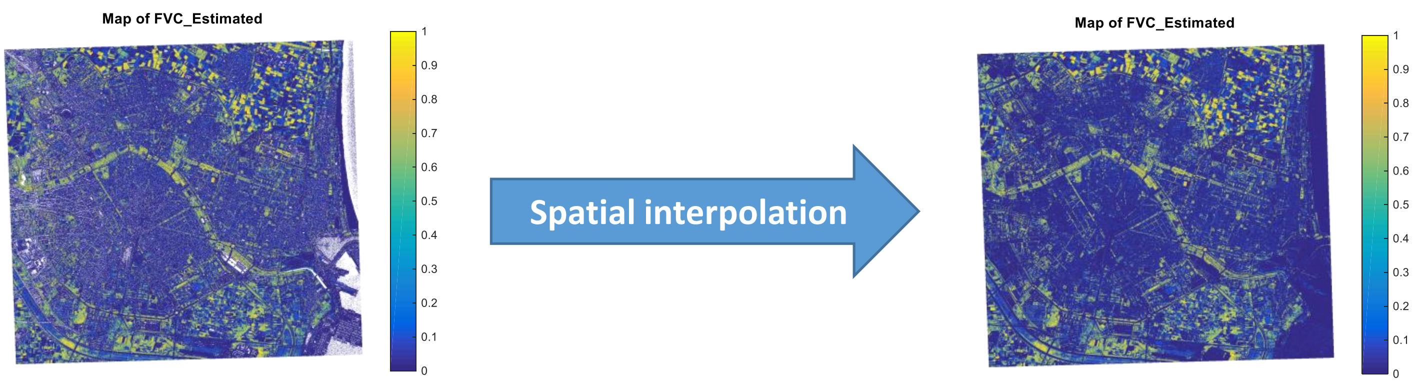



# ARTMO MLRA's toolbox




To ensure robust models, a cross-validation (CV) of k-fold=3 was applied.

**Post-processing**: 0 NDVI vlaues led to NaNs. They have been cleaned up with linear interpolation.



# Model CV results:

## LCC

| MLRA | RMSE | NRMSE (%) | R2 | Time [s] |
|---|---|---|---|---|
| **Weighted k-nearest neighbors regression** | **7.502** | **14.530** | **0.862** | **0.032** |
| Gradient Boosting/Boosted Trees | 7.608 | 14.734 | 0.858 | 0.077 |
| Principal components regression | 7.702 | 14.916 | 0.853 | 0.003 |
| *Least squares linear regression* | *7.702* | *14.916* | *0.853* | *0.002* |
| VH. Gaussian Processes Regression | 7.727 | 14.964 | 0.852 | 0.883 |
| K-nearest neighbors regression | 7.786 | 15.079 | 0.852 | 0.031 |
| Bagging trees | 7.832 | 15.169 | 0.849 | 1.081 |
| Canonical Correlation Forests | 7.888 | 15.276 | 0.847 | 1.943 |
| Gaussian Processes Regression | 7.904 | 15.308 | 0.845 | 0.192 |
| Relevance vector Machine | 7.927 | 15.353 | 0.844 | 12.682 |
| Regression tree | 8.557 | 16.573 | 0.823 | 0.008 |
| Boosting trees | 8.564 | 16.587 | 0.822 | 1.029 |
| Support Vector Regression | 8.585 | 16.627 | 0.826 | 2.259 |
| Neural Network | 8.662 | 16.776 | 0.818 | 3.622 |
| Extreme Learning Machine | 8.689 | 16.829 | 0.814 | 4.590 |
| Kernel ridge Regression | 11.307 | 21.899 | 0.734 | 0.028 |
| Twin gaussian process | 15.384 | 29.794 | 0.827 | 1.837 |

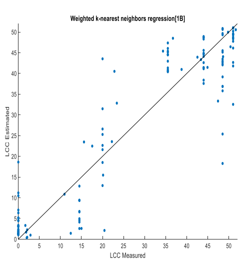
## LAI

| MLRA | RMSE | NRMSE (%) | R2 | Time [s] |
|---|---|---|---|---|
| **Neural Network** | **0.834** | **8.968** | **0.851** | **4.029** |
| Gradient Boosting/Boosted Trees | 0.835 | 8.973 | 0.850 | 0.071 |
| Bagging trees | 0.846 | 9.097 | 0.845 | 1.151 |
| Canonical Correlation Forests | 0.847 | 9.104 | 0.845 | 1.811 |
| Extreme Learning Machine | 0.903 | 9.712 | 0.823 | 8.486 |
| Gaussian Processes Regression | 0.904 | 9.718 | 0.823 | 0.182 |
| Boosting trees | 0.913 | 9.818 | 0.822 | 1.057 |
| Regression tree | 0.915 | 9.838 | 0.825 | 0.008 |
| Weighted k-nearest neighbors regression | 0.921 | 9.907 | 0.819 | 0.027 |
| Kernel ridge Regression | 0.936 | 10.063 | 0.814 | 0.028 |
| Support Vector Regression | 0.944 | 10.146 | 0.808 | 4.704 |
| VH. Gaussian Processes Regression | 0.956 | 10.280 | 0.803 | 0.842 |
| Relevance vector Machine | 0.979 | 10.522 | 0.797 | 13.292 |
| K-nearest neighbors regression | 0.996 | 10.709 | 0.785 | 0.026 |
| Principal components regression | 1.028 | 11.057 | 0.771 | 0.002 |
| *Least squares linear regression* | *1.028* | *11.057* | *0.771* | *0.002* |
| Twin gaussian process | 1.540 | 16.561 | 0.789 | 1.788 |

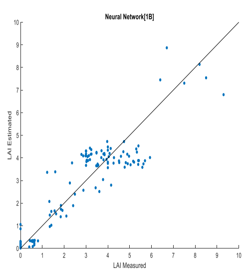
## FVC

| MLRA | RMSE | NRMSE (%) | R2 | Time [s] |
|---|---|---|---|---|
| **Gradient Boosting/Boosted Trees** | **0.156** | **15.645** | **0.834** | **0.473** |
| Canonical Correlation Forests | 0.158 | 15.761 | 0.829 | 1.699 |
| Weighted k-nearest neighbors regression | 0.158 | 15.803 | 0.828 | 0.026 |
| Bagging trees | 0.160 | 16.006 | 0.823 | 1.199 |
| Kernel ridge Regression | 0.162 | 16.234 | 0.818 | 0.028 |
| VH. Gaussian Processes Regression | 0.165 | 16.538 | 0.811 | 0.768 |
| Boosting trees | 0.166 | 16.547 | 0.815 | 1.056 |
| Gaussian Processes Regression | 0.166 | 16.595 | 0.809 | 0.188 |
| Relevance vector Machine | 0.167 | 16.741 | 0.806 | 13.214 |
| K-nearest neighbors regression | 0.169 | 16.872 | 0.803 | 0.026 |
| Support Vector Regression | 0.172 | 17.151 | 0.798 | 6.038 |
| Regression tree | 0.176 | 17.568 | 0.792 | 0.008 |
| *Least squares linear regression* | *0.178* | *17.769* | *0.782* | *0.002* |
| Principal components regression | 0.178 | 17.769 | 0.782 | 0.003 |
| Neural Network | 0.183 | 18.277 | 0.770 | 3.886 |
| Extreme Learning Machine | 0.264 | 26.384 | 0.629 | 5.956 |
| Twin gaussian process | 0.728 | 72.842 | 0.712 | 1.900 |

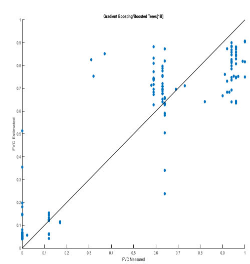
**LCC**



**LAI**

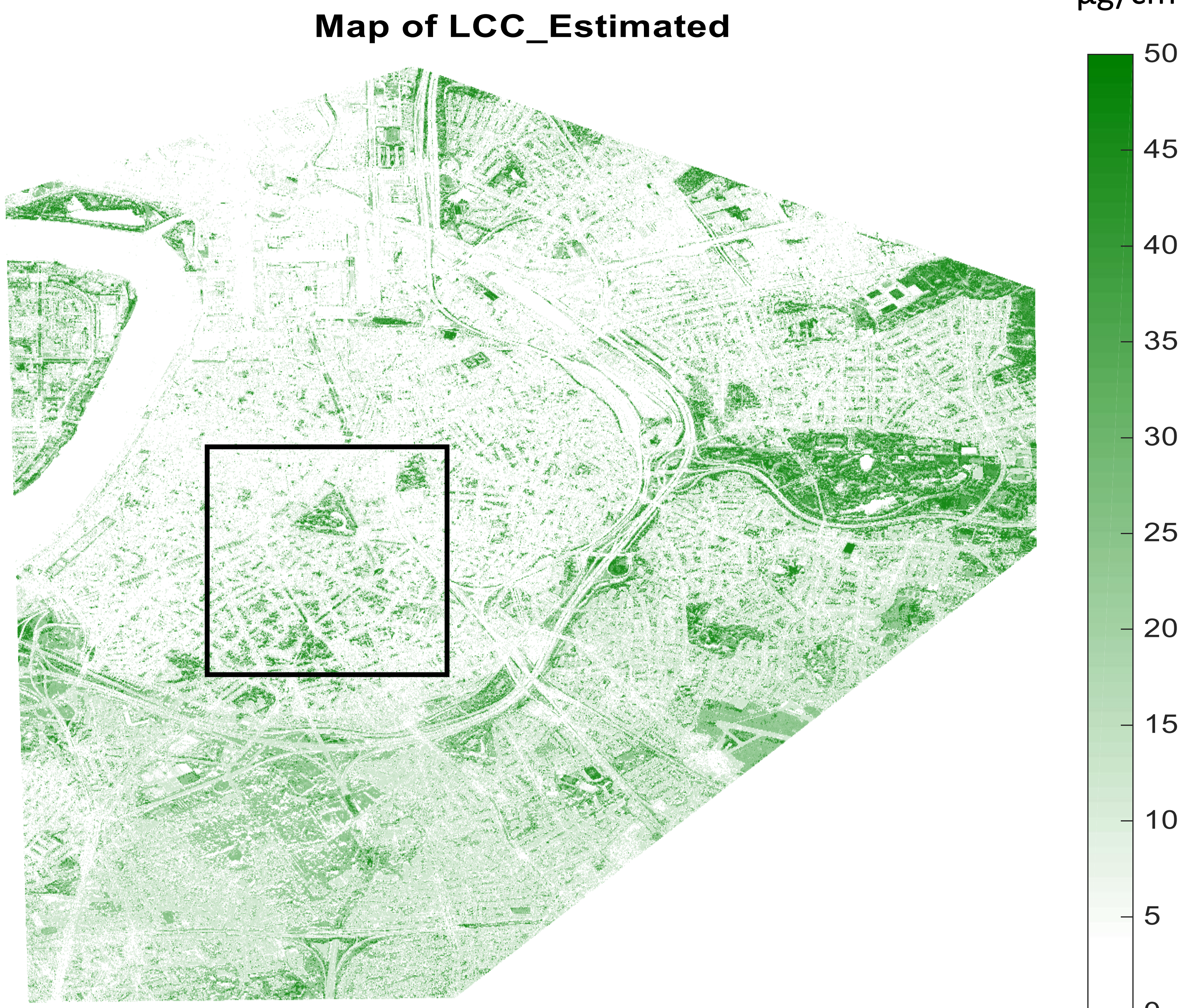


**FVC**

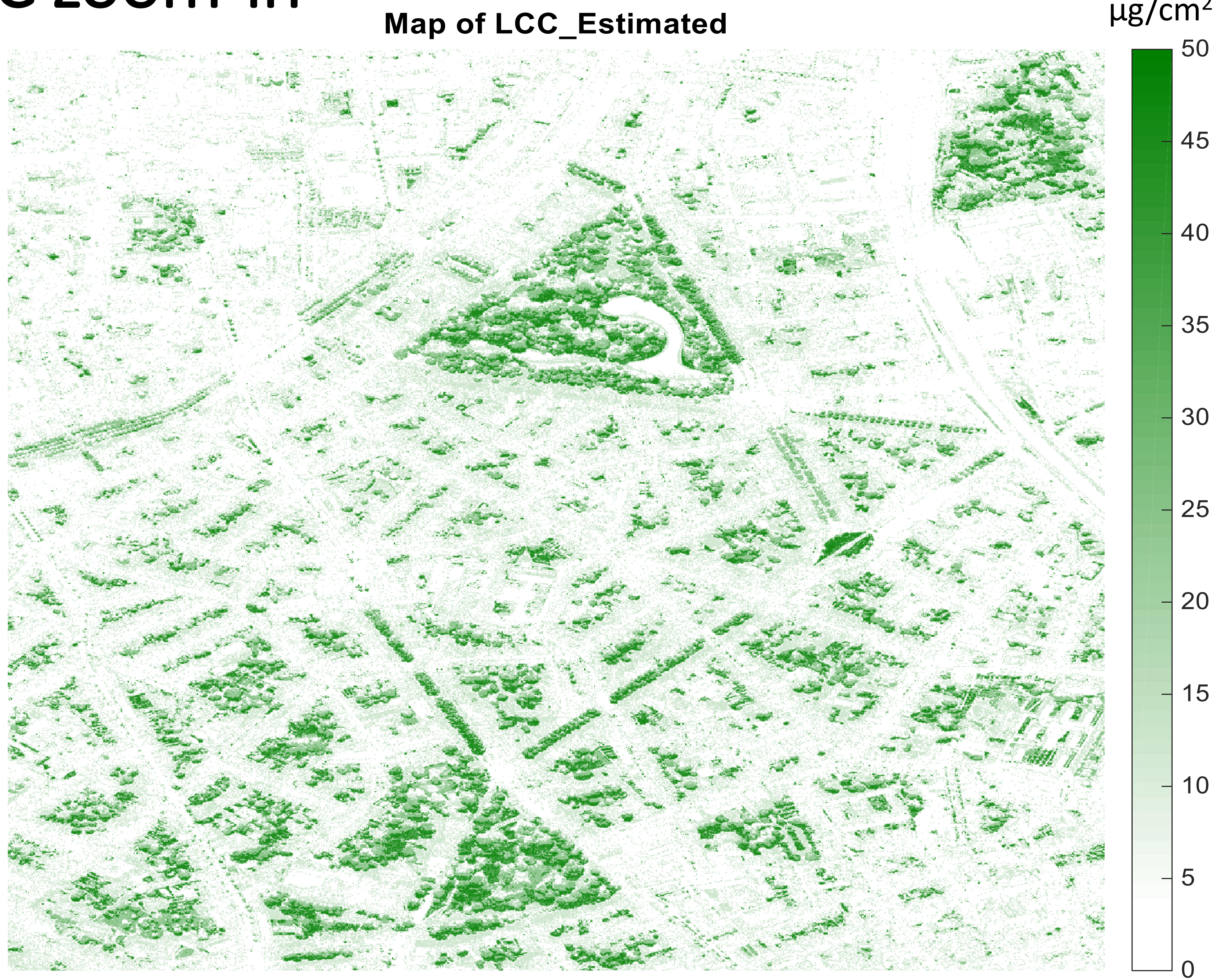


Although adequately validated, training data suboptimally distributed.

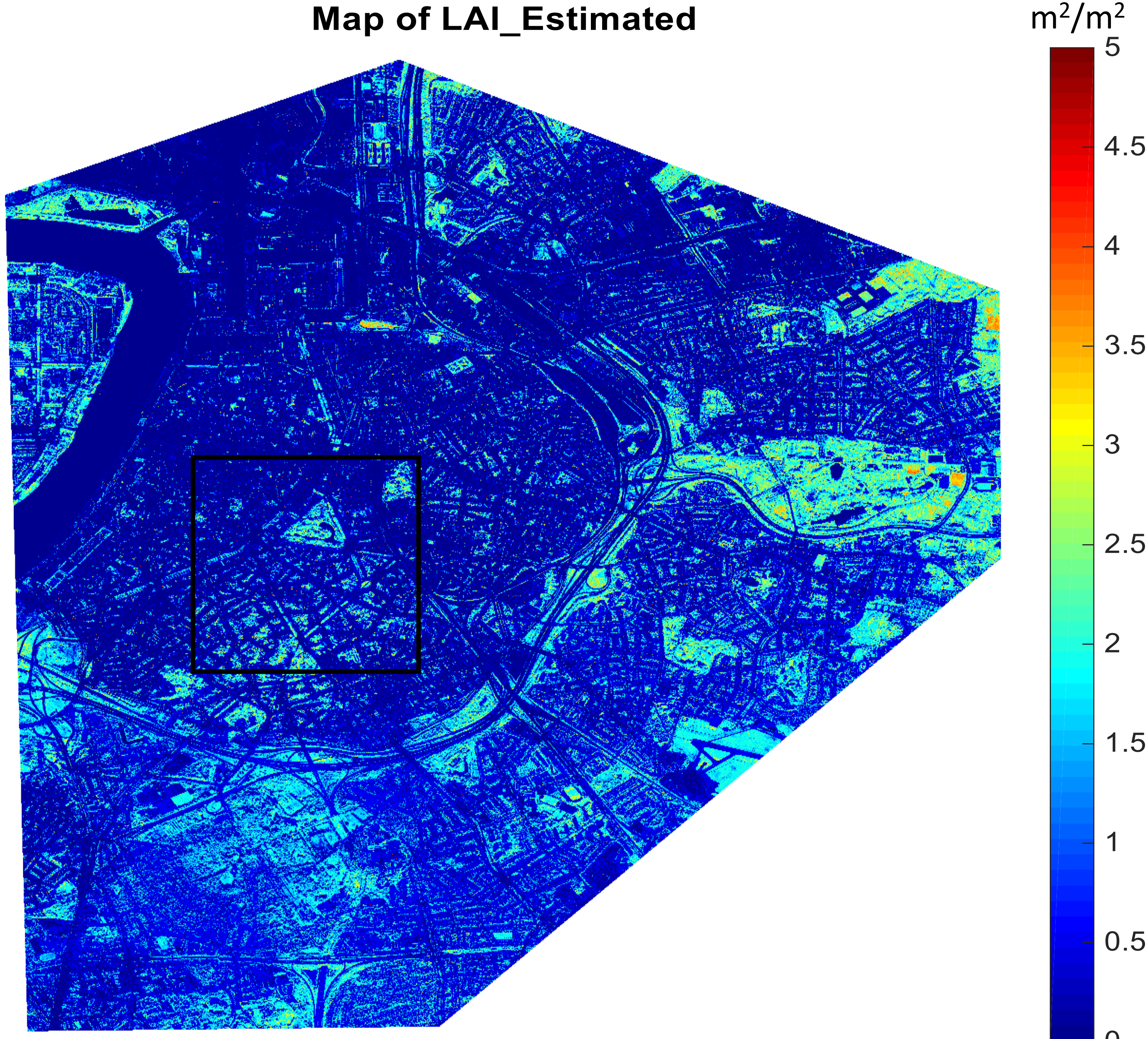
# Final maps: Antwerp LCC



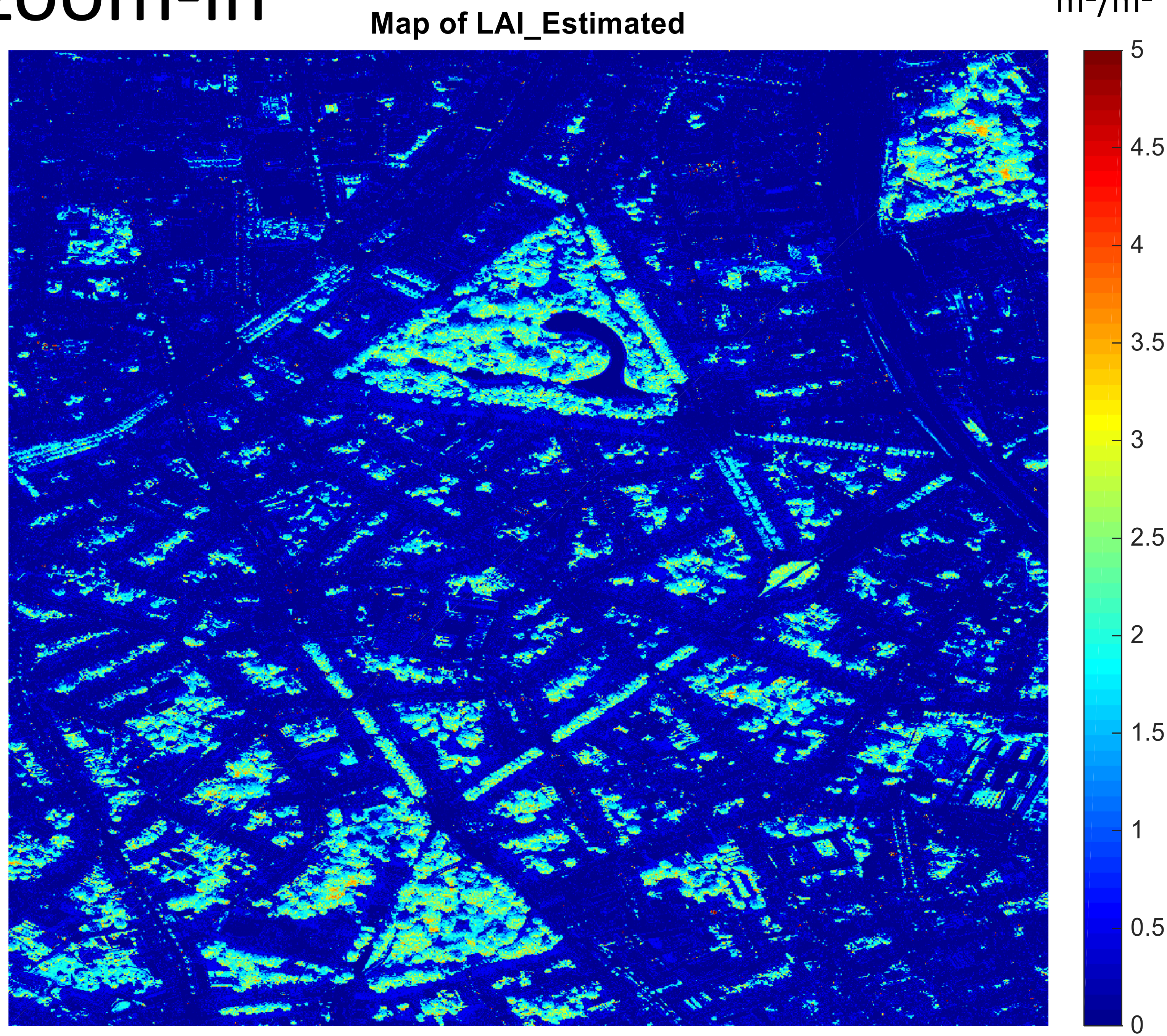
# LCC zoom-in



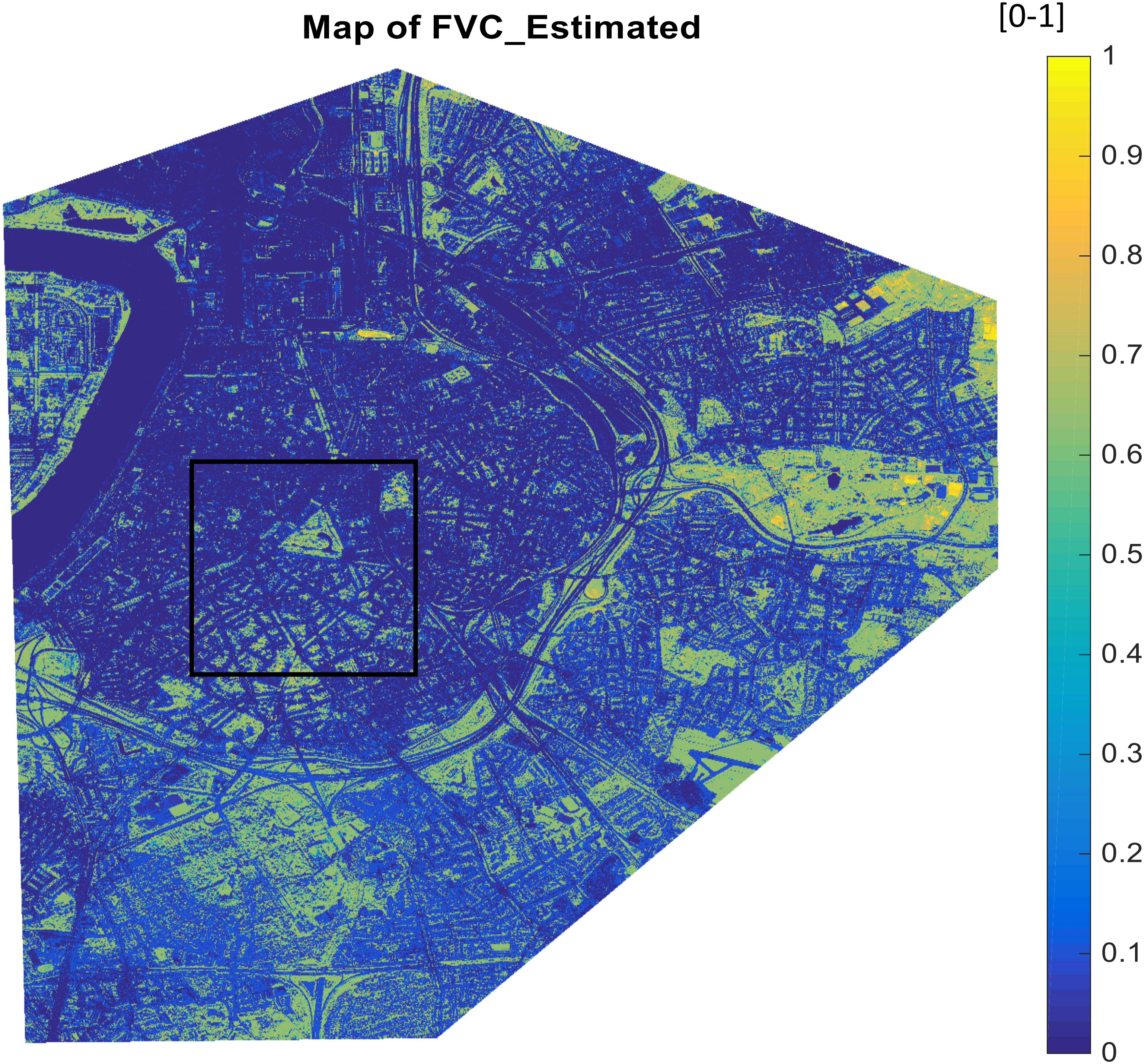
# LAI



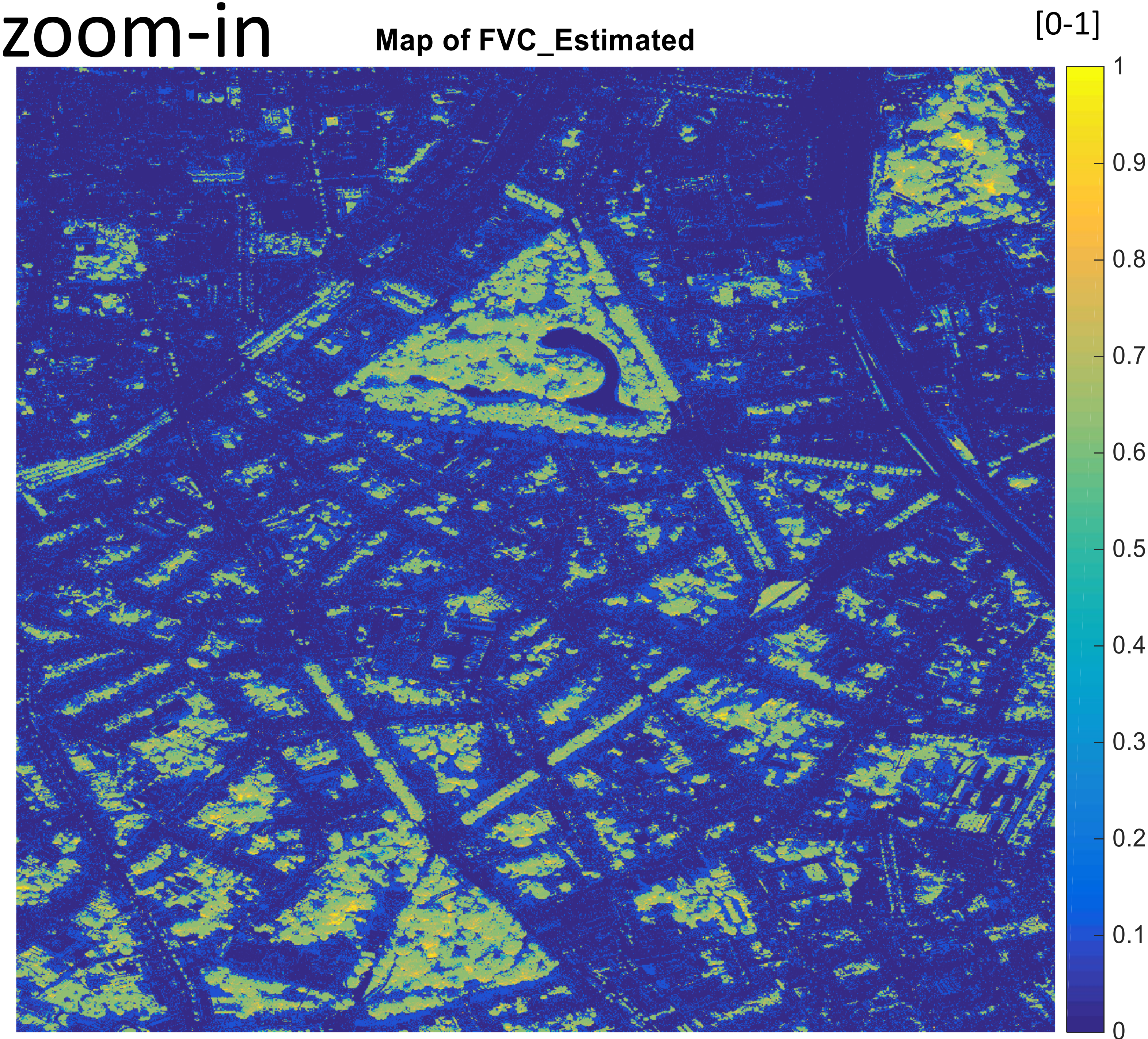
# LAI zoom-in



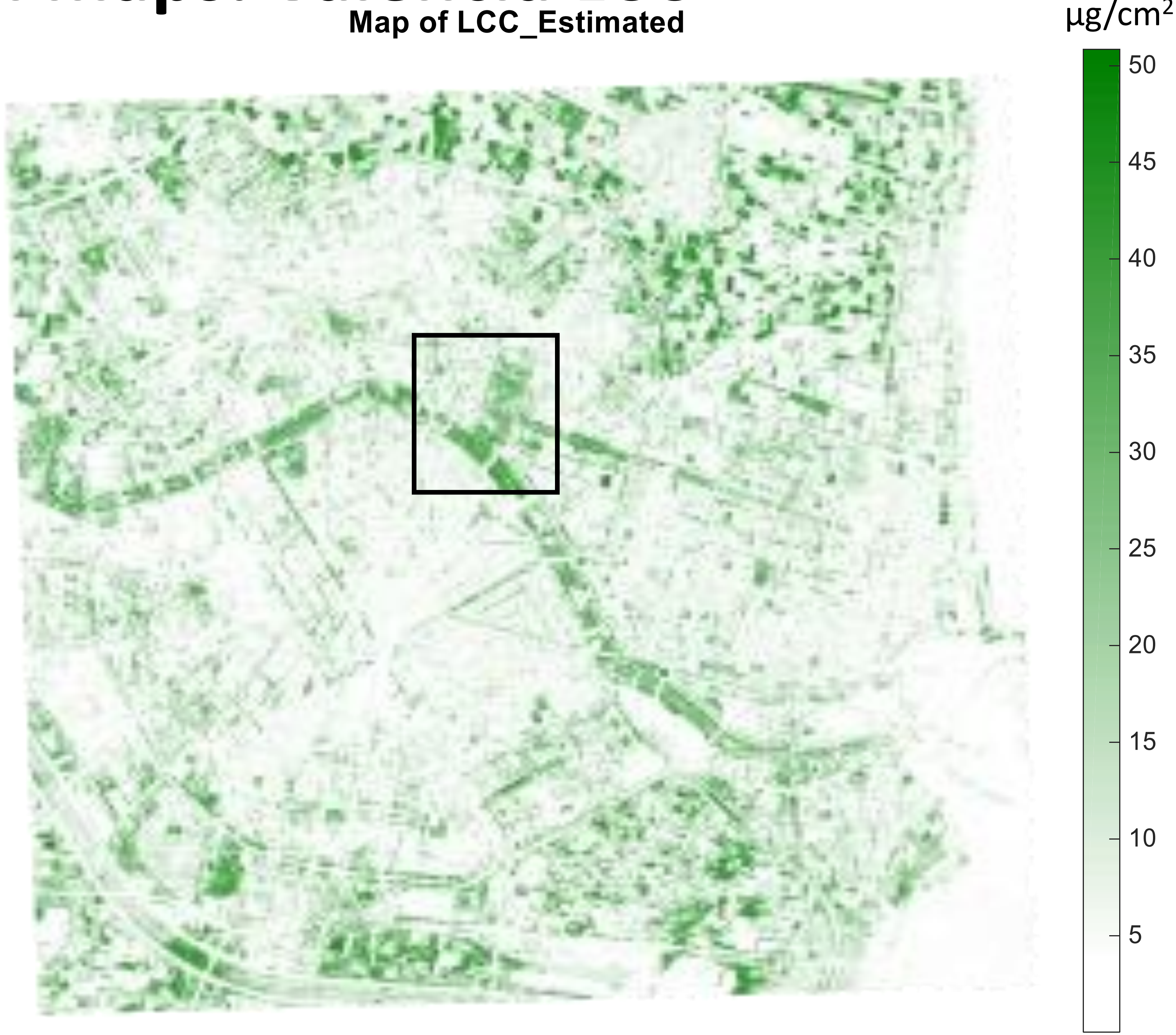
# FVC



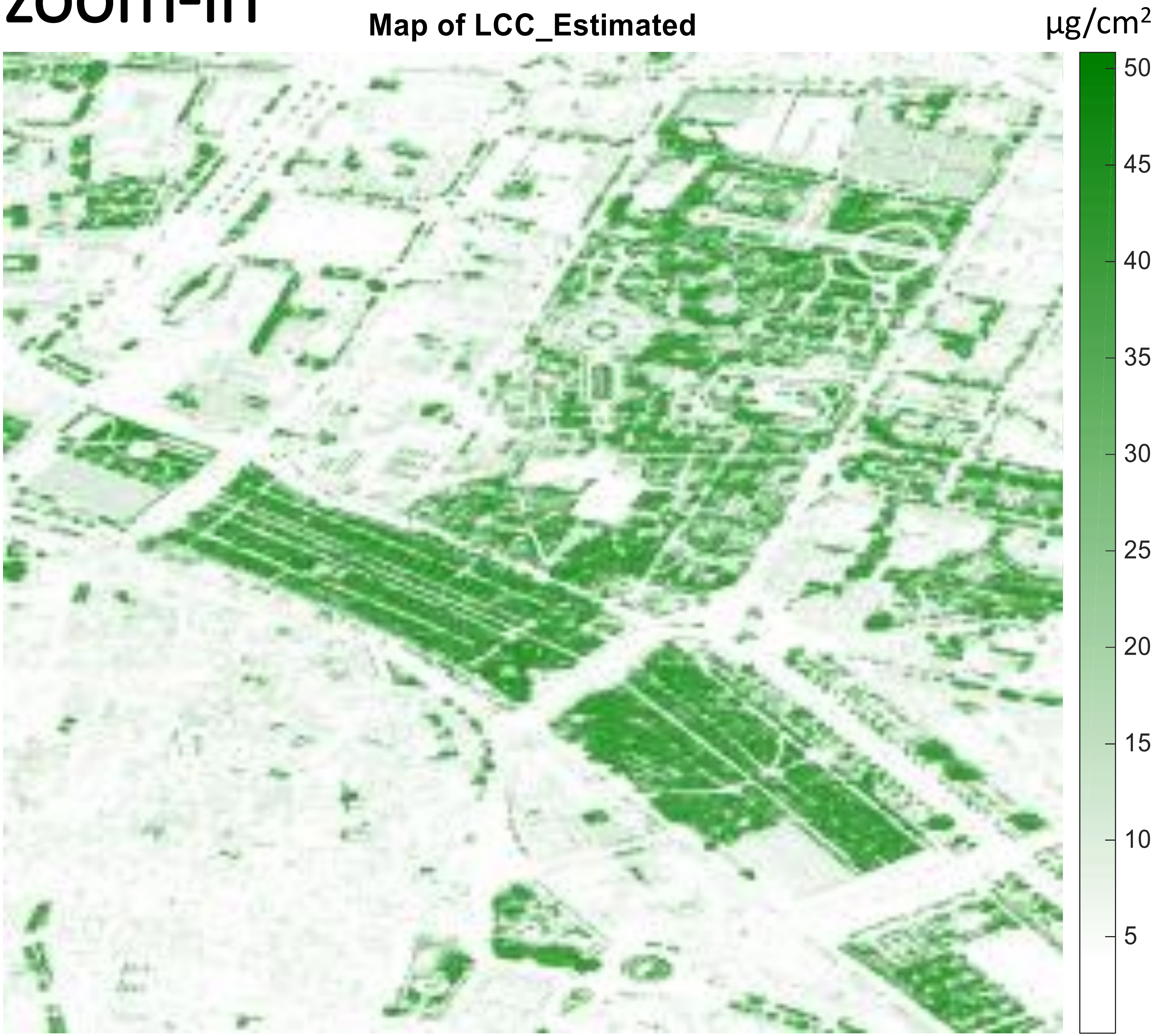
# FVC zoom-in



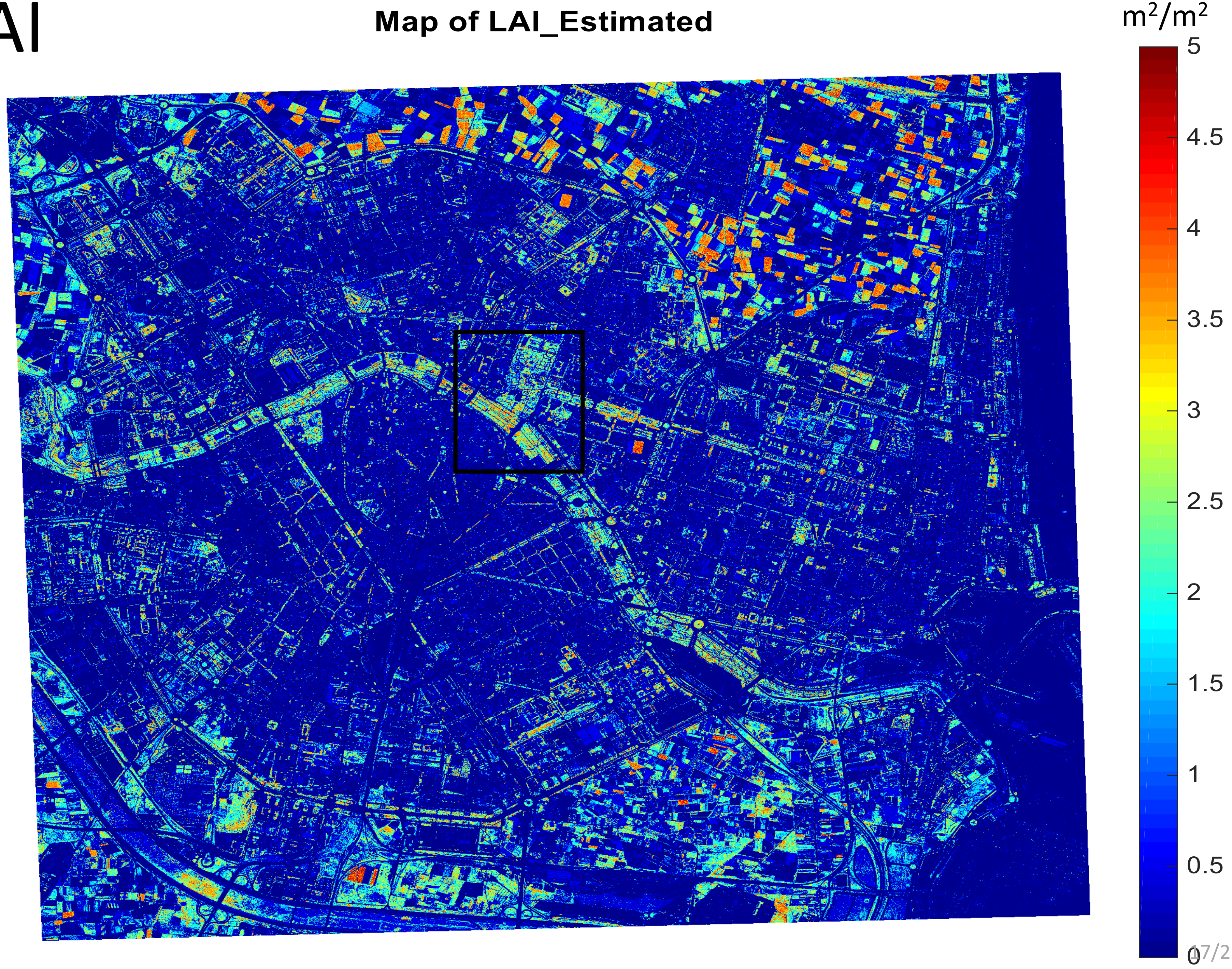
# Final maps: Valencia LCC



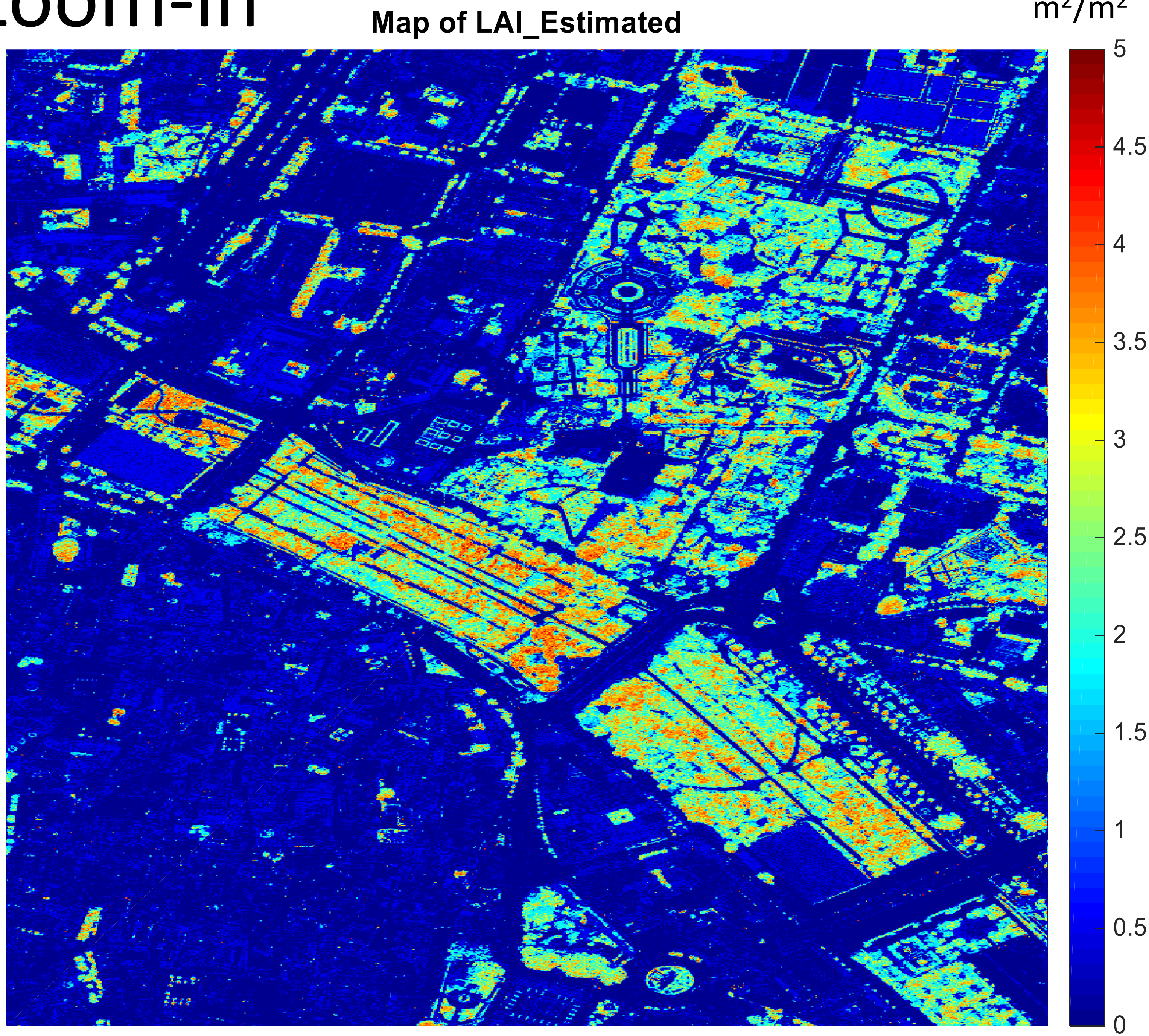
# LCC zoom-in



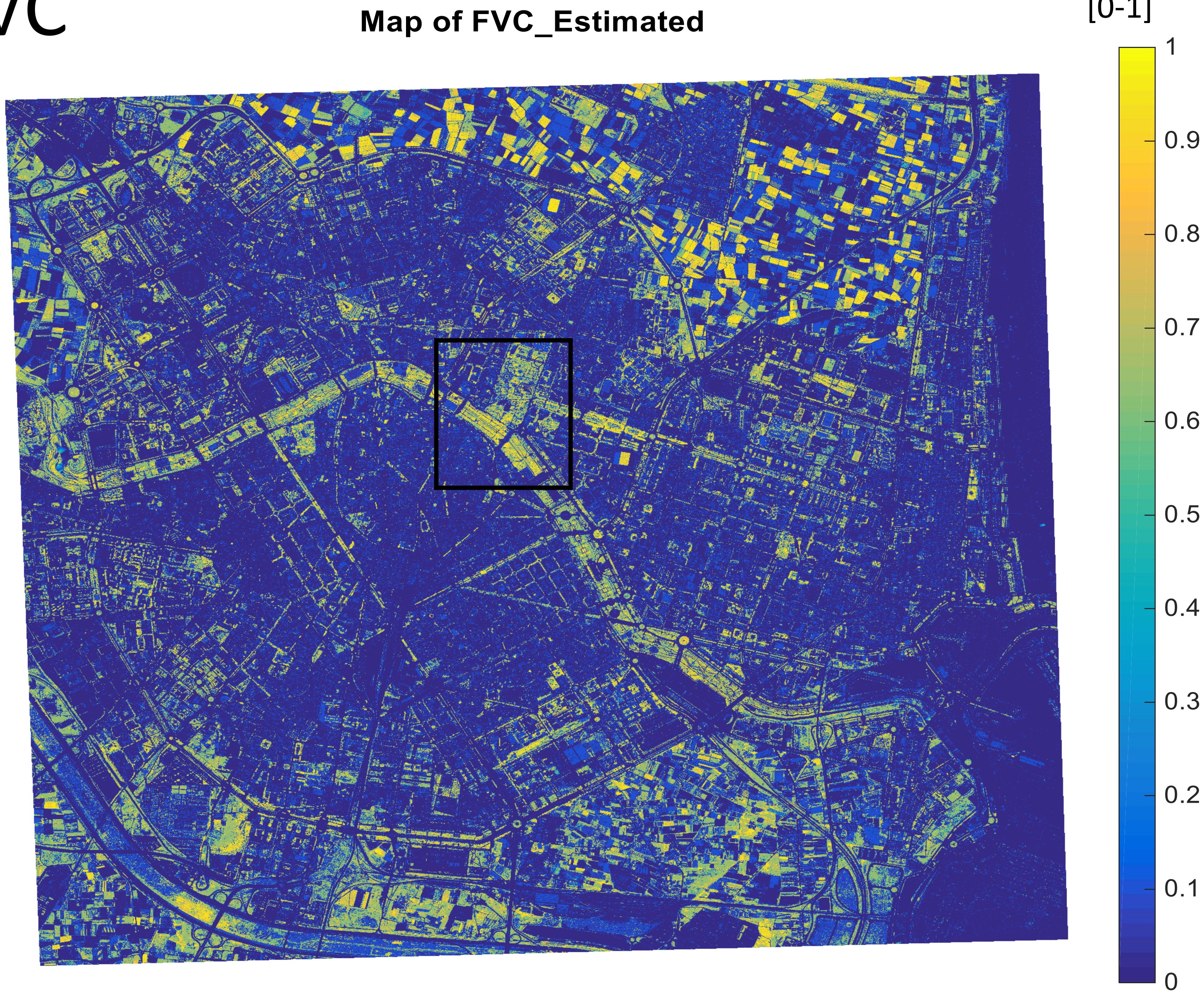
# LAI



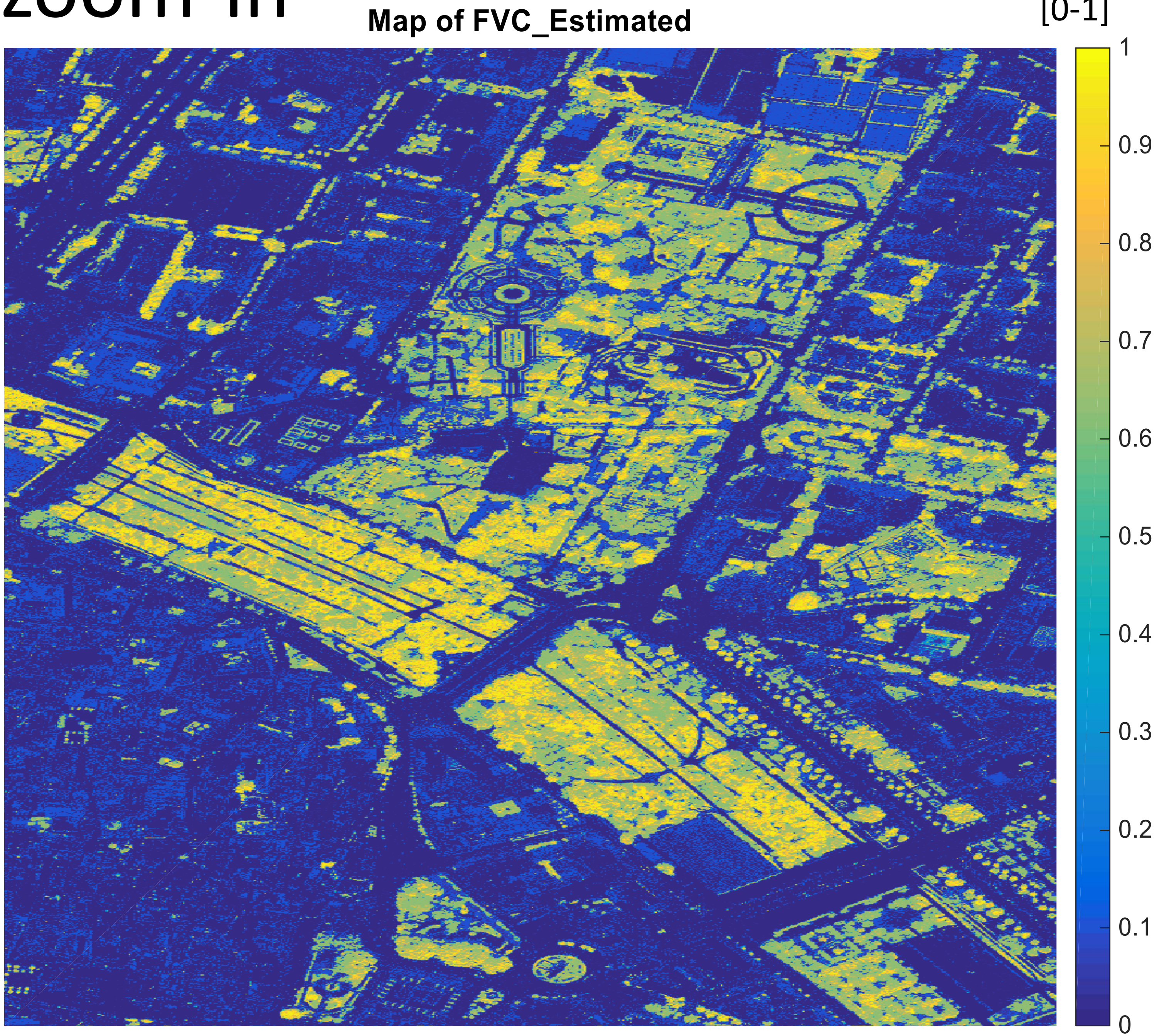
# LAI zoom-in



# FVC



# FVC zoom-in

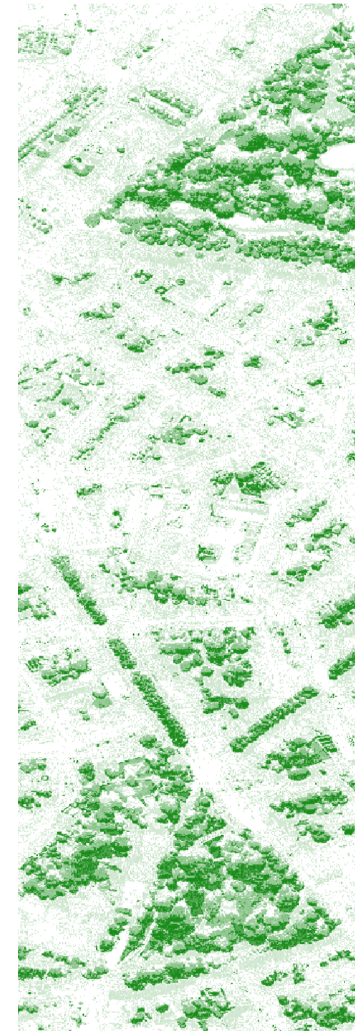


# Conclusions & recommendations

- NDVI maps successfully converted into maps of biophysical variables.

- Urban vegetation clearly identified. Within-crown variability of vegetation density quantified.

- Although maps are good, retrieval models can still be improved:
  - ✓Apply an atmospheric correction algorithm so that all spectral data can be used
  - ✓Add LAI, LCC, FVC measurements of urban vegetation to training data.




Thanks
?