Prototyping vegetation traits models in the context of the hyperspectral CHIME mission preparation

Jochem Verrelst, Charlotte De Grave, Eatidal Amin, Pablo Reyes, Miguel Morata, Enrique Portales, Santiago Belda, Giulia Tagliabue, Cinzia Panigada, Mirco Boschetti, Gabriele Candiani, Karl Segl, Stephane Guillasso, Katja Berger, Mattias Wocher, Tobias Hank, Uwe Rascher, Claudia Isola
CHIME: Copernicus Hyperspectral Imaging Mission

CHIME will carry a visible to shortwave infrared spectrometer to provide global routine hyperspectral observations. The mission will support new and enhanced services for sustainable agricultural and biodiversity management, as well as soil property characterisation.

Technical concept:
Routine spectroscopic observation in contiguous spectral bands:
- Instrument: Pushbroom Imaging Spectrometer 400 – 2500 nm, $\Delta \lambda \leq 10$ nm
- Revisit 10 – 15 days
- GSD (spatial resolution): 30 m
- Sun synchronous orbit (LTDN 10:30 – 11:30)
- Nadir view covering land and coastal areas
- High radiometric accuracy, low spectral/spatial misregistration

CORE Data Products:
The mission shall provide access to Level-1B, Level-1C and Level-2A products accessible via DIAS and with API support:
- Bottom-of-Atmosphere (BOA) reflectance (atmospherically corrected)
- Ortho rectified geometry
- Basic pixel classification (opaque clouds, thin clouds, cloud shadows, vegetation, water, snow etc.)
- Additionally -> Vegetation products (Level-2B)

http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Copernicus_High_Priority_Candidates
CHIME End-2-End Simulator (© GFZ)

- **OGM**: Simulate CHIME orbit and attitude
- **SGM**: Simulates the scene to be observed at fine spectral and spatial resolution.
- **OIP**: Models the CHIME sensor behavior.
- **L2GPP**: Performs the atmospheric correction
- **L2B**: Implements the algorithms for retrieval of bio-/geophysical variables, objective of the mission (e.g., retrieval of vegetation/soil/mineral contents).
- **PAM**: Evaluates the product performance at different processing levels
Hybrid workflow CHIME vegetation models

RTMs

Use RTMs (e.g. SCOPE, PROSPECT-PRO/DyN – SAIL) to generate a LUT composed by pairs (e.g. 1000) of vegetation parameters and spectra.

Active learning

Select the most representative samples from the training dataset via a diversity or entropy criteria. Later, add non-vegetated spectra.

PCA

Dimensionality reduction with PCA (20 components) or band selection.

Train GPR algorithms

With the LUT optimized for vegetation and non-vegetated surfaces, train probabilistic ML algorithms.

Maps + uncertainties

Final outputs of the workflow.

Apply to new observations

PRISMA images resampled to CHIME band settings

Validate the models

Assess models’ performance against field data and vegetation reference scenes.

Verification and Validation vegetation models through E2E using CHIME-like data
Exploitation of campaign data \textit{(in situ and images)} \textcopyright{} Italian team

Field measurements (Grosseto campaign, Italy) used for tuning the models using \textbf{Active Learning (AL)}:

\textbf{HyPlant-DUAL:}

<table>
<thead>
<tr>
<th>Date</th>
<th>07/07/2018</th>
<th>31/07/2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length</td>
<td>~7 km</td>
<td>~8 km</td>
</tr>
<tr>
<td>Surface Area</td>
<td>~18 km$^2$</td>
<td>~20 km$^2$</td>
</tr>
<tr>
<td>Runs Swath</td>
<td>380 m</td>
<td>1800 m</td>
</tr>
<tr>
<td>Pixel Size</td>
<td>1 m</td>
<td>4.5 m</td>
</tr>
<tr>
<td>Altitude (ASL)</td>
<td>600 m</td>
<td>3050 m</td>
</tr>
<tr>
<td>Flight conditions</td>
<td>Cloud Free</td>
<td>Cloud Free</td>
</tr>
</tbody>
</table>

Two fields selected $H_{FC}$ and $L_{FC}$ (High & Low fractional cover - different phenological stage)

3rd-7th July 2018 (66 plots)  
31st July 2018 (22 Plots)
Towards GPR v.1.8 models Validation of CHIME vegetation models with L2A (1/2)

- LUT simulated by SCOPE - 1000 samples spectrally resampled to CHIME with some bands left out.
- Noise added.
- Optimized with AL against **Grosseto** dataset.

<table>
<thead>
<tr>
<th>Variable</th>
<th>AL method</th>
<th>#samples</th>
<th>ML method</th>
<th>Spectral noise</th>
<th>R2</th>
<th>RMSE</th>
<th>RRMSE (%)</th>
<th>NMRSE (%)</th>
<th>Bands number</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>RSAL</td>
<td>386</td>
<td>VHGPR</td>
<td>0</td>
<td>0</td>
<td>21.23</td>
<td>47.43</td>
<td>43.15</td>
<td>247</td>
</tr>
<tr>
<td>LWC</td>
<td>RSAL</td>
<td>362</td>
<td>GPR</td>
<td>4.5</td>
<td>0.05</td>
<td>0.0056</td>
<td>37.83</td>
<td>23.06</td>
<td>207</td>
</tr>
<tr>
<td>LDMC</td>
<td>EBD</td>
<td>292</td>
<td>GPRm</td>
<td>0</td>
<td>0.13</td>
<td>0.013</td>
<td>27.20</td>
<td>25.99</td>
<td>207</td>
</tr>
<tr>
<td>LAI</td>
<td>EBD</td>
<td>404</td>
<td>VHGPR</td>
<td>3.5</td>
<td>0.76</td>
<td>0.8529</td>
<td>37.17</td>
<td>14.88</td>
<td>247</td>
</tr>
<tr>
<td>CCC</td>
<td>EBD</td>
<td>283</td>
<td>VHGPR</td>
<td>4.5</td>
<td>0.56</td>
<td>114.86</td>
<td>118.38</td>
<td>35.35</td>
<td>247</td>
</tr>
<tr>
<td>CWC</td>
<td>EBD</td>
<td>264</td>
<td>VHGPR</td>
<td>0</td>
<td>0.89</td>
<td>0.041</td>
<td>135.06</td>
<td>67.00</td>
<td>210</td>
</tr>
<tr>
<td>CDMC</td>
<td>-</td>
<td>1035</td>
<td>GPRm</td>
<td>0</td>
<td>0.46</td>
<td>0.013</td>
<td>87.55</td>
<td>44.95</td>
<td>247</td>
</tr>
<tr>
<td>CNC*</td>
<td>EBD</td>
<td>148</td>
<td>GPRm</td>
<td>0</td>
<td>0.65</td>
<td>3.6315</td>
<td>30.69</td>
<td>17.68</td>
<td>210</td>
</tr>
</tbody>
</table>

* LUT and validation data kindly shared by Ludwig-Maximilians-Universität (LMU) München
Towards GPR v.1.8 models Validation of CHIME vegetation models with L2A (2/2)

- For some variables no field data available.
- Theoretical validation of CHIME vegetation models with L2A
- LUT simulated by SCOPE and PROSAIL-PRO/DyN - 1000 samples resampled to CHIME with some bands left out.
- Noise added.

<table>
<thead>
<tr>
<th>Variable</th>
<th>AL method</th>
<th>#samples</th>
<th>ML method</th>
<th>Spectral noise</th>
<th>R2</th>
<th>RMSE</th>
<th>RRMSE (%)</th>
<th>NMRSE (%)</th>
<th>Bands Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCLC</td>
<td>--</td>
<td>1025</td>
<td>GPRm</td>
<td>0</td>
<td>0.3576</td>
<td>0.0009</td>
<td>44.81</td>
<td>15.88</td>
<td>251</td>
</tr>
<tr>
<td>LNC</td>
<td>--</td>
<td>1025</td>
<td>GPR</td>
<td>0</td>
<td>0.7537</td>
<td>0.0001</td>
<td>21.89</td>
<td>11.19</td>
<td>251</td>
</tr>
<tr>
<td>CCLC</td>
<td>--</td>
<td>1025</td>
<td>GPRm</td>
<td>0</td>
<td>0.5991</td>
<td>0.0040</td>
<td>51.72</td>
<td>9.88</td>
<td>251</td>
</tr>
<tr>
<td>FAPAR</td>
<td>--</td>
<td>1024</td>
<td>VHGPR</td>
<td>0</td>
<td>0.9752</td>
<td>0.0372</td>
<td>4.51</td>
<td>3.80</td>
<td>247</td>
</tr>
<tr>
<td>FVC</td>
<td>--</td>
<td>1024</td>
<td>VHGPR</td>
<td>0.5</td>
<td>0.9742</td>
<td>0.0425</td>
<td>5.25</td>
<td>4.25</td>
<td>247</td>
</tr>
</tbody>
</table>

To all training datasets, additionally ±30 non-vegetated spectra added.
L2BV v1.6 Processor

Radiative Transfer Model
SCOPE (v 1.70)

Machine Learning Algorithm
Gaussian Process Regression (GPR) + DR method (PCA, 20 components)

<table>
<thead>
<tr>
<th>version</th>
<th>Last release date</th>
<th>Revisions &amp; Changes</th>
</tr>
</thead>
</table>
| v1.6    | 29/03/2021        | - Error fixed in writing geoinformation (datatype) from the input image to the output product.  
- Error fixed when selecting GPR models for application.  
- Error fixed in activating/deactivating GPR models’ application.  
- Switch on/off products bands SD, CV and QF (Local.Config.File).  
- Implementation of a QF band per each variable (Local.Config.File).  
- A logfile (txt) added.  
- Variables written in the output product ordered from leaf to canopy level.  
- Negative values changed from NaN to 1e-5.  
- Code execution continues in case a GPR model does not work or an error appears. |
Main features processor v.1.6: 4 layers per variable

- (1) Trained to process full images (in principle no masking required)
- Two uncertainty layers: (2) absolute (SD), and (3) relative (CV in %)
- (4) Out-of-range quality flag

All these layers can be switched on/off in local xml file.
Verification V1.8: small simulated scene (outside E2E)

- GPR Models v1.8: After tuning against campaign data
- Verified against a 900 pixels reference image
- SCOPE & PROSPECT-PRO/DyN-SAIL simulated scenes

Pass: ≤10% NRSME

Canopy models are generally well trained.
Earlier attempts applied to airborne data resampled to CHIME (v1.7)

Julich
2018 FLEXSense
Earlier attempts applied to airborne data resampled to CHIME (v1.7)

HyPlant
2018 FLEXSense campaign
V1.8: PRISMA & data preparation (😊 Italian team)

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbit altitude reference</td>
<td>615 km</td>
</tr>
<tr>
<td>Field of View (FOV)</td>
<td>2.77°</td>
</tr>
<tr>
<td>Instantaneous FOV</td>
<td>48.34 mrad</td>
</tr>
<tr>
<td>Swath</td>
<td>30 km</td>
</tr>
<tr>
<td>Ground Sampling Distance</td>
<td>Hyperspectral camera: 30 m</td>
</tr>
<tr>
<td></td>
<td>Panchromatic camera: 5 m</td>
</tr>
<tr>
<td>Spectral range</td>
<td>VNIR: 400–1010 nm (66 bands)</td>
</tr>
<tr>
<td></td>
<td>SWIR: 920–2500 nm (173 bands)</td>
</tr>
<tr>
<td></td>
<td>PAN: 400–700 nm</td>
</tr>
<tr>
<td>Signal-to-noise ratio</td>
<td>VNIR: &gt;160 (450 at 650 nm)</td>
</tr>
<tr>
<td></td>
<td>SWIR: &gt;100 (&gt;360 at 1550 nm)</td>
</tr>
<tr>
<td></td>
<td>PAN: &gt;240</td>
</tr>
<tr>
<td>Spectral Width</td>
<td>≤14 nm</td>
</tr>
<tr>
<td>Spectral Calibration Accuracy</td>
<td>±0.1 nm</td>
</tr>
<tr>
<td>Radiometric quantisation</td>
<td>12 bits</td>
</tr>
<tr>
<td>Repeat cycle</td>
<td>29 days (450 orbits)</td>
</tr>
<tr>
<td>Relook time</td>
<td>&lt;7 days</td>
</tr>
<tr>
<td>Lifetime</td>
<td>5 years</td>
</tr>
</tbody>
</table>

Cleaning and smoothing of the PRISMA L2A images:

a) Raw PRISMA image

b) Smoothened PRISMA image

c) Reflectance vs. Wavelength (nm) graph
V1.8 Maps: results with 20 PCA

- **PRISMA** image resampled to CHIME band settings, heterogeneous spatial subset to test the performance in vegetation, buildings and water
- Zoom-in Munich-North-Isar campaign location from LMU, foreseen for EnMAP product validation
- Canopy variables 😊
- Some leaf variables 😞

GPR hybrid models powerful for vegetation trait mapping (with inclusion of uncertainty estimates)
Validation through L2BV Results: SCOPE scene

Pass: $\leq 10\%$NRSME

(Limited number of pixels due to masking)

- Variables that are doing great: LCC, CCC
- Very poor: LAI, LDMC
- E2E processing degrades retrieval performances.
- Probably artefacts in some spectral regions affect some variables (LAI, LDMC)
- Models need to be made robust for noise: *backup algorithms in preparation*
PRISMA vs CHIME E2E:

**CWC** [g/cm²]

**CNC** [g/m²]

**FVC** [-]

**Map of CWC**

**Map of CNC**

**Map of FVC**

**CHIME E2E**
Conclusions:

- L2BV v.16 processor with GPR models v1.8 delivered to ESA. All models provide uncertainty estimates.
- Majority of canopy models work fine on real data (e.g., APEX, PRISMA). Excellent performance for CNC, FVC, FAPAR.
- Leaf models are more challenging.
- First maps through E2E suggested artifacts but improvements in processing chain are expected.
- Models and software improvements are in the pipeline.
Thanks