UNDERSTANDING THE FUTURE MARKET FOR NOVASAR-S FLOOD MAPPING PRODUCTS USING DATA MINING AND SIMULATION

Samantha Lavender[(1)], Kajal Haria[(2)], Geraint Cooksley[(2)], Alex Farman[(2)] and Thomas Beaton[(2)]

[(1)]Pixalytics Ltd, 1 Davy Road, Plymouth Science Park, Derriford, Plymouth, Devon, PL6 8BX, UK. E-mail: slavender@pixalytics.com
[(2)]Telespazio VEGA UK Ltd., 350 Capability Green, Luton, Bedfordshire, LU1 3LU, UK. E-mail: Kajal.Haria@telespazio.com, Geraint.Cooksley@telespazio.com, Alex.Farman@telespazio.com and Thomas.Beaton@telespazio.com

ABSTRACT

The aim was to understand a future market for NovaSAR-S, with a particular focus on flood mapping, through developing a simple Synthetic Aperture Radar (SAR) simulator that can be used in advance of NovaSAR-S data becoming available.

The return signal was determined from a combination of a terrain or elevation model, Envisat S-Band Radar Altimeter (RA)-2, Landsat and CORINE land cover information; allowing for a simulation of a SAR image that's influenced by both the geometry and surface type. The test sites correspond to data from the 2014 AirSAR campaign, and validation is performed by using AirSAR together with Envisat Advanced (ASAR) and Advanced Land Observing Satellite "Daichi" (ALOS) Phased Array type L-Band Synthetic Aperture Radar (PALSAR) data.

It's envisaged that the resulting simulated data, and the simulator, will not only aid early understanding of NovaSAR-S, but will also aid the development of flood mapping applications.

1. INTRODUCTION

NovaSAR-S is a revolutionary concept of a low cost medium resolution SAR mission, being developed by Surrey Satellite Technology Ltd; in single polarisation mode the spatial resolution is 6 to 30 m, with swath widths ranging from 15 to 150 km, while alternative modes offer multi-polarisation at either a coarser spatial resolution or narrower swath width [1].

There is currently little literature on S-Band SAR imagery as the C Satellite of Environment and Disaster Monitoring and Forecasting Small Satellite Constellation (HJ-1-C) is the only active mission, launched 19 November 2012 [2], with a previous mission being the Russian ALMAZ-1 that operated in 1991-1992. In contrast, there have been a number of X-, C- and L-Band missions. X-band is the shortest-length, highest frequency, of these missions with a need for the smallest antenna and so a focus on high resolution surface mapping; 1 m resolution in both directions for missions such as TerraSAR-X and Cosmo-SkyMed. The medium-length (C-Band) microwaves have improved atmospheric transmission (i.e., penetration through clouds, dust, smoke, snow and rain), whilst the longer L-Band microwaves can also penetrate canopy cover and so provide information about the underlying land surface. S-Band, lying between the C- and L-Bands, offers improved atmospheric transmission and vegetation penetration over C-Band, whilst requiring a smaller antenna and so having an improved spatial resolution compared to L-Band missions. Therefore a simulator shall not only aid an early understanding of NovaSAR-S data, but also provide foresight into some of the future applications S-Band data may have for water assessment and analysis during a flood event.

The NovaSAR-S project, with a particular focus on flood mapping, was funded through a call for pathfinder projects from the UK's Centre for Earth Observation Instrumentation and Space Technology (CEOI-ST). The principal objectives were to develop a simple simulator that could produce S-Band SAR datasets from a terrain model, supported by data mining of Envisat S-Band RA-2 data alongside land surface characteristics.

2. METHODOLOGY

The initial research focused on the analysis of S-Band data from RA-2, to see whether it could provide insights into what a SAR system might be sensitive to.

2.1. Analysis of Radar Altimeter Data

Envisat's RA-2, functioned mainly at a nominal Ku-Band frequency, but also had an S-Band frequency; this dual frequency operation was initially designed to aid in the correction of ionospheric delay. A report [3] looking into exploiting individual RA-2 echoes and S-Band data for ocean, coastal zone, land and ice/sea ice altimetry (project name RAIES) in 2007 found that the S-Band measurements over the ocean and in coastal zones appeared much noisier than Ku-Band and European Remote Sensing (ERS) measurements, when validated against Significant Wave Height buoy data. However, S-Band sigma nought values were also found to be less affected by rain, more sensitive than Ku to ionospheric perturbation and less sensitive to the effects of

atmospheric liquid water. It was found that S-Band also allowed observations of greater surface slopes (up to 3 as opposed to Ku-Band's limit of 0.5) and it yielded a better penetration depth. Therefore, the use of the altimeter signal for understanding the interaction of S-Band with land surfaces was considered promising.

Analysis was performed for the Crop and Peak District test sites corresponding to AirSAR flightlines (see Tab. 1), which correspond to S- and X-Band data collected during airborne overflights during the 2014 AirSAR campaign (a UK partnership between Airbus Defence and Space, the Natural Environment Research Council and Satellite Applications Catapult). This involved extracting altimeter echo waveforms, from the RA-2 Level 2 Sensor Geophysical Data Record (SGDR) products, and then comparing them to Normalized Difference Vegetation Index (NDVI) values, which depend on a high reflectance in the Near InfraRed by plant matter contrasting to the strong absorption by Chlorophyll-a in the red wavelengths that's termed the 'red edge' [4], calculated from Landsat imagery collected as concurrent as possible.

Table 1. Crop and Peak District test site locations.

Day / Run	Location	Near edge pixel	Far edge pixel
26/06/14 Crop 2	Worcester shire	1°52.233'W 52°9.721'N	1°47.017'W 52°10.803'N
26/06/14 Crop 3	Worcester shire	1°42.010'W 52°11.492'N	1°36.843'W 52°12.065'N
26/06/14 Peak	Peak District	2°1.282'W 53°38.067'N	1°49.154'W 53°21.890'N

Fig. 1 shows an example for the Crop test site with the left plot showing the echo groups (numbered 1995, 1996 and 1997) as small images; x-axis is echo waveform and y-axis is the set of echoes, with a rainbow colour palette indicating the echo magnitude. On the right a Landsat waveband is overlaid by the interpolated positions of the individual altimetry echoes, which is where the NDVI values were extracted from.

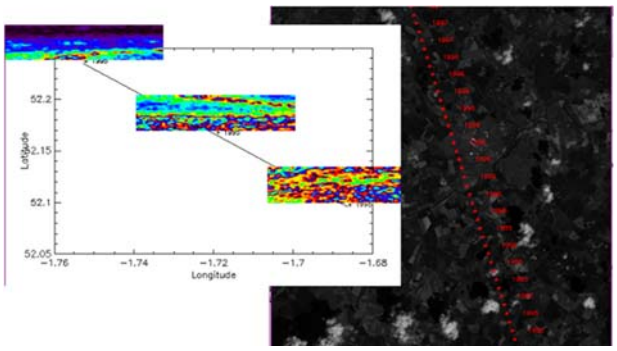

Figure 1. Extraction of the RA-2 echo waveforms (left, shown by plotting the bursts on a Latitude/Longitude plot) overlaying the altimetry track on the Landsat image (right).

Fig. 2 shows the RA calculated magnitude (height of the echo waveform peak) and peak position (gate number) plotted against the Landsat NDVI values, with the closest comparisons in time plotted in red. Although the results don't appear to be random, there is not a simple relationship between NDVI and the two echo waveform metrics chosen. It was therefore not clear how these results could be used to build a look-up-table to provide radar return strength for different vegetated surfaces. Therefore, as the project was only a short term (6 month study) the focus shifted to still using NDVI, but with the S-Band radar backscatter values coming from peer-reviewed papers.

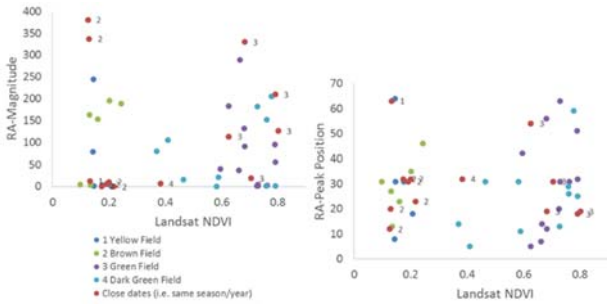

Figure 2. Comparison of the RA-2 echo waveform metrics to the corresponding Landsat NDVI values.

2.2. Stage 1 Simulator

The simulator was initially developed using the approach adopted in the Next ESA SAR Toolbox / Sentinel-1 Toolbox (S1TBX) to encompass terrain variations [5]. As an outline, Envisat Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) data is used to determine the satellite position with ASAR metadata providing the SAR geometry, and then the return signal is determined from a combination of a terrain or elevation model. Together these inputs allow for a simulation of a SAR image that is influenced by both the geometry, with the results being output in azimuth and slant range co-ordinates (slant range image).

Stage 1 initially used a combination of the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), but there were gaps in the resulting simulation that were thought to be caused by the resolution of the input DEM data. Therefore, in a second version, the DEM data was swapped to a higher spatial resolution; Ordnance Survey (OS) DEM at 5 m spatial resolution.

In Fig. 3, the OS DEM (Top) has a distinctive feature than can be seen within both the initial and adjusted simulation results. The simulation result looks reasonable (Bottom), but only the bottom left quarter of the DEM was in the output; the reasons for this are not currently fully understood as the mathematics is complex.

As implementing the DEM derived simulation was

taking time and there was difficultly with using the adjusted simulator outputs in the validation activity, because another step needs to be applied (resampling back to a map projected grid) in order to convert the data to a format that can be inserted into a GIS etc., it was decided to put this step on hold and work on the next step.

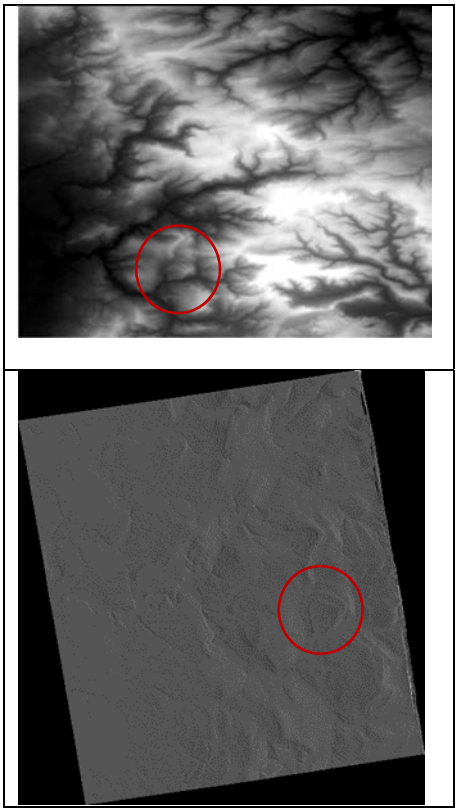

Figure 3. Peak District test site produced from the (Top) OS DEM and (Bottom) simulation of the bottom left quadrant from the Pixalytics code.

2.3. Stage 2 Simulator

TPZ VEGA provided an ENVI SARScape DEM simulation output and then Pixalytics concentrated on adding a surface roughness element. In the first test the SARScape output was sigma nought, but this limited the usefulness and so the Local Incidence Angle (LIA) was exported instead and the processing shown in Fig. 4 applied.

A Landsat image was processed to create maps of the NDVI and Normalized Difference Water Index (NDWI) [6], where the first indicates the fraction of green vegetation and the second characterises the amount of water held in the vegetation. The date of the Landsat acquisition was chosen so that it was as close in time as possible to the target simulation date, but also cloud free over the test site.

The classification of water was set to where the NDVI is less than or equal to 0.075, and then urban was values of between 0.075 and 0.3 with vegetation being NDVI values greater than 0.3. As the input Landsat image has been atmospherically corrected these values should be relatively stable, but this needs to be tested for a greater range of inputs.

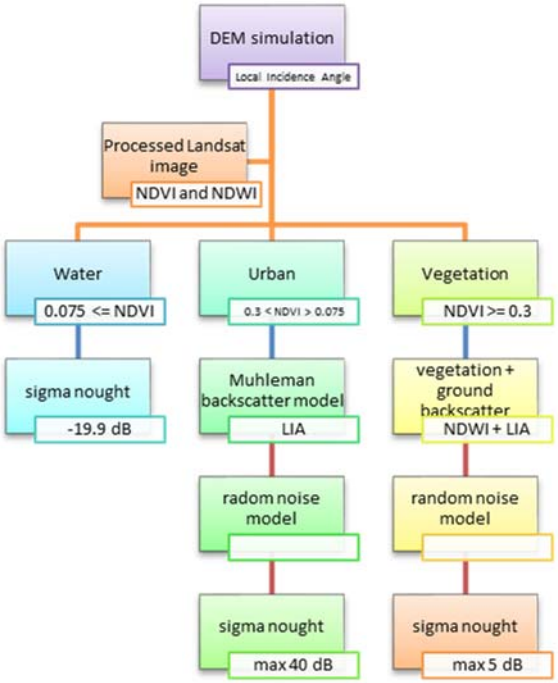

Figure 4. Processing flow diagram for assigning surface texture information.

If the target is water then a sigma nought value of 0.0102329 (-19.9 dB) is assigned. If the target is classed as urban then a version of the Muhleman backscatter model [5] (as used with NEST/S1TBX) is applied i.e. the surface is assumed to be rock like with backscattering being strongly dependent on the LIA. The code then applies random noise, which is scaled according to the estimated backscatter, and a maximum value of 40 dB is allowed.

If the target is classed as vegetation the NDWI is used to estimate the vegetation water content, and then the backscatter that would result from this vegetation; this is currently based on a paper [7] that developed a semi-empirical model for soya bean canopies at L and C bands. Also, the LIA is used to estimate the surface backscatter using a version of the Muhleman backscatter model and then the values are combined according the amount of vegetation cover; also estimated from the NDVI. After this, the random noise is applied and a maximum value of 5 dB is allowed.

3. RESULTS

3.1. Stage 2 Simulation Outputs

Fig. 5 shows the inputs involved for the Peak District test site, and final result. The NDVI map (Top Left) shows that lower values are seen over the higher ground where the underlying terrain is more exposed, but with the lowest values occurring for water and urban areas. The surface type classification shows that the separation is working well, but it is difficult to separate water and urban in the current approach. The LIA (Bottom Left) shows the strong influence of the terrain variation while the final simulation has this reduced over lower / flatter ground where the influence of the surface type is causing variations in the backscatter. The current formulation has been setup so that the output is an HH polarized S-Band image, but further work is needed on the parameterisations involved.

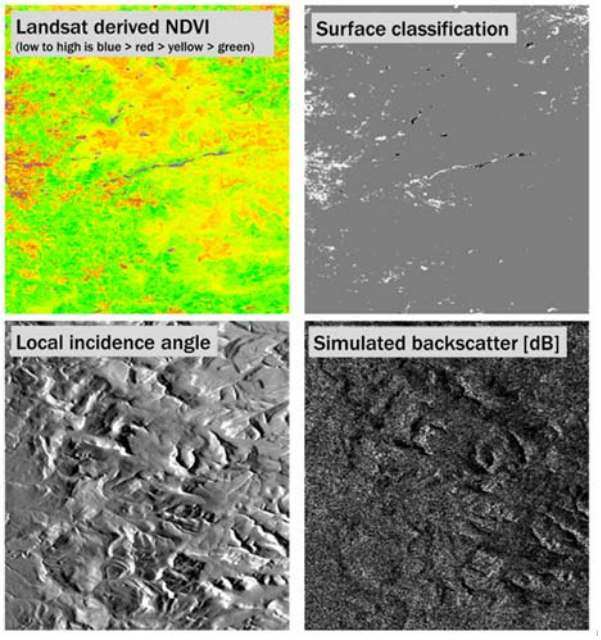

Figure 5. (Top Left) NDVI calculated from Landsat (low to high is blue > red > yellow > green), (Top Right) Surface type classification (black for water, grey for vegetation and white for urban), (Bottom Left) Local Incidence Angle and (Bottom Right) Simulated backscatter (sigma nought in dB to show range of values).

3.2. Validation Approach

Validation of the simulated S-Band imagery was conducted using the PALSAR (L), Envisat (C) and mean L and C band composite datasets, see Tab. 2, and by comparing the land cover classes.

The mean composite image of L and C band SAR data was derived from PALSAR and Envisat imagery in an attempt to simulate similar wavelength (λ) / dB values to S-Band SAR imagery i.e. ~10cm λ; taken after consideration of the scattering properties [8, 9 and 10]. The penetration capabilities of the datasets vary: L-Band being able to significantly penetrate forest canopy while C-Band only partially penetrating the forest canopy.

Table 2. Validation datasets.

Peak

Satellite/Sensor	Acquisition Date	Acquisition Time	SAR band
ALOS/PALSAR	16/06/2010	22:22	HH L-Band (23cm)
Envisat/ASAR	17/06/2010	21:39	HH C-Band (6cm)
Mean L and C Band	N/A	N/A	HH (L+C)/2 (~14cm)
AirSAR	26/06/2014	N/A	HH S-Band (~9cm)

Crop

Satellite/Sensor	Acquisition Date	Acquisition Time	SAR band
ALOS/PALSAR	15/08/2009	22:25	HH L-Band (23cm)
Envisat/ASAR	06/08/2009	21:39	HH C-Band (6cm)
Mean L and C Band	N/A	N/A	HH (L+C)/2 (~14cm)
AirSAR	26/06/2014	N/A	HH S-Band (9cm)

Localised comparisons between the S-Band simulated datasets and the SAR datasets were conducted on a set of sub-areas defined to represent a variety of land cover types as determined from the CORINE land cover classification [11]:

1 - Continuous Urban Fabric
2 - Discontinuous Urban Fabric
3 - Industrial and Commercial
6 - Airports
11 - Sport and Leisure
12 - Non-Irrigated Arable Land
18 - Pastures
23 - Broad leaf Forest
24 - Coniferous Forest
26 - Natural Grassland
27 - Moors and Heathland
29 - Sclerophyllous Vegetation
36 - Peat Bog
41 - Water Bodies

Non-urban land cover classes were analysed using 20-30 Vectorised Regions of Interest (ROIs) to obtain precise averages of dB values, whereas urban land cover classes were analysed with >50 ROIs due to the inconsistent

nature of urban environments and as a consequence erratic dB values. The mean pixel values from each ROI, in a land cover class, were collated to obtain a representative mean pixel value of the individual land cover classes across the two test sites (Peak and Crop).

RGB basemap imagery (derived from the satellite data within Google Earth™ displayed using ArcGIS) was included within the qualitative validation phase, see Fig. 6, allowing for easier visual interpretations between SAR images. In addition, Fig. 6 shows the Simulated S-Band dataset alongside the RGB imagery and Mean L and C band validation datasets over the second test site 'Crop'. A greater diversity of visible land cover classes is visible within the 'Mean L and C Band' than within the simulated S-Band image. However, within the simulated S-Band image there are visible differences in the backscatter strength between urban and other land cover classes.

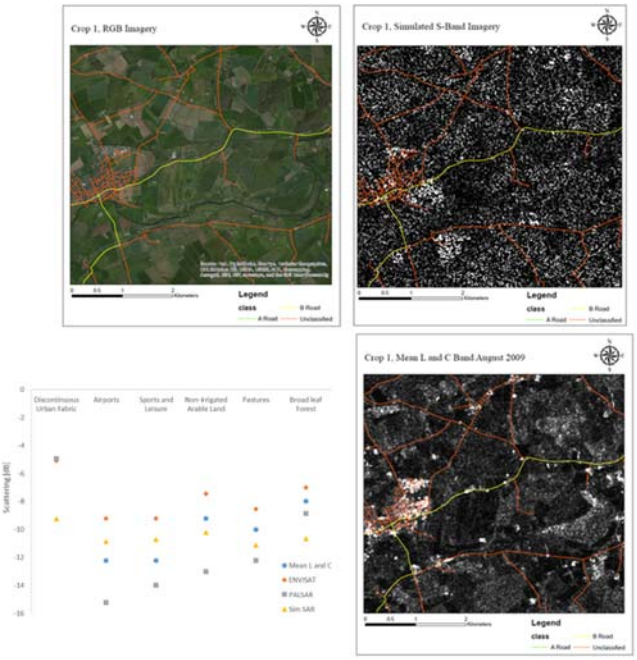

Figure 6. RGB basemap imagery (Top Left) alongside the Simulated S-Band imagery (Top Right) and Mean L and C Band imagery (Bottom Right) plus a plot of the scattering values for the different CORINE land cover classes.

Between validation datasets it is observed that there are strong similarities for the simulated backscatter values within classes of intensively managed grassland e.g. 'Sports and Leisure', 'Airports' and 'Pastures' and other vegetation classes such as 'Non-irrigated Arable Land' and 'Broad leaf forest'. This may be due to the simulation methodology being based on research on leaf dominant vegetation: soya bean.

A qualitative validation using S-Band AirSAR imagery was also conducted for the two test sites, but quantitative validation could not be achieved due to the loss of calibrated dB values in the processed AirSAR data.

4. CONCLUSION

Validation of the simulated S-Band SAR data with historical/current SAR data over the same locations was undertaken to ensure that:

- Geolocation is nominal
- Features in the simulated SAR product appear as expected

The project met these requirements i.e. the generation of a simple SAR simulated dataset with accurate geocoding when compared to an independent SAR validation dataset. In addition, the quantitative comparison of the simulated S-Band against the C- and L-Band SAR and mean of C- and L-Band SAR against a number of land cover types as defined by CORINE data showed that the simulated S-Band is different to both but more similar to L than C-Band.

The qualitative comparison made between simulated S-Band and AirSAR S-Band aerial data showed that the simulated S-Band did replicate the backscatter appearance of AirSAR (although at lower definition since the AirSAR source data is of much higher resolution). The performance varied as might be expected given that the backscatter variance was based on the level of NDVI/NDWI indices and not taking into account structural information of the canopy type (cf. inability to distinguish forest from unmanaged vegetation). Nevertheless it is felt the simulator as it stands is useful as learning tool and its further use with comparison to further datasets would allow its further refinement. As was reported, the simulator performed well on unmanaged land where its core of being built on NDVI lends itself well to leafy vegetation. It also simulates urban against natural again reflecting the decision tree division in the simulation model.

4.1. Future Steps and Roadmap

There are a number of areas for further development, which include:

- Envisat RA-2 data: Some interesting results were achieved, and it would be useful to return to this component of the study and extract further signals based on the ROIs used for validation. From discussions at the ESA Living Planet Symposium 2016 (Blarel *pers. comm.* and poster), backscatter maps derived from the RA-2 S-band data are strongly correlated with soil moisture in addition to surface roughness and land cover.

- DEM simulation: Return to this code and implement the additional step so that the output can be

converted to geographical co-ordinates, and then the roughness applied.

- Use of Landsat within the simulation: consider including Sentinel-2 so that a higher spatial resolution classification can be produced.

- Separation of water and urban land cover types: Include an additional step to the classification e.g. a requirement for a low value in the Mid InfraRed to make sure water and urban land cover types are correctly separated.

- Vegetation surface type: The current approach is based on a single paper, although significant background reading has been undertaken. This is because there is limited research published where both L- and C- band SAR data have been considered, and hence we can estimate the S-Band response; there are even fewer papers that discuss S-Band itself. By further analysing existing L- and C- band scenes taken in close time proximity to each other, further knowledge can be gained and so built into the vegetation modelling.

- Urban surface type: A better approach is needed for more accurate determination of the backscatter in urban areas (e.g., ray tracing within a 3D model) as the current approach will be underestimating the return signal because a significant proportion of the backscatter will be coming from double bounces. This could be implemented using the Environment Agency's Lidar data that is available as a Digital Surface Model (DSM), rather than elevation model (DEM), except that the data will not exist for higher ground as it was collected for flood mapping purposes.

5. REFERENCES

1. Davies, P., Whittaker, P., Bird, R., Gomes, L., Stern, B., Sweeting, M., Cohen, M & Hall, D. (2012). NovaSAR – Bringing Radar Capability to the Disaster Monitoring Constellation. *Proceedings of the 26th Annual AIAA/USU Conference on Small Satellites*, Logan, Utah, USA, August 13-16, 2012. Online at http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1020&context=smallsat (as of 11 May 2016).

2. Liang, L., Wei, Z., Xiaoguang, J., Xianbin, L., Hongyuan, H. & Xuehua, Z. (2016). Water Body Extraction Research Based on S Band SAR Satellite of HJ-1-C. *International Journal of Advanced Remote Sensing and GIS*, **5**(2), 1514-1523.

3. Challenor, P.G. et al. (2004). Exploitation of the Envisat radar altimeter individual echoes and S-Band data for ocean, coastal zone, land and ice/sea ice altimetry (RAIES), *Task 6: Scientific applications of S-Band data*. Southampton, UK, Southampton Oceanography Centre, pp77.

4. Curran, P.J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, **29**, 271-278.

5. Liu, H., Zhao, Z. & Jezek, K.C. (2004). Correction of Positional Errors and Geometric Distortions in Topographic Maps and DEMs Using a Rigorous SAR Simulation Technique. *Photogrammetric Engineering & Remote Sensing*, **70**(9), 1031–1042.

6. Gao, B-C. (1996). NDWI - A Normalized Difference Water Index for remote sensing of vegetation liquid water from space. *Remote Sens Environ*, **58**, 257-266.

7. De Roo, R.D., Du, Y., Ulaby, F.T. & Dobson, M.C. (2001). A semi-empirical backscattering model at L-Band and C-Band for a soybean canopy with soil moisture inversion. *Geoscience and Remote Sensing. IEEE Transactions*, **39**(4), 864-872.

8. Chu, H. & Ge, L. (2010). Land Cover Classification Using Combinations of L and C Band SAR and Optical Satellite Images. *Proceedings of Asian Association on Remote Sensing (ACRS)*, Hanoi. pp1-5.

9. Haarpainter, J., Davids, C., Hindberg, H., Zahabu, E. & Malimbwi, R.E. (2015). Forest and Forest change Mapping with C and L Band SAR in Liwale, Tanzania. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **40**(7), 391-xxx.

10. Li, G., Lu, D., Moran, E., Dutra, L. & Batistella, M. (2012). A comparative analysis of ALOS PALSAR L-Band and RADARSAT-2 C-Band for land cover classification in a tropical moist region. *ISPRS Journal of Photogrammetry and Remote Sensing*, **70**, 26-38.

11. Copernicus. (2016). CORINE land cover. Online at http://land.copernicus.eu/pan-european/corine-land-cover (as of 3 May 2016).

6. ACKNOWLEDGEMENTS

The Envisat ASAR, DORIS and RA-2 data is courtesy of ESA, with the Landsat data courtesy of the USGS/NASA and the ALOS PALSAR data from JAXA. The AirSAR campaign data was provide by the Satellite Applications Catapult and the CORINE land cover data came from Copernicus Land Service.

In addition, the NEST and S1TBX packages have been developed under funding from ESA and the OS DEM data is provided through an OS license: © *Crown copyright and database rights (2015) OS (100057045)*.

The project activities were funded by the CEOI-ST (where Telespazio VEGA UK were the prime), with funding coming from the UK Space Agency.