Defining Brightness-Shape-Moisture soil parameters for Southern Africa from Hyperion hyperspectral imagery

Wessel Bonnet, Moses Azong Cho, Cecilia Masemola

***Abstract*— An effective methodology is needed to simulate soil spectra on a large scale. The brightness-shape-moisture (BSM) radiative transfer model is used to simulate soil spectra for different semi-arid and arid biomes within Southern Africa based on hyperspectral imagery obtained from the Hyperion satellite. Such simulation based on hyperspectral data is especially relevant in light of newer hyperspectral missions such as Prisma providing ongoing data streams. In this particular study, Hyperion's data is cleaned using the SUREHYP procedure, segmented using the SLIC algorithm, filtered to exclude photosynthetic and senescent vegetation, and parameterised via a Hyperion band calibrated BSM model look-up table to obtain simulation parameter distributions for different biomes. This provides a means to better simulate soil spectra using each biome's obtained parameter distributions in the BSM forward model.**

***Index Terms*— Africa; Hyperion; Radiative Transfer; Soil properties**

I. Introduction

Maps of soil types and their biophysical and biochemical properties are essential tools in agriculture, land use planning, environmental conservation, water management and infrastructure planning. In general, accurate maps of soil properties over large areas can be generated through geostatistical analysis of data from numerous representative sample points. However, obtaining soil data from many field sample points is not viable because manual soil collection and laboratory analysis is a time-consuming, destructive, laborious and expensive exercise. Remote sensing (RS) offers a more efficient, cost-effective alternative.

Remote sensing of soil properties generally follows two methods, namely, empirical and physically-based modelling approaches. In the Southern African semi-arid context, this is especially important, since soil erosion, highly variable rainfall patterns and large diurnal temperature fluctuations cause variability in soil moisture contents and structural properties, and regular measurement is needed to gauge changes. Despite this, no consolidated soil spectrum mapping has been performed for the particular study area.

W. Bonnet is with the Precision Agriculture Research Group, Advanced Agriculture and Food, CSIR, Tshwane, and also with the Department of Plant and Soil Sciences, University of Pretoria (email: wbonnet@csir.co.za)

M.A. Cho is with the Precision Agriculture Research Group, Advanced Agriculture and Food, CSIR, Tshwane, and also with the Department of Plant and Soil Sciences, University of Pretoria, Tshwane

C. Masemola is with the Precision Agriculture Research Group, Advanced Agriculture and Food, CSIR, Tshwane, and also with the Department of Geography, University of KwaZulu Natal, Pietermaritzburg.

In literature, soil spectrum analysis for the purposes of soil moisture level estimation has largely been done through empirical methods [1-5], primarily correlating visible and near -infrared soil spectra and their first-order differences or trend-removed values against soil moisture levels. Radiative transfer methods have been less common in this field, although these vector-based models are typically more flexible than traditional spectrum simulations [6]. By contrast, radiative transfer model (RTM) methodologies have been used extensively to model leaf and canopy characteristics and have proven to show more robustness across sites and seasons compared to empirical methods [7-9]. RTM methods are often also able to directly estimate soil moisture content without additional correlation modelling, which empirical methods cannot do [10,11].

Soil RTMs can broadly be categorized into three categories, namely beam tracing RTMs, multi-flux radiative transfer (MFRT) models, and bidirectional reflection distribution function (BRDF) RTMs. Beam tracing models analyse surface refraction and reflectances of sun-induced radiation between air and soil media and soil spectral absorptions using Fresnel equations or Snell's law. The multilayer radiative transfer model of soil reflectance (MARMIT) model was proposed by Bablet et al. [12] is based on beam tracing, and models wet soil as dry soil surrounded by liquid water film, with light refracted and scattered between the soil and the water film. The MARMIT-2 model [13] improves on this by considering the state of mixing of soil particles within the water film and calculating a mixed soil-water reflectivity, more closely aligned to the true beam soil-water interactions.

MFRT models ignore the multiple radiance interactions between the media, only considering the bidirectional incident fluxes before and after absorption, yielding a more tractable and user-friendly RTM. Sadeghi et al. [14] laid out a model for the estimation of soil moisture based on the Kubelka-Munk model [15,16], where surface radiance scattering and absorption characteristics are deconstructed into the individual scattering and absorption properties of dry soil with surrounding moisture and air. Ou et al. [17] put forward an equation for soil thicknesses using the Kubelka-Munk model, and estimated soil organic matter content based on scattering coefficients for soil spectra of different thicknesses. These models only work at specific wavelengths, decreasing their usefulness in hyperspectral imagery analysis, and making them intractable for the modelling of continuous or high spectral resolution radiance responses.

BRDF models consider incident and reflected light distributions within a hemispherical space. The most widely used of these is the Hapke model [18-20], which separates total radiance into single-scattering and multiple-scattering radiance components. Jacquemoud et al. [21] proposed the SOILSPECT radiative transfer model for soils, based on the Hapke model using a Legendre polynomial approximation to model bidirectional scattering from the soil. Both the beam tracing and the BRDF RTMs can gauge soil moisture content if dry soil spectra are known, but are unable to model dry soil.

In this study, soil spectra were modelled using the Brightness-Shape-Moisture (BSM) radiative transfer model as first used by Verhoef et al. [22] and described by Yang et al. [23,24], which can model full spectral signatures between 400 and 2500 nm as a beam tracing method, while also having the capability of modelling dry soil spectra before gauging soil moisture content. In previous studies utilising the BSM model, soil spectra are not explicitly modelled, but rather approximated as a combination of specified dry and wet soil spectra [25] or modelled from existing dry soil spectra [26,27].

In terms of RTM inversion for soil spectra retrieval, Eon et al. [28] performed soil RTM inversion using the MARMIT model to estimate soil moisture in bare dunes in the USA from UAS imagery, and Bayat et al [29] focused on retrieving land surface properties during a drought episode from Landsat imagery using BSM within a multi-layer RTM, while Prikaziuk et al. [30] inverted ASD-retrieved bare soil spectra to obtain BSM parameters for SCOPE-modelled potato fields.

However, based on research and literature sources consulted, no large-scale soil spectra inversion endeavour has been undertaken using the BSM methodology applied directly to hyperspectral spaceborne satellite imagery. Furthermore, it is posited that localised BSM soil parameter distributions will perform better than worldwide soil distributions in inverting or simulating Southern African semi-arid biome soil spectra. In this study, due to the rarity of hyperspectral spaceborne data, imagery obtained from the Hyperion imaging instrument is used to find BSM parameter distributions specific to the study area. Thus, the objective of the study is to explore whether soil spectra from semi-arid regions in Southern Africa can be modelled more accurately from localised BSM parameters.

II. Pre-processing

The Hyperion satellite project commissioned by Earth Observing-1 ran from 2000 to 2017, and was the first spaceborne hyperspectral imaging instrument. The sensor utilised two spectrometers capturing radiance in the SWIR, VNIR and visible wavelengths, resulting in 242 band imagery representing wavelengths of between 356 and 2577 nm, with a 10 nm spectral resolution and a 30m spatial resolution. Hyperion imagery is offered at three processing levels, namely L1R, L1Gst, and L1T. L1R data contains images that are radiometrically corrected, but not terrain corrected. L1T and L1Gst imagery are georeferenced and corrected radiometrically and topographically, but because of the georeferencing, the row and column orientations no longer align to track directions.

A large component of this study involved preprocessing Hyperion images to get atmospherically corrected surface reflectances. Hyperion-specific calibrations included removal of the spectral smile (desmiling) and track-direction illumination stripes (destriping) on the L1R data, and alignment of coincident VNIR and SWIR radiometric bands. Once the at-satellite radiance was corrected in this way, top-of-canopy surface reflectance could be obtained through atmospheric correction algorithms. SUREHYP, developed by Miraglio and Coops [31], was used as the consolidated procedure of algorithms for the end-to-end preprocessing of the Hyperion imagery prior to analysis. The procedure is laid out below.

A. Band removal and extraction

Out of the 242 bands in the imagery, only 200 are calibrated, namely bands 8 to 57 and 77 to 224 [32]. Other bands had to be excluded prior to preprocessing. Secondly, Hyperion data appears as digital number values, from which radiances could only be obtained by dividing the numbers by the scaling factors provided in the image metadata.

B. Desmiling

For the correction of spectral smiles present within Hyperion imagery, SUREHYP utilises San and Süzen's across-track correction algorithm [33] on the raw L1R images. The data is first transformed into the Minimum Noise Fraction (MNF) space [34], and a 2nd-degree polynomial is fitted to each MNF band's column means. If the 2nd-order coefficient was over three standard deviations from the band's mean, the polynomial is subtracted from every row of the band readings as $x'_{ij} = x_{ij} - p_i$, where x_{ij} and x'_{ij} are the smiled and desmiled MNF values for column i and row j, and p_i is column i's polynomial value. The desmiled MNF data x'_{ij} is then transformed back to the original spectral radiance space.

C. Destriping

The destriping algorithm described by Pal et al. [35] available in the SUREHYP procedure was used next. This algorithm has two parts. Firstly, global destriping was performed. For each image band k, the image column means $P_{ca,ik}$ were calculated, and the breadth of the biggest trough or ridge n_c was found via the automated *find_peaks* algorithm in SciPy [36] with n_c given by the median of the 20% largest breadths, rounded up to a whole number. Afterwards, a 2nd-order Savitzky-Golay filter [37] with a window length of $10n_c$ was applied on each band's column means to get a smoothed curve $P_{fit,ik}$. The globally-destriped image could then be found as $x'_{ijk} = x_{ijk} + \left(P_{fit,ik} - P_{ca,ik}\right)$, where x_{ijk} and x'_{ijk} are the original and updated values for a pixel in row *i* and column *j*.

Secondly, local image striping not spanning entire columns was removed. For a given band, a moving window of size $[3n_c, 3n_c]$ was utilised to find $I_{Corr\_mean,ijk}$ and $I_{Corr\_stdev,ijk}$, the local mean and standard deviation of windowed pixels. Any

pixel x'_{ijk} where $\left|x'_{ijk} - I_{Corr\_mean,ijk}\right| > I_{Corr\_stdev,ijk}$ was flagged as an outlier. If 90% of a moving window column's pixels were flagged in this way, the full column was considered a local stripe, and its values x'_{ijk} were replaced by $I_{Corr\_mean,ijk}$.

D. Alignment of VNIR and SWIR and georeferencing

A one-pixel along-track geometric misalignment is found between the 128th and 129th pixels in Hyperion's SWIR bands. This was corrected by shifting all the SWIR array elements after 128 up by one pixel. The SWIR spectrometer is also marginally misaligned from the VNIR spectrometer [38]. The spectrally corrected L1R data is georectified to the georeferenced L1T data using the 833nm bands, and applying a homography between the images with the corners as key points.

E. Atmospheric correction

Hyperion imagery data is captured as at-satellite top-of-atmosphere radiance, and requires an atmospheric RTM to calibrate the imagery to the top-of-canopy level. SUREHYP utilises the freely available SMARTS RTM (https://www.nrel.gov/grid/solar-resource/smarts.html), a physically-based model commonly used for atmospheric correction [39-41]. The Hyperion imagery metadata provided the atmospheric and illumination input variables corresponding to the image capture. These included atmospheric vapour and ozone concentrations, solar and acquisition geometries, topographic slope and altitude data and acquisition dates. SMARTS-applicable terrain altitudes and slope angles were calculated from publicly available Google Earth Engine data. Air water vapour levels were gauged from vapour absorption spectra through a SMARTS-generated Look-Up Table. The impact of neighbouring pixels was not considered, and the top-of-atmosphere and the top-of-canopy reflectances, ρ_{TOA} and ρ_{surf}, was calculated as

$$\begin{aligned} \rho_{TOA} &= \frac{\pi L d^2}{E_{\text{sun}}\cos(\theta_z)} \\ \rho_{\text{surf}} &= \frac{\pi(L - L_{\text{cirrus}} - L_{\text{haze}})}{\tau_{gs}\left(\tau_{\text{sg}} E_{\text{sun}}\cos(\theta_z) + E_{\text{dif}} + \hat{\rho}_{\text{terrain}} E_g\right)}, \end{aligned} \quad (1)$$

where L is the at-sensor radiance, d is the correction factor for Sun-Earth distance, E_{sun} is the out-of-atmosphere irradiance, θ_z is the solar zenith angle, L_{cirrus} is the radiance of transparent thin cirrus clouds, L_{haze} is the aerosol haze radiance, τ_{gs} is the optical path transmittance between the ground and the sensor, τ_{sg} is the optical path transmittance between the sun and the ground, E_{dif} is diffuse irradiance, $\hat{\rho}_{\text{terrain}}$ is the reflectance of the local terrain, and E_g is the ground global irradiance which is zero when the surface is flat. SMARTS provides τ_{sg}, τ_{gs}, E_{sun}, and $E_{\text{dif}} + \hat{\rho}_{\text{terrain}} E_g$, while θ_z is a sensor acquisition parameter. Removal of the L_{cirrus} and L_{haze} signals was the next step needed for atmospheric correction to be performed.

F. Thin cirrus and haze correction

Gao et al. proposed a method using the 1380nm band to correct for the presence of thin cirrus clouds. This is the method incorporated in SUREHYP and implemented in this study. ρ_{TOA} was computed for the entire radiance pixel set, and reflectances at the 1380nm wavelength were plotted against those at other wavelengths to get the slope factor K_a. Cirrus reflectances could then be corrected with the equation $x'_{ij} = x_{ij} - \frac{\rho_{TOA,1380}}{K_{a,k}}$, and L_{cirrus} was subtracted as per the ρ_{surf} equation (1).

Haze removal was performed through SUREHYP's implemented dark object subtraction as formulated by Chavez [42]. The basis for this technique is that if a scene is sufficiently large, it will likely include dark pixels with reflectances (and therefore measured radiances) close to zero throughout the spectrum. Minimum radiance values are stored for all bands, and since haze is caused by atmospheric photon scattering, the relative scattering follows a wavelength power scaling law λ^c, where c is a negative value informed by atmospheric haze conditions. The aerosol radiance L_{haze} could then be computed by fitting the largest possible $A\lambda^c$ function bounded above by the stored series of minimum radiances. This aerosol radiance was subtracted from the at-sensor radiance as per equation (1).

G. Water vapor retrieval

SUREHYP's water vapour retrieval method makes use of the 940nm and 1120nm bands where vapour absorbs light. A SMARTS Look-up table was created using known parameters with vapour concentrations simulated between 0cm and 12cm. Water vapour concentrations are linked with the relative spectral absorption depths, defined as the relative amplitude between the absorption band and the mean radiance of neighbouring vapour non-absorbing bands. By default, SUREHYP uses the 1120nm band, but uses the 940nm band as a backup if the 1120nm band is saturated. Average water vapour levels were computed for the scene by retrieving LUT parameters yielding the closest absorption depths.

H. Terrain correction

As per the SUREHYP procedure, elevations and the slope angles and azimuths were found using another SMARTS look-up table simulated with varying terrain altitude, slope and azimuth values. The SMARTS outputs were then used to get each pixel's irradiance values ρ_{rough}. SUREHYP uses the modified Minnaert approach suggested by Richter [43] to calculate the bidirectional reflectance distribution function for non-Lambertian rough terrains. The corrected terrain reflectance ρ_{MM} was then calculated using the solar incidence angle β, the solar zenith angle θ_S, and a wavelength λ as

$$\rho_{MM} = \rho_{\text{rough}}\left(\frac{\cos(\beta)}{\cos(\beta_T)}\right)^b, \quad (2)$$

$$\text{where } \beta_T = \begin{cases} \theta_S + 20^\circ \text{ if } \theta_S < 45^\circ \\ \theta_S + 15^\circ \text{ if } 45^\circ \le \theta_S < 55^\circ \\ \theta_S + 10^\circ \text{ elsewhere} \end{cases}$$

$$\text{and } b = \begin{cases} 0.5 \text{ without vegetation} \\ 0.75 \text{ for vegetation; } \lambda < 720 \text{ nm}, \\ 0.33 \text{ for vegetation; } \lambda \ge 720 \text{ nm} \end{cases}$$

Following this correction, the Hyperion image could be utilised for further processing. This started with segmenting the processed Hyperion images into smaller zones to obtain soil spectra for images containing diverse landscapes, soil types or soil moisture content levels.

III. Algorithms

A. SLIC segmentation

Simple Linear Iterative Clustering (SLIC) is a segmentation algorithm put forward by Achanta et al. [44,45], which implements a k-means grouping method to demarcate image segments containing pixels in the same vicinity with similar spectral characteristics. These segments are called superpixels. Using superpixels was appropriate in this study, because delineating pixel groups allowed for more refinement in spectral simulation, and allowed for a more definite discrimination between different soil types within a single image. An input parameter L informs approximately how many superpixels the full image is divided into. The full image is then split into a grid of L clusters $l \in \{1, 2, \ldots, L\}$, where G represents the grid step size. For each cluster l, a cluster centre $C_l \in [R_l, y_l, x_l]$ is chosen, where R_l is the full signature of pixel l of the image, while y_l and x_l represent the row and column values of l in the image. Then for every pixel i in a 2G × 2G neighbourhood of C_l, the distance from i to C_l is computed, and if C_l is the cluster centre to which pixel i's spectral signature is closest, i is reassigned to cluster l. The cluster centres C_l can then be recalculated as the average spectral signature and location of pixels assigned to the cluster. This process is repeated until pixel distances fall below a specified error threshold. Obtained superpixels can then be analysed individually to gauge soil characteristics.

B. BSM Model

The BSM model [23,24] is a radiative transfer model used to simulate soil spectra based on various physical and structural parameters. The model characterises soil spectra through brightness and shape transformations, as well as a calibration for moisture content. Firstly, the model simulates dry soil spectra through a linear combination of three base soil spectra calculated from the global soil vector library (GSV) developed by Jiang and Fang [6], as shown in Fig. 1. The linear combination of these spectra is given by $R_{dry} = \sum_{i=1}^{3} f_i V_i$, where V_i are the base soil vectors, and f_i are the three coefficients calculated as

$$f = \begin{bmatrix} B.\sin(\varphi) \\ B.\cos(\varphi)\sin(\lambda) \\ B.\cos(\varphi)\cos(\lambda) \end{bmatrix},$$

where B is a dimensionless soil brightness parameter, and φ and λ inform spectral shape. The model then uses these dry vectors together with a soil moisture content paramater SM_p to simulate the wet soil reflectance spectra R_{soil} via a physically-based approach as

$$R_{\text{soil}} = R_{\text{dry}} P(k=0) + \sum_{k=1}^{\infty} R_{\text{swet}}(k) P(k), \quad (3)$$

$$\text{where } R_{\text{swet}}(k) = \rho_{12} + \frac{(1-\rho_{12}) \exp(-2\kappa_w k \Delta) R_{\text{bac}} (1-\rho_{21})}{1 - \rho_{21} \exp(-2\kappa_w k \Delta) R_{\text{bac}}}$$

where ρ_{12} and ρ_{21} are the air-water and water-air Fresnel reflectances, R_{bac} is the background soil reflectance, which is the dry soil reflectance spectrum plus the water-soil Fresnel reflectance. κ_w is the water absorption spectrum, Δ is the optical thickness of a single water film (usually set to a constant), and k is the number of water films around the soil particles, which follows a Poisson distribution given by

$$P(k) = \frac{e^{-\mu} \mu^k}{k!} \; ; \; \mu = \frac{w - 5}{SM_c}, \quad (4)$$

where w is the volumetric moisture carrying capacity of the soil, usually set to 25%.

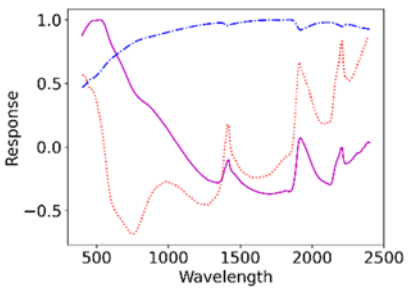

***Fig. 1**. GSV base soil vectors combined to model dry soils*

IV. Simulation (forward modelling) of soil spectra

Hyperion imagery was obtained for Southern Africa from the USGS Earth Explorer database, as shown in Fig. 2. This included the countries of South Africa, Lesotho, Swaziland, Namibia and Botswana.

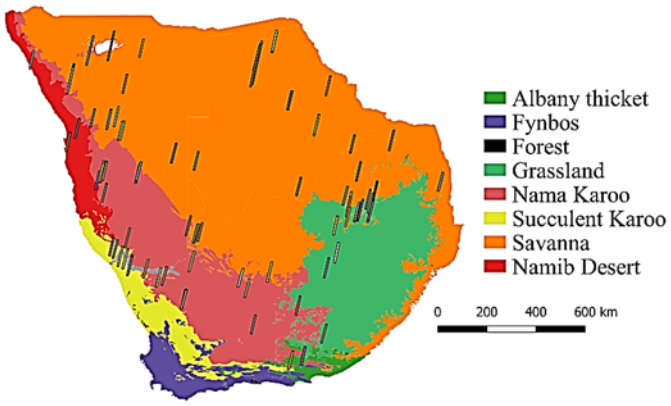

Fig. 2. Hyperion imagery within Southern African Biomes:

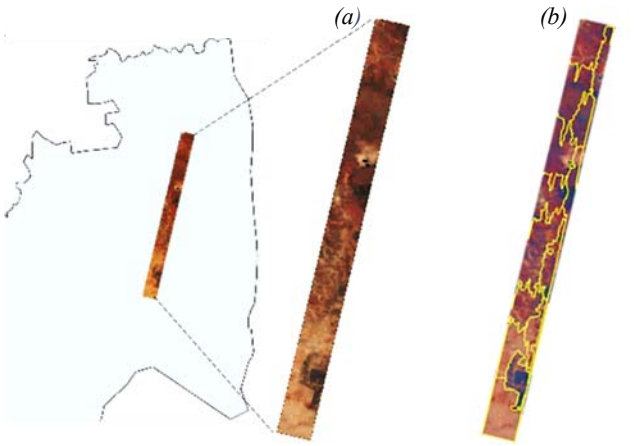

Fig. 3. *(a) Pre-processed Hyperion image in Mpumalanga, South Africa (b) SLIC segmented Hyperion image*

TABLE I:
BIOME CLASSIFICATIONS WITH TRAINING/TESTING SPLIT

Biome	Superpixels	Training set	Testing set
Desert	98	**66**	**32**
Grassland	217	**145**	**72**
Nama Karoo	446	**298**	**148**
Savanna	905	**604**	**301**
Succulent Karoo	99	**66**	**33**

Hyperion pixel responses were regularised to values between 0 and 10, and bands with anomalous spectra were excluded through simple thresholding such that responses are between 0.01 and 0.6, and the absolute differences between consecutive bands remain below 0.1. Each image was segmented using the SLIC algorithm with 20 superpixels as shown in Fig. 3. A total of 1765 superpixel parameter sets were obtained like this way, which were categorised into 5 different biomes based on the geographic location of the original Hyperion imagery. These biomes are listed in Table I with their superpixel look-ups.

Average spectra are found for each segment's pixels with cellulose absorption index (CAI) values below -1% and normalized difference vegetation index (NDVI) values between -0.046 and 0.25, where lower NDVI values represent water, and higher values represent photosynthetic vegetation [47]. As in [48], Hyperion-derived CAI is given by

$$\text{CAI} = \frac{\hat{\rho}_{2000} + \hat{\rho}_{2200}}{2} - \hat{\rho}_{2100},$$
$$\text{where} \quad \hat{\rho}_{2000} = \frac{\rho_{1982}+\rho_{1992}+\rho_{2002}}{3}, \quad (5)$$
$$\hat{\rho}_{2100} = \frac{\rho_{2103}+\rho_{2113}+\rho_{2214}}{3},$$
$$\hat{\rho}_{2200} = \frac{\rho_{2194}+\rho_{2204}+\rho_{2214}}{3}.$$

The CAI measure tests for increased spectral absorption around the 2100nm wavelength brought about by cellulose and lignin in senescent vegetation but not soil. The -1% threshold was suggested by Daughtry [49] to discriminate between their spectral signatures.

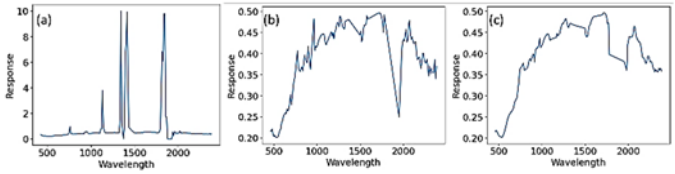

Fig. 4. Example of Hyperion superpixel spectrum
(a) Regularised; (b) Thresholded; (c) Filtered (Savitzky-Golay)

NDVI for Hyperion data is given by Pervez and Khan [46] as

$$\text{NDVI} = \frac{\rho_{925} - \rho_{651}}{\rho_{925} + \rho_{651}}. \quad (6)$$

These average spectra are then smoothed using a linear Savitzky-Golay filter with a window size of 5 to decrease spectral noise. This regularising, thresholding and filtering process is shown in in Fig. 4.

The BSM model as implemented in the SPART package in Python [24] was used in the RTM simulation of soil spectra. The soil moisture capacity parameter SM_c was set to the default value of 25%, while the water film optical thickness parameter was set to 0.01cm. 10 000 different soil spectra were simulated using B, φ, λ and w ranges as shown in Table II. Simulated look-up table spectra are then recalibrated to Hyperion spectral band responses, estimated using the provided band central wavelengths and full width at half maximum (FWHM) measures as provided in the Hyperion L1R metadata. Average spectra for each superpixel in the training set are then associated with parameter values from the look-up table with the closest spectra to the superpixel ones. An example of this is shown in Fig. 5, where a processed spectrum is matched to a simulated one from the look-up table to obtain its parameters. Although some spectral noise is still present, each superpixel's mean soil spectrum could be modelled to a relative root mean squared error of 1%, affirming the ability of BSM and the basis vectors to simulate Southern African spectra.

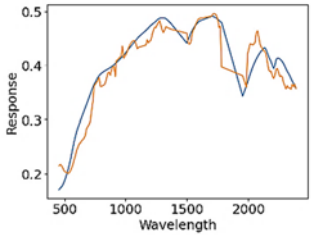

Fig. 5. Example of a Processed Hyperion spectrum (orange) matched to its closest look-up table spectrum (blue)

TABLE II:
LOOK-UP TABLE PARAMETERS FOR THE BSM MODEL

Parameter	Minimum	Maximum	Distribution
B	0.25	0.9	Uniform
φ	-30	30	Uniform
λ	80	120	Uniform
w	5%	75%	Uniform
SM_c	25%	25%	Constant
Δ	0.01cm	0.01cm	Constant

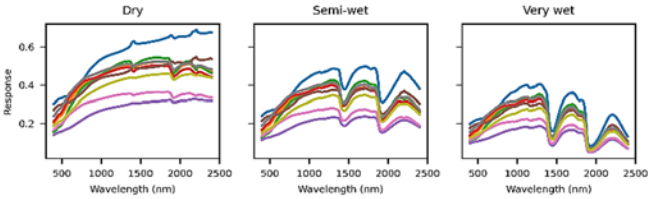

Fig. 6. Example of biome-specific simulated spectra for grassland soil in dry, semi-wet and very wet states

Each superpixel's mean soil spectrum could be modelled to a relative root mean squared error of 1%, affirming the ability of BSM and the basis vectors to simulate Southern African spectra. A multivariable distribution function for each biome's parameter vector $\Theta = [B, \varphi, \lambda, w]$ could then be modelled based on each biome's retrieved parameters, together with their covariances. The estimated multivariate distribution functions $\widehat{\Theta}$ for each of the biomes are

$$\widehat{\Theta}_{desert} \sim \mathcal{N}\left(\begin{bmatrix} 0.596 \\ -1.77 \\ 88.8 \\ 14.4 \end{bmatrix}, \begin{bmatrix} 0.018 & 0.253 & 0.112 & -0.855 \\ 0.253 & 10.2 & 3.62 & -42.0 \\ 0.112 & 3.62 & 15.0 & -19.3 \\ -0.855 & -42.0 & -19.3 & 192.4 \end{bmatrix}\right)$$

$$\widehat{\Theta}_{grassland} \sim \mathcal{N}\left(\begin{bmatrix} 0.498 \\ -2.50 \\ 85.0 \\ 52.9 \end{bmatrix}, \begin{bmatrix} 0.012 & -0.077 & -0.112 & 0.491 \\ -0.077 & 6.38 & -0.724 & 3.21 \\ -0.112 & -0.724 & 17.6 & -33.1 \\ 0.491 & 3.21 & -33.1 & 364.4 \end{bmatrix}\right)$$

$$\widehat{\Theta}_{nama\ karoo} \sim \mathcal{N}\left(\begin{bmatrix} 0.54 \\ -1.32 \\ 88.4 \\ 22.3 \end{bmatrix}, \begin{bmatrix} 0.012 & -0.07 & 0.142 & -0.301 \\ -0.07 & 1.93 & -1.94 & 6.33 \\ 0.142 & -1.94 & 25.7 & -48.0 \\ -0.301 & 6.33 & -48.0 & 166.9 \end{bmatrix}\right) \quad (6)$$

$$\widehat{\Theta}_{savanna} \sim \mathcal{N}\left(\begin{bmatrix} 0.549 \\ -3.31 \\ 86.1 \\ 38.3 \end{bmatrix}, \begin{bmatrix} 0.019 & 0.06 & 0.119 & -1.23 \\ 0.06 & 19.8 & -0.379 & 7.15 \\ 0.119 & -0.379 & 25.2 & -60.0 \\ -1.23 & 7.15 & -60.0 & 50.1 \end{bmatrix}\right)$$

$$\widehat{\Theta}_{succ.\ karoo} \sim \mathcal{N}\left(\begin{bmatrix} 0.513 \\ -1.82 \\ 90.5 \\ 14.2 \end{bmatrix}, \begin{bmatrix} 0.01 & -0.096 & -0.044 & 0.043 \\ -0.096 & 3.69 & -0.399 & 1.38 \\ 0.119 & -0.399 & 22.7 & -17.3 \\ 0.043 & 1.38 & -17.3 & 50.5 \end{bmatrix}\right),$$

where once again, SM_c and Δ are kept constant. An example of biome-specific BSM simulated spectra is provided in Fig. 6 for grassland biome soils with soil moisture levels of 5%, 25% and 65%, representing dry, semi-wet and very wet soils.

As shown in Table II, a biome-specific look-up table of 1000 values was created to simulate each biome's soil spectra more accurately than a 1000-value general look-up. More specifically, for each spectrum R in the testing data set belonging to biome b, a relative root mean relative square error (RMRSE) is found between R and $\hat{R}$, and between R and $\hat{R}_b$, where $\hat{R}$ is the closest spectrum from the full BSM table, and $\hat{R}_b$ is the closest spectrum from the biome-specific look-up table. For any test spectrum j, the RMRSEs are given by

$$RMRSE_{b,j} = \sqrt{\frac{\sum_i (R_i - \hat{R}_{b,i})^2}{n \cdot \sum_i \hat{R}_i^{\,2}}}$$
$$RMRSE_{f,j} = \sqrt{\frac{\sum_i (R_i - \hat{R}_i)^2}{n \cdot \sum_i \hat{R}_i^{\,2}}}, \quad (7)$$

where $RMRSE_{b,j}$ and $RMRSE_{f,j}$ denote the RMRSE's under the full and biome-specific parameter regimes respectively, and the subscript i denotes each spectral value comprising the spectrum vector.

A null hypothesis could therefore be created as follows:

$$\mathbf{H_0}: \exists\ b \in biomes: \mathbb{E}[RMRSE_b \mid b\,] = \mathbb{E}\big[RMRSE_f \mid b\,\big]$$
$$\mathbf{H_1}: \forall\ b \in biomes: \mathbb{E}[RMRSE_b \mid b\,] < \mathbb{E}\big[RMRSE_f \mid b\,\big] \quad (8)$$

Since the full and biome-specific BSM simulations are used to simulate the same spectral vectors, a dependent t-test for paired samples is used to test the hypothesis, since simulation outputs are highly correlated and centred around the same target. p-scores for each biome b's t-test is provided in Table III. Assuming a 99% one-sided confidence interval is required, a p-score above 0.01 for any biome b would be required not to reject the null hypothesis. Since this is not the case, the null hypothesis is rejected, and the authors conclude that the biome-specific parameter regime does indeed yield more accurate simulations than the full parameter regime.

TABLE III:
SPECTRA SIMULATION ERRORS UNDER FULL AND SPECIFIC PARAMETER REGIMES

Biome	RRMSE - full		RRMSE - specific		Student's
	Mean	Std error	Mean	Std error	t-test p-score
Desert	1.02%	0.03%	0.71%	0.05%	2.93×10^{-11}
Grassland	1.57%	0.12%	1.34%	0.10%	1.78×10^{-10}
Nama Karoo	1.00%	0.03%	0.84%	0.02%	9.23×10^{-45}
Savanna	2.01%	0.41%	1.29%	0.15%	8.32×10^{-3}
Succulent Karoo	0.97%	0.05%	0.8%	0.04%	2.03×10^{-11}

V. DISCUSSION AND CONCLUSION

In terms of the objective of simulating soil spectra for specified semi-arid and arid biomes within Southern Africa, the brightness-shape-moisture radiative transfer modelling of soils was the topic of focus. Hyperion data was pre-processed using the SUREHYP procedure, filtered via NDVI and CAI thresholding, segmented using the SLIC algorithm, filtered and parameterised via a Hyperion band calibrated BSM model look-up table to obtain biome-specific simulation parameter distributions. These parameter distributions were tested against the full BSM parameter distributions for simulation of actual preprocessed Hyperion spectra, and the biome-specific parameter distributions were found to yield higher simulation accuracies than the full distributions for a smaller look-up table.

This provides a strong computational advantage over full BSM soil spectral simulations, and will allow for improved mapping of soil moisture content and zoning of soil types in the studied semi-arid zones of Southern Africa. Since soil moisture content is one of the model's parameters, it can also be adjusted to simulate the soils at any given moisture level. Simulated pure soil spectra can then be used as a more accurate basis in canopy RTMs such as PROSAIL and SCOPE, where local soil spectra inform the true overall canopy spectra separately from green or senescent plant matter for a given study area.

Although simulation results align well to Hyperion spectra, true ground truth validation for spectra, moisture, or other properties is made inviable due to the legacy nature of Hyperion data. The accuracy and efficiency of Hyperion-derived BSM parameters in modelling biome-specific soil sample spectra or tested water content remains a suitable topic for further investigation. Characteristic soil parameters may also provide further insight into biophysical and biochemical properties of soils in different biomes, and the modelling of the relationship between soil physical properties and characteristics against local BSM parameters presents a suitable topic for further research. The methodology could also be extended to retrieve BSM parameters from multispectral imagery with fewer bands.

ACKNOWLEDGEMENTS

The authors are grateful to the CSIR and the National Research Foundation New Earth Observation Frontiers innovation funding mechanism for funding the research contained herein.

REFERENCES

[1] Ben-Dor, E.; Banin, A. Near-infrared Analysis as a Rapid Method to Simultaneously Evaluate several Soil Properties. Soil Sci. Soc. Am. J. **1995**, *59*, 364-372.

[2] Haubrock, S.; Chabrillat, S.; Kuhnert, M.; Hostert, P.; Kaufmann, H.J. Surface Soil Moisture Quantification and Validation Based on Hyperspectral Data and Field Measurements. Journal of applied remote sensing **2008**, *2*, 023552.

[3] Khanna, S.; Palacios-Orueta, A.; Whiting, M.L.; Ustin, S.L.; Riaño, D.; Litago, J. Development of Angle Indexes for Soil Moisture Estimation, Dry Matter Detection and Land-Cover Discrimination. Remote Sens. Environ. **2007**, *109*, 154-165.

[4] Chabrillat, S.; Ben-Dor, E.; Rossel, R.A.V.; Demattê, J.A. Quantitative Soil Spectroscopy. Applied and Environmental Soil Science **2013**, *2013*.

[5] Zhang, J.; Zhang, Q.; Bao, A.; Wang, Y. A New Remote Sensing Dryness Index Based on the Near-Infrared and Red Spectral Space. Remote Sensing **2019**, *11*, 456.

[6] Jiang, C.; Fang, H. GSV: A General Model for Hyperspectral Soil Reflectance Simulation. International Journal of Applied Earth Observation and Geoinformation **2019**, *83*, 101932.

[7] Jacquemoud, S.; Baret, F. PROSPECT: A Model of Leaf Optical Properties Spectra. Remote Sens. Environ. **1990**, *34*, 75-91.

[8] Schlerf, M.; Atzberger, C. Inversion of a Forest Reflectance Model to Estimate Structural Canopy Variables from Hyperspectral Remote Sensing Data. Remote Sens. Environ. **2006**, *100*, 281-294.

[9] Darvishzadeh, R.; Atzberger, C.; Skidmore, A.; Schlerf, M. Mapping Grassland Leaf Area Index with Airborne Hyperspectral Imagery: A Comparison Study of Statistical Approaches and Inversion of Radiative Transfer Models. ISPRS Journal of Photogrammetry and Remote Sensing **2011**, *66*, 894-906.

[10] Stafford, J.V. Remote, Non-Contact and in-Situ Measurement of Soil Moisture Content: A Review. J. Agric. Eng. Res. **1988**, *41*, 151-172.

[11] Petropoulos, G.P.; Ireland, G.; Barrett, B. Surface Soil Moisture Retrievals from Remote Sensing: Current Status, Products & Future Trends. Physics and Chemistry of the Earth, Parts A/B/C **2015**, *83*, 36-56.

[12] Bablet, A.; Vu, P.V.H.; Jacquemoud, S.; Viallefont-Robinet, F.; Fabre, S.; Briottet, X.; Sadeghi, M.; Whiting, M.L.; Baret, F.; Tian, J. MARMIT: A Multilayer Radiative Transfer Model of Soil Reflectance to Estimate Surface Soil Moisture Content in the Solar Domain (400–2500 nm). Remote Sens. Environ. **2018**, *217*, 1-17.

[13] Dupiau, A.; Jacquemoud, S.; Briottet, X.; Fabre, S.; Viallefont-Robinet, F.; Philpot, W.; Di Biagio, C.; Hébert, M.; Formenti, P. MARMIT-2: An Improved Version of the MARMIT Model to Predict Soil Reflectance as a Function of Surface Water Content in the Solar Domain. Remote Sens. Environ. **2022**, *272*, 112951.

[14] Sadeghi, M.; Jones, S.B.; Philpot, W.D. A Linear Physically-Based Model for Remote Sensing of Soil Moisture using Short Wave Infrared Bands. Remote Sens. Environ. **2015**, *164*, 66-76.

[15] Kubelka, P. Ein Beitrag Zur Optik Der Farbanstriche. Z.tech.Phys **1931**, *12*, 593-601.

[16] Kubelka, P. New Contributions to the Optics of Intensely Light-Scattering Materials. Part I. Josa **1948**, *38*, 448-457.

[17] Ou, D.; Tan, K.; Wang, X.; Wu, Z.; Li, J.; Ding, J. Modified Soil Scattering Coefficients for Organic Matter Inversion Based on Kubelka-Munk Theory. Geoderma **2022**, *418*, 115845.

[18] Hapke, B. Bidirectional Reflectance Spectroscopy: 1. Theory. Journal of Geophysical Research: Solid Earth **1981**, *86*, 3039-3054.

[19] Hapke, B. Bidirectional Reflectance Spectroscopy: 4. the Extinction Coefficient and the Opposition Effect. Icarus **1986**, *67*, 264-280.

[20] Hapke, B. *Theory of Reflectance and Emittance Spectroscopy*.; Cambridge university press, 2012.

[21] Jacquemoud, S.; Baret, F.; Hanocq, J.F. Modeling Spectral and Bidirectional Soil Reflectance. Remote Sens. Environ. **1992**, *41*, 123-132.

[22] Verhoef, W.; van der Tol, C.; Middleton, E.M. Hyperspectral Radiative Transfer Modeling to Explore the Combined Retrieval of Biophysical Parameters and Canopy Fluorescence from FLEX – Sentinel-3 Tandem Mission Multi-Sensor Data. Remote Sens. Environ. **2018**, *204*, 942-963.

[23] Yang, P.; Prikaziuk, E.; Verhoef, W.; van Der Tol, C. SCOPE 2.0: A Model to Simulate Vegetated Land Surface Fluxes and Satellite Signals. Geoscientific Model Development **2021**, *14*, 4697-4712.

[24] Yang, P.; van der Tol, C.; Yin, T.; Verhoef, W. The SPART Model: A Soil-Plant-Atmosphere Radiative Transfer Model for Satellite Measurements in the Solar Spectrum. Remote Sens. Environ. **2020**, *247*, 111870.

[25] Döpper, V.; Rocha, A.D.; Berger, K.; Gränzig, T.; Verrelst, J.; Kleinschmit, B.; Förster, M. Estimating Soil Moisture Content Under Grassland with Hyperspectral Data using Radiative Transfer Modelling and Machine Learning. International Journal of Applied Earth Observation and Geoinformation **2022**, *110*, 102817.

[26] Bayat, B.; van der Tol, C.; Verhoef, W. Retrieval of Land Surface Properties from an Annual Time Series of Landsat TOA Radiances during a Drought Episode using Coupled Radiative Transfer Models. Remote Sens. Environ. **2020**, *238*, 110917.

[27] Wang, N.; Yang, P.; Clevers, J.G.P.W.; Wieneke, S.; Kooistra, L. Decoupling Physiological and Non-Physiological Responses of Sugar Beet to Water Stress from Sun-Induced Chlorophyll Fluorescence. Remote Sens. Environ. **2023**, *286*, 113445.

[28] Eon, R.S.; Bachmann, C.M. Mapping Barrier Island Soil Moisture using a Radiative Transfer Model of Hyperspectral Imagery from an Unmanned Aerial System. Scientific Reports **2021**, *11*, 3270.

[29] Bayat, B.; van der Tol, C.; Verhoef, W. Retrieval of Land Surface Properties from an Annual Time Series of Landsat TOA Radiances during a Drought Episode using Coupled Radiative Transfer Models. Remote Sens. Environ. **2020**, *238*, 110917.

[30] Prikaziuk, E.; Ntakos, G.; ten Den, T.; Reidsma, P.; van der Wal, T.; van der Tol, C. Using the SCOPE Model for Potato Growth, Productivity and Yield Monitoring Under Different Levels of Nitrogen Fertilization. International Journal of Applied Earth Observation and Geoinformation **2022**, *114*, 102997.

[31] Miraglio, T.; Coops, N.C. SUREHYP: An Open Source Python Package for Preprocessing Hyperion Radiance Data and Retrieving Surface Reflectance. Sensors **2022**, *22*, 9205.

[32] Datt, B.; McVicar, T.R.; Van Niel, T.G.; Jupp, D.L.; Pearlman, J.S. Preprocessing EO-1 Hyperion Hyperspectral Data to Support the Application of Agricultural Indexes. IEEE Trans. Geosci. Remote Sens. **2003**, *41*, 1246-1259.

[33] San, B.T.; Süzen, M.L. Evaluation of Cross-Track Illumination in EO-1 Hyperion Imagery for Lithological Mapping. Int. J. Remote Sens. **2011**, *32*, 7873-7889.

[34] Green, R.O.; Eastwood, M.L.; Sarture, C.M.; Chrien, T.G.; Aronsson, M.; Chippendale, B.J.; Faust, J.A.; Pavri, B.E.; Chovit, C.J.; Solis, M. Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Remote Sens. Environ. **1998**, *65*, 227-248.

[35] Pal, M.K.; Porwal, A.; Rasmussen, T.M. Noise Reduction and Destriping using Local Spatial Statistics and Quadratic Regression from Hyperion Images. Journal of Applied Remote Sensing **2020**, *14*, 016515.

[36] Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature methods **2020**, *17*, 261-272.

[37] Savitzky, A.; Golay, M.J. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. **1964**, *36*, 1627-1639.

[38] Khurshid, K.S.; Staenz, K.; Sun, L.; Neville, R.; White, H.P.; Bannari, A.; Champagne, C.M.; Hitchcock, R. Preprocessing of EO-1 Hyperion Data. Canadian Journal of Remote Sensing **2006**, *32*, 84-97.

[39] Zarco-Tejada, P.J.; González-Dugo, V.; Berni, J.A. Fluorescence, Temperature and Narrow-Band Indices Acquired from a UAV Platform for Water Stress Detection using a Micro-Hyperspectral Imager and a Thermal Camera. Remote Sens. Environ. **2012**, *117*, 322-337.

[40] Sola, I.; González-Audícana, M.; Álvarez-Mozos, J. Validation of a Simplified Model to Generate Multispectral Synthetic Images. Remote Sensing **2015**, *7*, 2942-2951.

[41] Zarco-Tejada, P.J.; Hornero, A.; Hernández-Clemente, R.; Beck, P. Understanding the Temporal Dimension of the Red-Edge Spectral Region for Forest Decline Detection using High-Resolution Hyperspectral and Sentinel-2a Imagery. ISPRS Journal of Photogrammetry and Remote Sensing **2018**, *137*, 134-148.

[42] Chavez Jr, P.S. An Improved Dark-Object Subtraction Technique for Atmospheric Scattering Correction of Multispectral Data. Remote Sens. Environ. **1988**, *24*, 459-479.

[43] Richter, R.; Kellenberger, T.; Kaufmann, H. Comparison of Topographic Correction Methods. Remote Sensing **2009**, *1*, 184-196.

[44] Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. IEEE Trans. Pattern Anal. Mach. Intell. **2012**, *34*, 2274-2282.

[45] Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels. EPFL Technical Report no. 149300 **2010**.

[46] Pervez, W.; Khan, S.A. Hyperspectral Hyperion Imagery Analysis and its Application using Spectral Analysis. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **2015**, *40*, 169-175.

[47] Mehta, A.; Shukla, S.; Rakholia, S. Vegetation Change Analysis using Normalized Difference Vegetation Index and Land Surface Temperature in Greater Gir Landscape. Journal of Scientific Research **2021**, *65*.

[48] Daughtry, C.S.; Doraiswamy, P.C.; Hunt Jr, E.R.; Stern, A.J.; McMurtrey Iii, J.E.; Prueger, J.H. Remote Sensing of Crop Residue Cover and Soil Tillage Intensity. Soil Tillage Res. **2006**, *91*, 101-108.

[49] Daughtry, C.S.T. Discriminating Crop Residues from Soil by Shortwave Infrared Reflectance. Agron. J. **2001**, *93*, 125-131.

Wessel Bonnet received a B.Sc. in Financial Mathematics from the University of Johannesburg in 2010, the B.Sc. degree (Hons.) in Advanced Mathematics of Finance from the University of the Witwatersrand in 2011, and the M.Sc. degree (cum laude) in eScience from the University of the Witwatersrand in 2011. He is currently pursuing the Ph.D. degree in Plant and Soil Sciences at the University of Pretoria. He joined the Council for Scientific and Industrial Research in 2022.

His research interests include remote sensing for soil and plant biophysical and biochemical mapping, machine learning, and remote sensing image processing.

Moses Azong Cho received the B.Sc. degree in natural sciences from the University of Yaoundé, Cameroon in 1991, the M.Sc. degree in biodiversity conservation from the University of Greenwich, London, U.K., in 2001 and the Ph.D. degree in Hyperspectral Remote Sensing from Wageningen University, The Netherlands and the International Institute from Geoinformation Science and Earth Observation, Enschede, The Netherlands in 2007.

He joined the Council for Scientific and Industrial Research, South Africa in 2007 where he currently serves as a Chief research scientist in Earth Observation technologies. He is also an extraordinary Professor in remote sensing of vegetation with the Department of Plant and Soil Sciences, University of Pretoria and a research fellow the University of KwaZulu-Natal since 2010. His research interests include the development of remote sensing algorithms for assessing and mapping vegetation biochemical and biophysical properties, and for discriminating vegetation types or species.

Moses is a member of the African Association of Remote Sensing of the Environment (AARSE) and African chapter president of the International Association of Landscape Ecologists (IALE). He is an associate editor of the International Journal of Applied Earth Observation and Geoinformation.

Cecilia Masemola received her B.Sc. degree in microbiology and biochemistry in 2007, a B.Sc. (Hons.) degree in environmental management in 2008, an M.Sc. degree (cum laude) in environmental science in 2015, and later completed her Ph.D at the University of South Africa. In 2015, she joined the Council for Scientific and Industrial Research (CSIR), initially as a Ph.D. student. Later, she became a researcher in earth observation technologies.

She is currently a senior researcher at the CSIR's Precision Agriculture Unit. Her research interests revolve around the improvement and development of remote sensing algorithms. Specifically, she focuses on quantifying and distinguishing invasive alien plant species, such as *Acacia spp*, based on their biochemical and biophysical properties, from the leaf to the landscape level.

Cecilia is an esteemed member of Agro-Geoinfomatics and the International Association of Landscape Ecologists (IALE).