

Automated mapping of regolith units with support vector machine and artificial neural network using data from Landsat-8 OLI, ALOS PALSAR DEM, and Sentinel-1A radar images: the case of the Sissingué Gold Project, Northern Côte d'Ivoire

H. A. Diaby, F. Kouamé, B. Saley & K. O. Ansah

To cite this article: H. A. Diaby, F. Kouamé, B. Saley & K. O. Ansah (2023) Automated mapping of regolith units with support vector machine and artificial neural network using data from Landsat-8 OLI, ALOS PALSAR DEM, and Sentinel-1A radar images: the case of the Sissingué Gold Project, Northern Côte d'Ivoire, International Journal of Remote Sensing, 44:22, 7126-7155, DOI: [10.1080/01431161.2023.2282406](https://doi.org/10.1080/01431161.2023.2282406)

To link to this article: <https://doi.org/10.1080/01431161.2023.2282406>



Published online: 27 Nov 2023.



Submit your article to this journal [↗](#)



Article views: 118



View related articles [↗](#)



View Crossmark data [↗](#)



Automated mapping of regolith units with support vector machine and artificial neural network using data from Landsat-8 OLI, ALOS PALSAR DEM, and Sentinel-1A radar images: the case of the Sissingué Gold Project, Northern Côte d'Ivoire

H. A. Diaby^a, F. Kouamé^a, B. Saley^a and K. O. Ansah^b

^aCentre Universitaire de Recherche et d'Application en Télédétection (CURAT), University of Cocody, Abidjan, Côte d'Ivoire; ^bDepartment of Earth Science, University of Ghana, Accra, Ghana

ABSTRACT

The Sissingué Gold Project is in Northern Côte d'Ivoire and part of the southern extension, a large peneplain of the landscape in South Mali. The area has experienced extensive weathering and erosion over a long time, resulting in a widespread and complex regolith cover. This diverse and vast regolith cover poses considerable difficulties for exploration. To overcome this problem, we adopted mapping of regolith units using machine learning (ML) with support vector machine (SVM) and artificial neural network (ANN) algorithms. The main objectives were to (1) make the predictive regolith map of the study region using these algorithms with data obtained from Landsat-8 operational land imager, Advanced Land Observing Satellite Phased Array Type L-Band Synthetic Aperture Radar digital elevation model and Sentinel-1A; (2) show the method of pre-processing and processing the required data and (3) develop the regolith landform unit (RLU) map from the predictive regolith map and collected data based on the Relict – Erosional – Depositional model. Tests using the SVM and ANN algorithms showed that both ML tools could accurately map regolith landforms. An innovative method of optimizing data using parameters is presented. The results showed that ANN outperformed SVM with an overall accuracy of 87.01% and a kappa coefficient of 0.84, whereas the corresponding values for SVM were 86.69% and 0.84, respectively. However, the validation data obtained from SVM-based prediction exhibited a better score than that of validation data obtained from ANN-based prediction. The Sentinel-1A radar band combined with Landsat-8 data reduced the vegetation-masking effect and improved the classification results. The RLUs of this area are composed primarily of relicts at 24.91%, including lateritic residuum and soil, depositional material at 36.39% with exotic

ARTICLE HISTORY

Received 7 February 2023
Accepted 29 October 2023

KEYWORDS

Support vector machines;
neural networks; remote
sensing

CONTACT H. A. Diaby ✉ alidiaby@gmail.com Centre Universitaire de Recherche et d'Application en Télédétection (CURAT), University of Cocody, 03 BP 224 Cidex Abidjan 03, Abidjan, Côte d'Ivoire

© 2023 Informa UK Limited, trading as Taylor & Francis Group

sediments and ferrierite and the remaining 38.7% of erosional material made of saprolite, colluvial fragments, and mottled zones on flanks.

1. Introduction

Regolith covers lands in Africa, Australia, India, and South America. It covers one-third of each continent (Tardy 1997). It has formed continuously for more than 100 million years in various climates and regions and is defined as the cover of altered and transported materials overlying sound rock. It also includes weathered and fissured rock, saprolite, soil, colluvium, alluvium and aeolian deposits (Anand and Paine 2002). Regoliths can act as a barrier to the discovery of mineralization and geological mapping (Metelka et al. 2018). This may lead to difficulties in the former case. Therefore, it is important to map regoliths in most mineral exploration projects. Regolith mapping involves the extensive use of remote sensing, field mapping, and satellite data (Anand and Paine 2002; Arhin et al. 2015; Butt 2016; Chardon et al. 2018; De Boissieu et al. 2017; Metelka et al. 2018; Sevin 2014).

Machine learning (ML) algorithms (MLAs) are increasingly used to perform supervised and unsupervised classifications, such as regolith (Cudahy 2016; De Boissieu et al. 2017; Metelka et al. 2018) and geological mapping (Cracknell and Reading 2014). Metelka et al. (2018) exploited artificial neural networks (ANNs) using different types of information. Radiometric data and images obtained from Landsat, Advanced Spaceborne Thermal Emission Reflection Radiometer (ASTER), Advanced Land Observing Satellite Phased Array Type L-Band Synthetic Aperture Radar (ALOS PALSAR), Radarsat-2 and Shuttle Radar Topography Mission (SRTM) were used to create a prediction map of regolith units. The data validation and training samples were obtained from the Gaoua region of Burkina Faso. This study achieved an overall accuracy (OA) of 94.21% with the confusion matrix and a kappa coefficient of 0.92 with the combined data obtained from γ -ray radiometry, the SRTM elevation model, Landsat imagery and polarimetric radar data from ALOS PALSAR/Radarsat-2. However, the author pointed out the difficulties associated with the presence of vegetation and the size of outcrops, which are often difficult to identify.

To produce a regolith map of a nickel prospect in New Caledonia, Sevin (2017) used HyVista hyperspectral data, 36 field-control data, SRTM data and a support vector machine (SVM). Six regolith units were mapped with an accuracy of 70%. The difficulties encountered were related to the discrimination of the pisolite, red laterite, and saprolite classes, which exhibited average accuracies of 67–69%. This performance is linked to the similarity of their mineralogical compositions to those of yellow laterite, ferricrete, and saprolite. It was concluded that vegetation acts as a brake in this method, which was onerous. Arhin et al. (2013) developed a regolith landform map to guide mineral exploration in the Lawra greenstone belt of Ghana. The authors used SRTM data to delineate alluvial zones and high-glacial or residual zones. Landsat-7 images were integrated into the Geographic Information System (GIS) and SRTM to visually interpret other areas through digitization. Finally, this study used radiometric data to delimit the residual and

non-residual units, which were transported to integrate them into the previous map, known as the preliminary map. This preliminary map was superimposed on the non-genetic regolith map using GIS to validate the interpretation. Merging the non-genetic map and interpretative or preliminary map resulted in a map being used to create the regolith landform map. The authors also confirmed that this methodology was only adoptable in savannah areas using the data. According to Anand and Paine (2002) and Arhin et al. (2015), two main models exist for creating regolith maps, which are divided based on their units, i.e. Relict – Erosional – Depositional (RED) and Ferruginous – Relict–Erosional–Depositional (FRED) models.

The Sissingué Gold Project (SGP), which hosts the Sissingué mine in northern Côte d'Ivoire in the Boundiali greenstone belt, is ideal for testing and evaluating MLAs in regolith mapping. It has moderate vegetation and deep weathering, similar to Mali and Burkina, and lacks regolith mapping covering the study area. Previous studies on MLAs have reported problems with vegetation and difficulty in identifying certain classes of regolith, such as outcrops, and the accuracy of the classes. According to Cardoso-Fernandes, Teodoro, and Lima (2019) and Cracknell and Reading (2014), an SVM is relatively more robust than an ANN, particularly when faced with limited training data.

Herein, we exploited ANN and SVM with data obtained from Landsat-8 operational land imager (OLI), Advanced Land Observing Satellite Phased Array Type L-Band Synthetic Aperture Radar (ALOS PALSAR) digital elevation model (DEM), and Sentinel-1A (S1A) to perform predictive regolith mapping and develop an interpretative regolith map of landforms respecting either the RED or FRED model. In this study, we also attempt to improve the classification results and develop a methodology accessible to geologists and explorers wishing to collect better geoscientific information on soils or materials with origins which remain problematic.

2. Regional setting

The SGP covers an area of 446 km² and is located approximately 600 km north of Abidjan, the capital city of Cote d'Ivoire (Figure 1). The study area is within the department of Tengrela, which is part of the Bagoué region, in the Savanna District. The SGP is limited to the east by the Bagoué River, which forms the geographical border between Mali and Cote d'Ivoire (Figure 2).

Temperatures range from 26° in July to 31° in April (source : Société d'Exploitation et de développement Aéroportuaire, Aéronautique et Météorologique [SODEXAM]) in the SGP area. Precipitation reached 3433 mm per annum, peaking at 523 mm in August (source: SODEXAM). The climate is equatorial with a dry winter (Kottek et al. 2006) in this part of the country. The SGP landform is shaped by the Bagoué River following low-lying terrains with an elevation of 310 m, whereas a few laterite hills can reach 370 m of elevation.

The project is situated in the southern part of the West African craton and is referred to as the Man-Leo Shield (Parra-Avila et al. 2016). The Man-Leo Shield comprises the western portion of the Archean Kenema-Man Domain (Figure 1), which dominates the geology of Liberia, Sierra Leone, and Guinea (Feybesse and Milési 1994). The SGP area lies on the northwest edge of a 100 km long Birimian greenstone belt, defined as the Bouaflé volcano-sedimentary belt (Figure 2). The greenstone belt is north-northeast – trending and is composed of a complex deformed assemblage of Palaeo-Proterozoic volcanic and sedimentary rocks intruded by mafic to felsic sills and dykes.

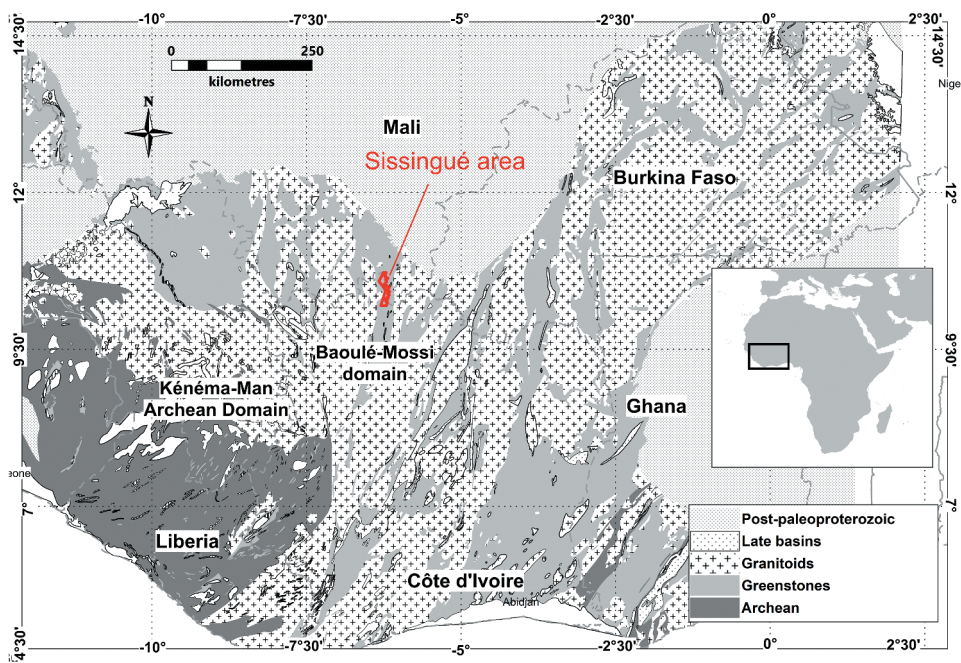


Figure 1. West African craton map showing study area (source: BRGM).

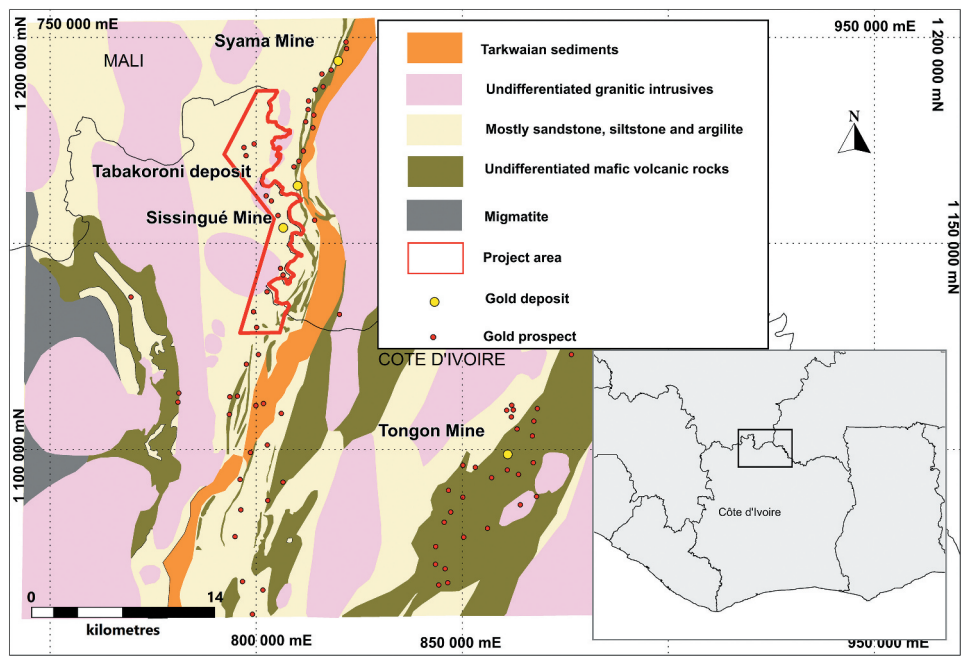


Figure 2. Local geology of Sissingué permit, showing gold occurrences (digitized by Hamed Diaby).

The Syama – Bagoué greenstone belt (Ballo et al. 2016; Traoré et al. 2016) comprises Birimian volcano-sedimentary units adjacent to mafic to felsic intrusions. The volcanic rocks have basaltic, andesitic or dacitic compositions. The sediments include sandstone, greywacke, black shale, cherty siliceous units and minor polygenic conglomerates. The volcano-sedimentary sequence is metamorphosed to low greenschist facies and exhibits NS- to NE-trending tectonic grains oriented parallel to the steeply dipping regional foliation. Major regional faults have developed along with volcanic and sedimentary units and have been interpreted as major sinistral strike-slip shear zones (Feybesse and Milési 1994).

Gold mineralization in the Syama – Bagoué greenstone belt is typically orogenic and hosted by various lithofacies, including sediments and intrusive rocks at the Sissingué deposit and volcanic and sedimentary rocks at the Syama mine (Figure 2). Brittle-ductile shear zones structurally control the vein type and disseminated gold mineralization in most known gold occurrences.

3. Materials and methods

3.1. Data correction

In this study, various data types were used to accurately map the regolith landforms. Landsat-8, ALOS PALSAR DEM, and S1A data were processed and used, as shown in the graph below (Figure 3).

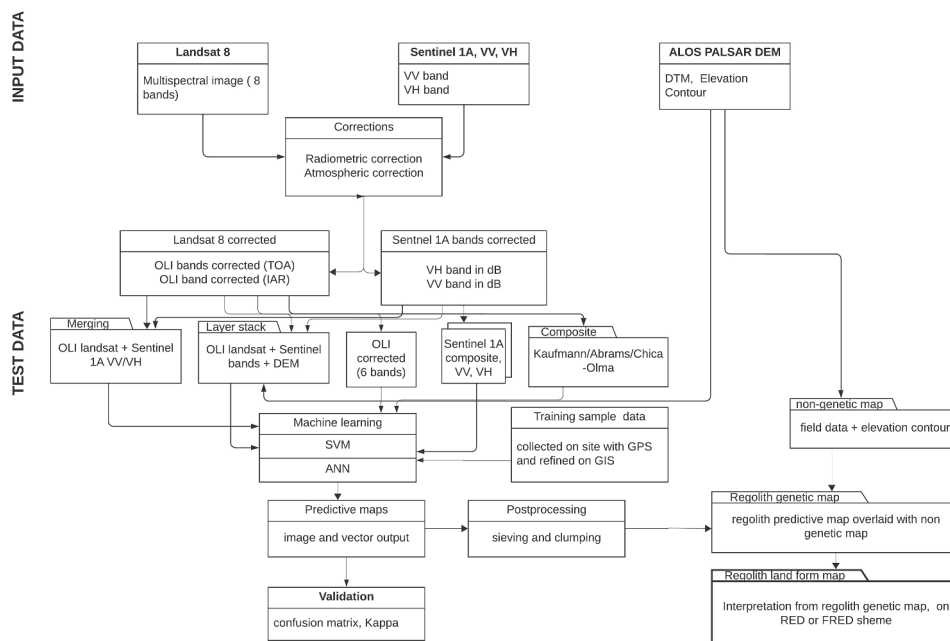


Figure 3. Methodology workflow.

3.1.1. Sentinel-1A data

Sentinel-1A data, dated 11 June 2017 was collected from the sentinel hub of the European Space Agency (ESA) website and corrected geometrically and radiometrically using the ESA Sentinel Application Platform (SNAP 7.0) software. Speckles were removed using the SNAP single product filter operator based on the Lee sigma filter of 7-pixel size and a target window of 3-pixel size matrix with a sigma value of 0.9 as the default value (Veci 2015). Data transmitted horizontally and received vertically representing horizontal – vertical (HV) polarization and data transmitted vertically and received vertically representing vertical – vertical (VV) polarization were downloaded for this project. The scene was clipped with a licence border and masked to generate a subset of the datasets.

3.1.2. Landsat-8 data

A Landsat-8 scene was acquired on 28 November 2017 as an LT1 product, implying that the data were terrain-corrected and registered geometrically and offer radiance data. This scene was made of a single file for each band on the US Geological Survey Earth Explorer web platform. After stacking the bands into a single file, the output file was converted to the band-interleaved-by-line (BIL) format. The resulting file was attributed to the high level of radiometric and reflectance corrections. Reflectance was corrected using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) module based on the MODTRAN radiative transfer model using Equation (1) (Prieto-Amparan et al. 2018). The output reflectance file was corrected using internal average relative (IAR) reflectance. This correction was performed to the image to the average scene spectrum (Merzah and Sabah Jaber 2020; Prieto-Amparan et al. 2018) as follows:

$$\rho_{\text{SUP}} = \frac{1}{a_1} \left(\frac{d^2 \pi L_{\text{TOA}}}{\cos \theta_i E_{\text{TOA}}} - a_0 \right), \quad (1)$$

where d is the distance to the sun, L_{TOA} is the spectral radiance of the satellite, E_{TOA} is the solar spectral radiance on a surface perpendicular to the rays of the sun outside the atmosphere, and θ_i is the solar zenith angle. The coefficients a_0 and a_1 are standard atmospheric parameters (water vapour and visibility in km), which can be automatically obtained in the FLAASH module. In this study, visibility was 35 km, and the atmospheric model used was tropical. The aerosol model used was that of a rural area. All the parameters were based on the date of image acquisition.

3.1.3. ALOS PALSAR DEM

The ALOS PALSAR DEM data were acquired on 28 June 2018. The file was also processed to obtain slope images in degrees and percentages using the ERDAS IMAGINE (2015) software. An elevation contour map from the DEM was generated at an equidistance of 10 m using the Datamine Discover (2019) software.

3.2. Field visits

Field visits were carried out from 11–22 December 2016. The team was equipped with a handheld Global Positioning System (GPS) receiver to record all landform regolith units

using code names. An ALOS PALSAR DEM of a 10 m contour map was created on a scale of 1/1000 to 1/5000 depending on the area to visit on A1 format paper and a Landsat-8 natural colour (bands 4, 3, and 2) map on the same scale. These maps were obtained on a board onsite. The main roads and paths were traversed to reach all unit types depending on accessibility. All landform or regolith unit changes were marked on both papers and recorded using a handheld GPS receiver.

Contouring a few medium size regolith landform units (RLUs), e.g. the laterite hills, area of outcrops and certain alluvial illegal mining sites, was conducted on-site with the GPS.

3.3. Data processing

Different datasets were used to test the two ML models, i.e. the ANN and SVM classifiers in ENVI 5.1 software, on the SGP. Each test was based on the same training sample and ground truth data used to evaluate the model. Landsat-8 data corrected for the IAR reflectance image was used as the input data for the MLA. This was the first input dataset used as Dataset 1 to test the proposed model.

Abrams colour composite is based on the band ratios of Landsat-7 images (Alasta 2011). In this study, we used the ratios of bands 5/7 in red, 3/1 in green, and 4/3 in blue obtained from Landsat-8 images (Diaby 2013; Ramadan, Abdelsalam, and Stem 2001). These ratios are known for delineating lithological features highlighting the presence of iron oxides (Alasta 2011; Bodruddoza Mia and Fujimitsu 2012; Cardoso-Fernandes, Teodoro, and Lima 2019; Pournamdari, Hashim, and Pour 2014; Ramadan, Abdelsalam, and Stem 2001). This was used as Dataset 2 to test the model.

The Chica – Olma band ratios (6/7, 6/4, and 4/2) assigned in RGB were tested in the classification workflow as Dataset 3. These ratios are well-known for highlighting ferrous minerals (Aden 2020; Diaby 2013; Kudamnya, Andongma, and Osumaje 2014). The band ratios of 4/2 and 6/7 are well known for lithological mapping and the detection of hydrothermal alteration zones (Banerjee et al. 2019).

Kaufmann band ratios comprised a composite of 7/5, 5/4, and 6/7 band ratios using Landsat-7 data. For Landsat-8 data, we obtained 7/6, 6/4, and 10/7 band ratios (Bodruddoza Mia and Fujimitsu 2012; Diaby 2013; Kudamnya, Andongma, and Osumaje 2014; Ramadan, Abdelsalam, and Stem 2001). These ratios were used to separate vegetation and altered zones (Kudamnya, Andongma, and Osumaje 2014). This dataset was used as Dataset 4.

The Landsat-8 image was combined with the ALOS PALSAR DEM to better highlight the influence of topography on the regolith and its impact on delineating high glacia and distal sediments. The two dataset bands were stacked together and used as Dataset 5.

Gramm – Schmidt Pan Sharpening tool is an algorithm that improves the multispectral data resolution using a high-resolution panchromatic image (Cánovas-García, Pesántez-Cobos, and Alonso-Sarría 2020; Diaby 2013). In our test, the high-resolution image was the S1A radar image (10 m resolution), and the multispectral image was the Landsat-8 bands (30 m resolution). This process generated a picture with 10 m resolution, with the six bands of Landsat-8 displaying the S1A radar pattern on the image. Different combinations were used, with the S1A band as the panchromatic band. In the first case, the data transmitted vertically and received horizontally representing the vertical – horizontal (VH) band was used as the input for the radar and merged with the OLI bands in Dataset 6. In the second case, VH bands from S1A multiplied by two were used as panchromatic images and merged

with the OLI bands in Dataset 7. The S1A VV bands in another test were used as panchromatic images and merged with Landsat-8 data to obtain Dataset 8. For Dataset 9, data from Landsat-8 were corrected for IAR reflectance, and the S1A bands VV and VH and ALOS PALSAR DEM were stacked into the same file and used as a single dataset to test ML. This was performed with the layer stacking tool within ENVI 5.1, and all data were resampled at 30 m resolution using the nearest neighbour resampling method.

3.4. Training and data validation sample selection

Although field observations were recorded onsite during the project in December 2016, an appropriate and organized sample selection was necessary to represent all categories of the regolith unit. Various studies have reported that an optimal choice of training samples is needed (MacLean and Congalton 2009; Metelka et al. 2018). During the field visits, observations were made regarding regolith types and descriptions, and then recorded in the handheld GPS receiver using point and linear features to contour all unit types with a code specifying each RLU. The GPS data were superimposed onto a Landsat-8 image to check and digitize the contours of the training samples. During digitization, we ensured that the unit boundaries did not overlap with other RLU, based on the colour or pattern of the image of each unit. In total, 4905 training sample pixels were evenly selected from the entire SGP area to cover each section of the project area. The validation samples were evenly selected from the study area and were not selected from the samples used to test the model.

3.5. Classification with ANN and SVM

ENVI (V.5.1, 2013) (Solutions 2013) was used to run the classification process on the data. In this study, two main ML classifiers, i.e. ANN and SVM, were used to classify the RLUs. Nine datasets were used to train the two MLAs. The ANN was tested as a nonlinear function by adding a hidden layer. A logistic function was used by default rather than a hyperbolic function. The training rate was set from 0.1 to 0.2, and the training momentum was set at 0.9000. The iteration started at 1000 with the first test and reached 4000 iterations. The root mean square error was set to 0.1000 for the training data; therefore, various data parameters were used to determine the optimal setting with the finest accuracy score. The input data for the ANN and SVM are the same as those reported in Table 1.

The SVM was used as a nonlinear classifier for training, and the kernel type was used as a radial basis function. The gamma kernel function was set at 0.333 by default for RGB bands, such as the colour ratio, and 0.166 for Landsat-8 bands by default. It is the inverse of the number of bands used in the input image, as recommended in the ENVI 5.1 software, and the penalty parameter was set at 100 to obtain the most optimal results. Increasing the value of the penalty parameter increases the cost of misclassifying points and causes ENVI 5.1 to create a more accurate model that may not generalize well (ENVI 2009). A pyramid level was not used to process the data, because the information size was sufficiently fine and did not require a hierarchical processing level.

Table 1. Data input used for the project and bands.

Dataset	Data input	Bands	Processing
1	OLI	6	Layer stack
2	Abrams composite	3	Colour composite
3	Colour composite, Chica Olma	3	Colour composite
4	Colour composite, Kaufman	3	Colour composite
5	OLI and ALOS PALSAR DEM	7	Layer stack
6	OLI pansharpened with Sentinel 1A VH band	7	Gram – Schmidt Pan Sharpening
7	OLI pansharpened with Sentinel 1A VH band X2	7	Gram – Schmidt Pan Sharpening
8	OLI pansharpened with Sentinel 1A VV band	8	Gram – Schmidt Pan Sharpening
9	OLI, Sentinel 1A band VV and VH, and ALOS PALSAR DEM	9	Layer stack

Abbreviations: OLI, operational land imager; ALOS PALSAR, Advanced Land Observing Satellite Phased Array Type L-Band Synthetic Aperture Radar; DEM, digital elevation model; VH, vertical – horizontal; VV, vertical – vertical.

3.6. Accuracy assessment and data validation

The most optimal prediction results obtained with the MLAs used for classifying images were evaluated using a confusion matrix. The confusion matrix or error matrix compares the ground truth data with the classification results and is calculated using an input validation sample (ground truth samples) and a classification image (Congalton 1991). Based on a recent study, error matrices were used in more than 74% of the studies over the last 10 years (Morales-Barquero et al. 2019). The OA was used in 80% of the published documents from 1997 to 2017, and the kappa coefficient was used in 50.4% of the cases published between 2012 and 2017 (Morales-Barquero et al. 2019). In the SGP, ground-truth data collected during field missions were used to assess the prediction images. The second set of data collected onsite with GPS was superimposed on the Landsat-8 image and contoured to create a validation dataset while visiting the site.

Kappa is a coefficient generated after the error matrix, and it assesses the extent to which the classification is compared with random classification. In addition, OA, producer accuracy (PA), and user accuracy (UA) are used to assess the data. The OA was obtained by dividing the sum of the correctly classified pixels by the total number of pixels. The UA was calculated by dividing the number of pixels accurately mapped for a class by the total number of pixels classified for this class. This is related to the reliability of the map and provides the probability that a pixel mapped to a group belongs to that group in the field (Cardoso-Fernandes, Teodoro, and Lima 2019). The PA was obtained by dividing the number of correct pixels in a particular class by the total number of pixels in the test set. This indicates the probability that the training samples were correctly mapped (Cardoso-Fernandes, Teodoro, and Lima 2019).

3.7. Data processing post-classification

After classification, the resulting file is often peppered and contains isolated pixels scattered throughout the study area. To remove these pixels, we used sieving with a tool with the same name as that in ENVI 5.1. We opted to remove pixels when they were one or two and, therefore, not part of a group. Once removed, they leave a gap and thus become unclassified.

We used a clamp-class tool to remove these unclassified pixels and close the gaps. This tool uses the classes surrounding the unclassified pixels to close the gap, providing more coherence to the image. The selected classes were grouped

Table 2. Class combinations and references based on the Relict–Erosional–Depositional (RED) scheme.

Class name	Combined as	RED name
High laterite	High laterite	Relict regime
Low laterite	Low laterite	Depositional regime
Outcrop	Area of outcrop	Erosional regime
Colluvium	Proximal colluvium	Erosional regime
Distal alluvium	Distal sediment	Depositional regime
Clear water	Distal sediment	Depositional regime
Dirty water	Distal sediment	Depositional regime
Building area	Building area	Replaced with surrounding RED unit, and based on a field visit

by performing dilation and erosion operations on the classification image using a kernel of the size specified in the parameter dialogue box (Research Systems 2001). In this study, seeds from two lines and two columns were used. A larger kernel erases certain details from the image.

The final step in processing the classification results was the combination of classes. Classes were combined into categories based on membership or origin (Table 2). Water, water with particles, and distal alluvial sediments were combined into one class called distal sediments. In our context, laterite hills (referred to as laterite) and low laterite areas were considered separately. However, they can also be merged. The outcrop areas were preserved. Colluvial areas were also retained as such. Proximal alluvium was retained as such. All units were merged into a GIS using the RED scheme (Table 2). The final file was converted to a vector file using the image-to-vector menu and saved in '.shp' format.

3.8. Developing regolith landform map using regolith scheme

The regolith landform maps were developed to comply with the schemes described above. This can be the FRED or RED model (Arhin et al. 2015). Data collected during the field visit allowed us to build a non-genetic map (Figure 4). All visited units were recorded, as explained in the methodology section. The digitized boundary from the factual data was superimposed onto the regolith map predicted by ANN or SVM. The predicted map was validated using factual and DEM data and edited using the GIS software. As the MLA classifies all unit boundaries, this work consists of choosing each unit group based on the scheme used (Table 2).

Once the regolith landform map was developed following the RED model, it was overlaid on the ALOS PALSAR DEM in the GIS software Datamine Discover (v.2019) and displayed in 3D. The topography was then exaggerated by multiplying the elevation 10 times its initial value, which allowed the landform features to be displayed on the regolith map.

4. Results

4.1. Regolith characteristics and the formation process

4.1.1. Area of outcrops

Outcrop areas are rare on the project site. The occurrences of outcrop are mostly on illegal mining sites created by the excavation of saprolite and hard rock. Outcrops mapped

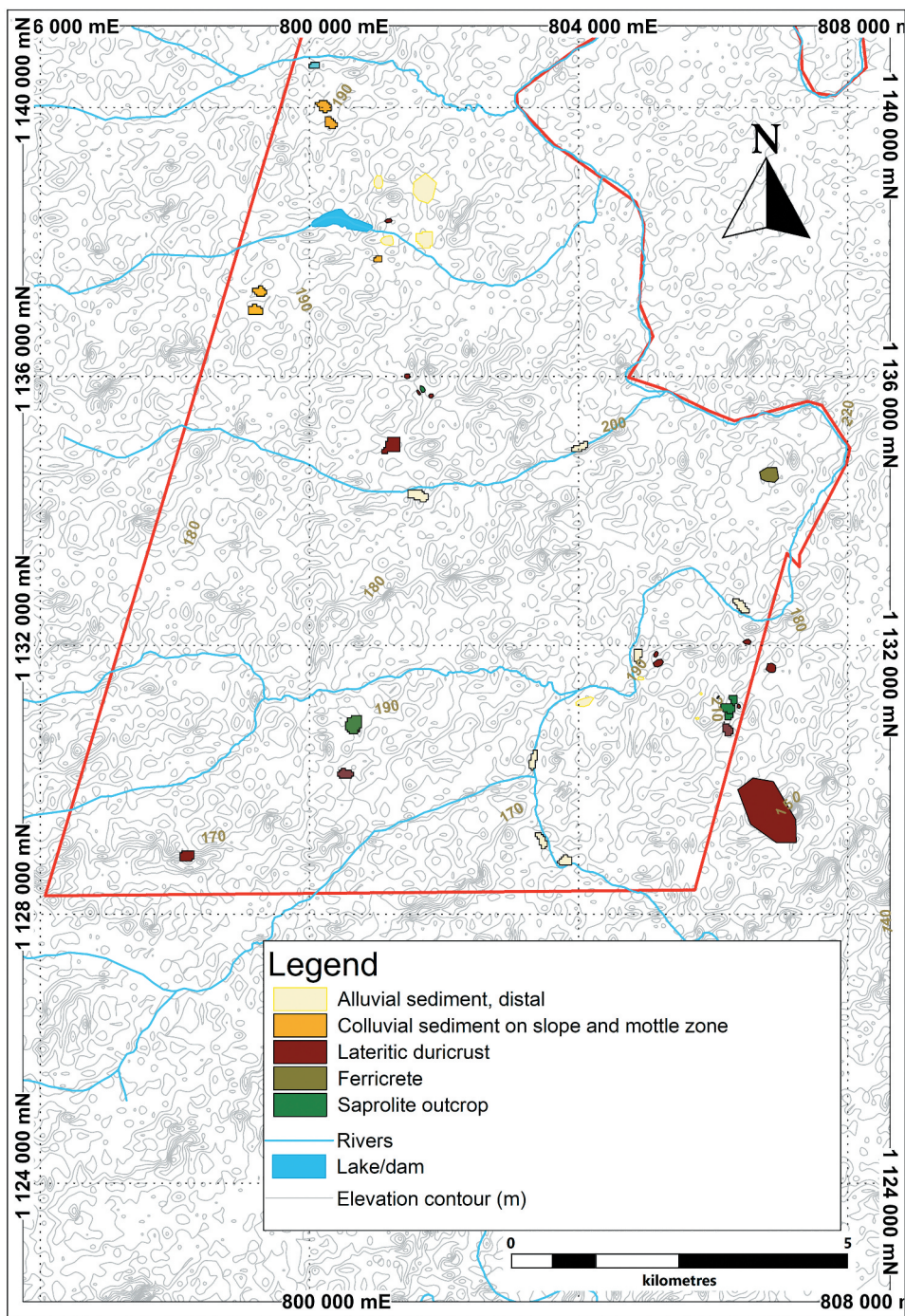


Figure 4. Non-genetic regolith map of the southern side of the Sissingué Gold Project (SGP), created from ground truth data overlaid on elevation contour.

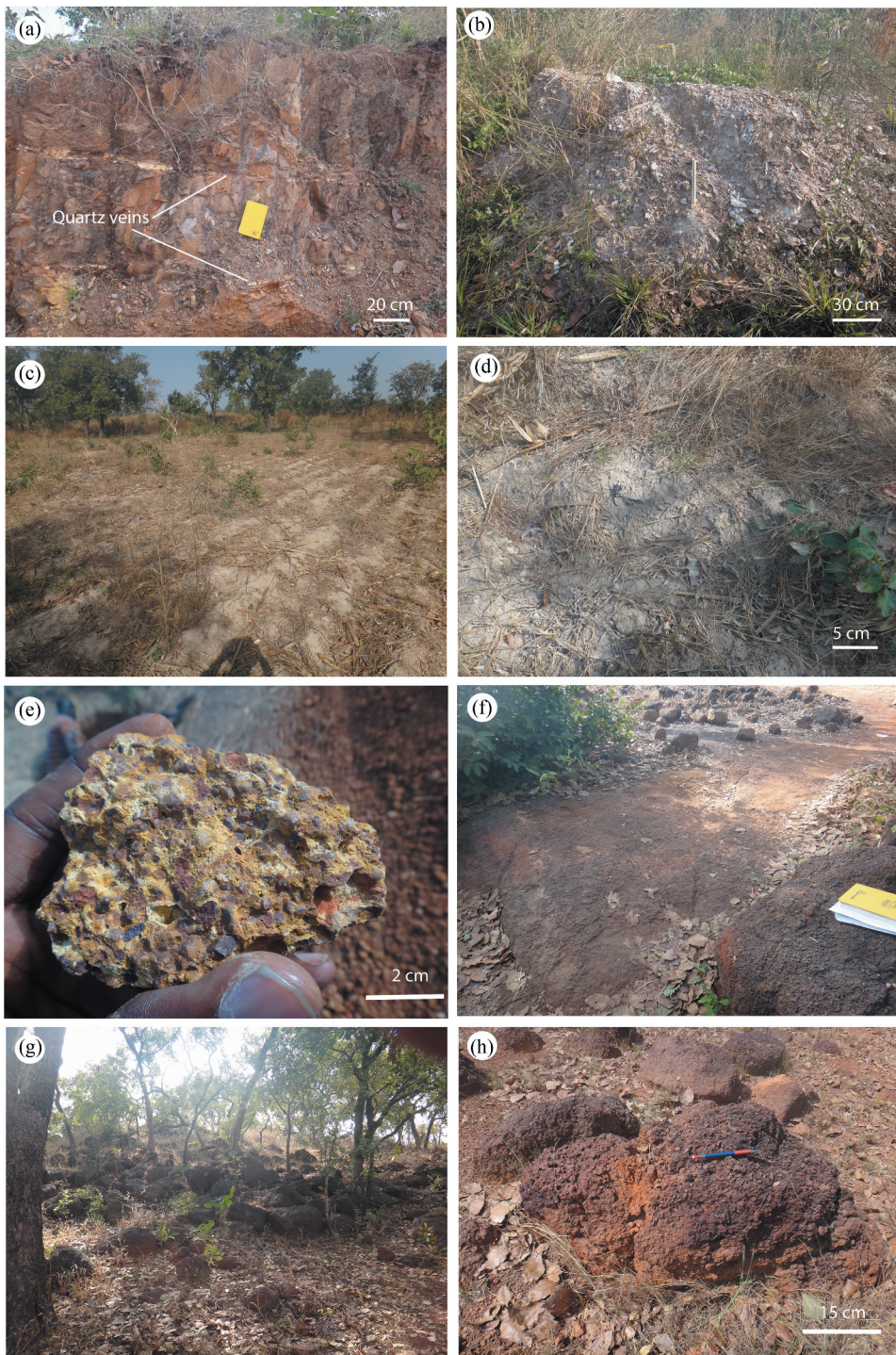


Figure 5. (a) Saprolite exposed by illegal mining activity showing quartz veins and geological structure fabric. (b) Rejects from pits from artisanal mining exposing underground rock fragments. (c) Plantation showing alluvium or transported regolith. (d) Alluvium is composed mostly of fine sand and clay. (e) Laterite fragment. (f) Indurated laterite cap metre size on top of the hill, preventing vegetation growth. (g) Colluvium block on hill slope coming from the hilltop, vegetation growing on a slope, between blocks. (h) Colluvium block of laterite centimetric size on the hill flank.

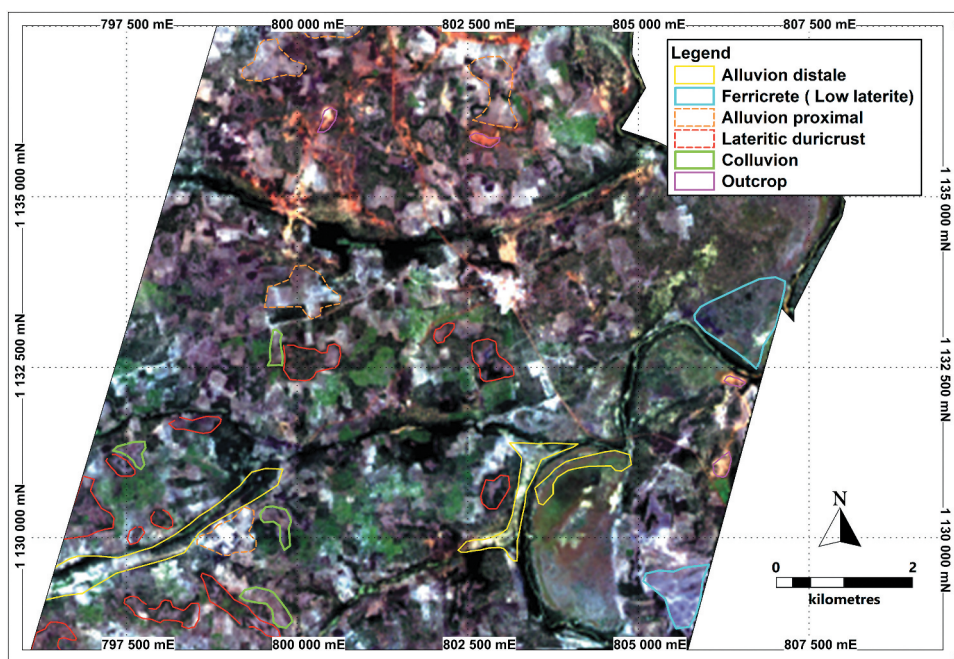


Figure 6. Landsat-8 image in natural colour, showcasing the regolith units identified on site.

hitherto on this site are generally sediments, specifically chert and siltstone (Figure 5(a)), and a few are granitoid. Quartz vein is bearing the mineralization in the main illegal mining site (Figure 5(b)). These illegal mining sites are noticeable on the Landsat-8 OLI colour composite false natural colour image with a bright brown colour (Figure 6).

4.1.2. Distal alluvium

Alluvial sediments are mostly present in low-elevation areas, readily noticeable in the valley, with the ALOS PALSAR DEM data showing a depression area. On gentle slopes, alluvium can contain gravel, pisolites and the fragments of lithic material. Colour varies from yellowish to whitish, depending on the terrain position and the original material (Figure 5(c,d)).

Alluvial areas appear bright yellow or light orange on the high-resolution false colours of Landsat 8 OLI images, whereas they appear dark green along the stream, represented by thick vegetation along the riverbed (Figure 6). This depositional unit, mostly used for farming by local people, can be located with ease. Farming areas are regular in shape. Two categories of river bed alluvium exist, i.e. the recent alluvium, called distal sediment, which mostly occurred during the Quaternary, and the old sediment, exposed after a secondary erosion and created from a paleochannel mostly during mid-to-late Eocene. These sediments were created via lateral displacement from the upper hill caused by water and wind action.

4.1.3. Low-lying laterite and ferricrete

In natural colour images, ferricretes appear dark purple. They are common in the eastern part of the project, near the Bagoué River, and cover large areas that can be identified in satellite images (Figure 6). This unit is darker compared to the residual laterite. The colour is generally red or dark brown. Lithic and rounded quartz fragments are present in the laterite matrix containing voids. The ferricrete appears vermicular because of pisolite containing haematite and goethite nodules (Figure 5(e,f)). It is formed by the accumulation of ferruginous sediments and distal colluvium, composed of fragments of other laterites cemented by water and extremely fine sediments. Ferricrete also appears on high altitudes as concave hills with no vegetation and few small termite mounds (nearly half a metre long). Its presence at high altitudes is due to inversion of the relief.

4.1.4. Colluvium

The colluvium is generally composed of blocks of lateritic duricrust, mostly proximal, i.e. not far from their source (Figure 5(g)). Lateritic gravels surround them, and the block sizes in this lateritic unit vary from 50 cm to 1 m (Figure 5(h)). A few trees have grown in this unconsolidated unit, occasionally masking the colluvium on satellite images. On Landsat-8 imagery, they may appear either green because of the vegetation or orange because of a mix of laterite blocks and laterite gravel surrounding them (Figure 6). The plasmic zone and saprolite lie beneath this unit and appear in places.

4.1.5. Laterite duricrust

Laterite duricrust commonly appears purple in natural OLI images and is surrounded by green vegetation. This unit is readily identifiable in natural colour images (Figure 6). On-site, they harbour little vegetation. At high altitudes, the shape of the hill is convex, with its slope generally covered by colluvium or saprolite.

The duricrust is reddish to brown coloured and does not contain voids or round quartz fragments. It contains haematite nodules and goethite and is formed by complete weathering of the underlying bedrock. On its surface lag is composed of the pisolites and lithorelics of the same composition from the same bedrock.

4.1.6. Proximal sediments or soil

The proximal sediments are generally covered by vegetation. In places, they develop on or near the saprolite. They are composed of fine sand and lateritic gravel, and their colour varies from orange to light yellow. In satellite images, they are light yellow when ploughed for planting crops (Figure 6). They are not present in the highlands nor in the lowlands but in a transition zone between the plains and high plateaux. This unit is rare, and therefore, difficult to locate on this site.

4.2. Regolith predictive map results

Nine regolith prediction maps were produced using SVM (Figure 7), whereas the ANN produced eight predictions (Figure 8). The predictive map with Dataset 9 could not generate results because of a persistent bug due to the size of the file used. The outputs, obtained from Datasets 1–9, were used to delineate the following regolith units: lateritic duricrust, low laterite areas, colluvium, light-coloured alluvium, lake sediment, clear water,

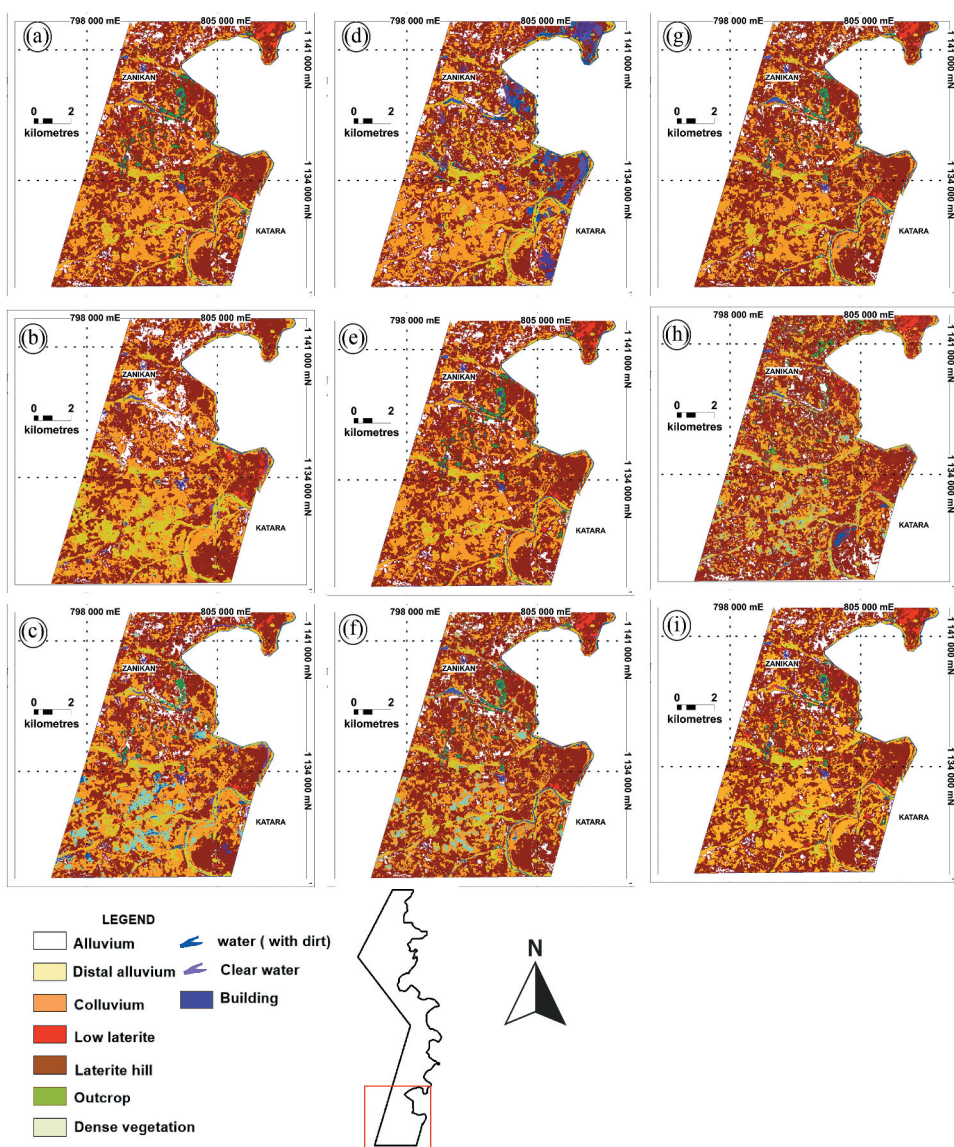


Figure 7. Results obtained from support vector machine (SVM) classification using various inputs based on the Datasets (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g) 7, (h) 8 and (i) 9.

polluted river water and outcrops. Urban and densely vegetated areas were classified using these algorithms.

The test results obtained with Dataset 1 were fine, and the classes appeared to be correctly identified with the ANN and SVM based on the repartition of units in terms of field visits and landforms (Figure 7(a) and 8(a)). The second test with Dataset 2 using the Abrams ratio was also consistent and identified all the regolith units but displayed more water classes even on the lateritic duricrust or in outcrop areas (Figure 7(b) and 8(b)).

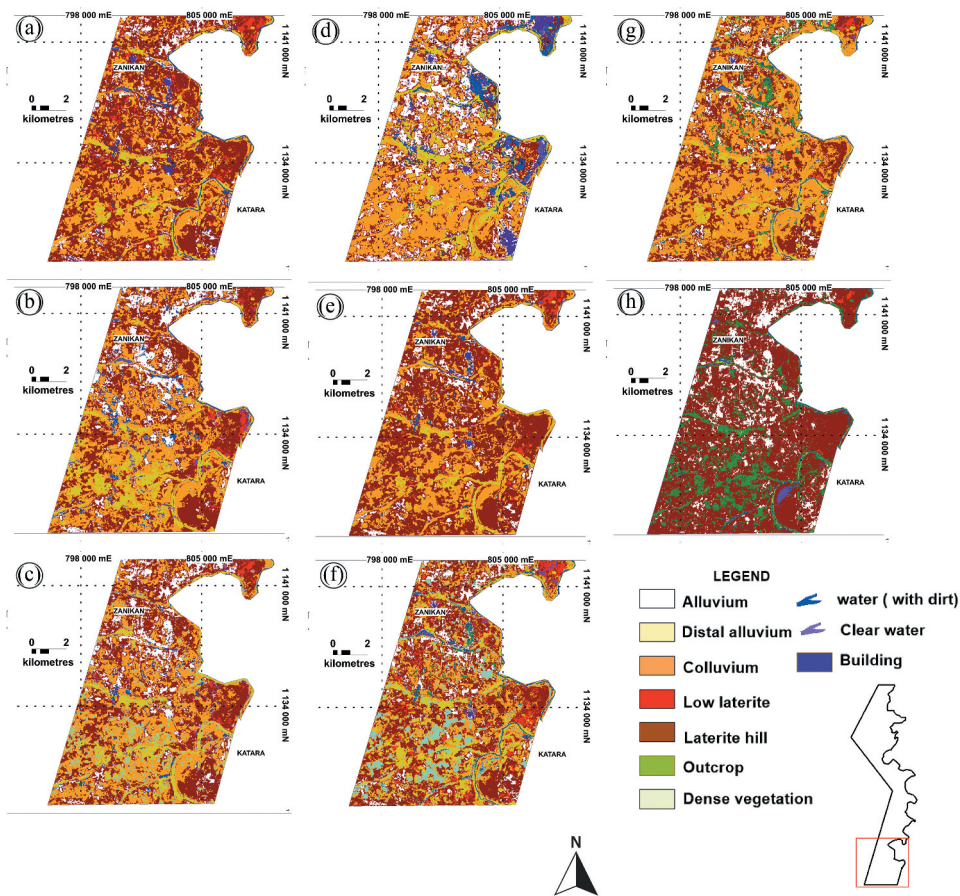


Figure 8. Results obtained from ANN classification using various inputs based on the Datasets (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g) 7 and (h) 8.

The test with Dataset 3 identified dense vegetation in addition to the other classes. The results for both ANN and SVM delineated the other units without major confusion, although water appeared in some places far from the rivers (Figure 7(c) and 8(c)).

The test with Dataset 4 did not correctly identify the outcrops and water classes, and the areas of the buildings appeared all over the predictive map with both MLAs (Figure 7(d) and 8(d)).

The test with Dataset 5 generated predictive maps with an ANN showing all classes but rarely outcrops and rivers. The SVM showed all classes without difficulty for outcrops or rivers (Figure 7(e) and 8(e)).

Using OLI and S1A data for Dataset 6, we predicted the regolith, vegetation, and building classes. The ANN seemed to identify relatively more low laterite areas compared with that identified by the SVM in the southern part of the work area (Figure 7(f) and 8(f)). The predictions with Dataset 7 identified all classes using both methods. The sediments were along the rivers, and the laterite plateau was at the top of the hill (Figure 7(g) and 8(g)). The test with Dataset 8 also identified all regolith classes, but the ANN generated mostly laterite

and outcrops (Figure 7(h) and 8(h)). However, the test with Dataset 9 was only conducted by the SVM with all units of the regolith identified.

4.3. Regolith landform map based on the RED scheme

The three RLUs that made up the RED model were perfectly mapped using the predictive regolith map created by the MLA and overlaid by a non-genetic regolith map. This section presents the results of the regolith landform map development.

The district landscape is mainly characterized by a low-lying, monotonous plateau (320–330 m above sea level) falling in the widespread depositional regolith units mapped as alluvium and distal colluvium widespread in the lower geomorphological unit relative to the local catena. At the surface, alluvial areas are often well-vegetated with grey-reddish silty soil. A considerable portion of this unit comprises fertile floodplain soils and crops. Lateritic caps transported and recemented in low-lying areas were also part of this depositional regime (D) (Figure 9). They mostly border the Bagoué riverbed in the eastern part of the study area.

The relict regime (R) is represented by small hills of *in situ* lateritic plateaus, with lateritic residuum representing high geomorphological units at elevations ranging from 350–370 m (Figure 10). The lateritic plateau consists of residual duricrust, which is commonly preserved at ridges and high pediments. Residual sediments and sediments of proximal origin are also part of this regolith domain.

Areas with bedrock outcrops and sub-outcrops are rare in the field and represent patchy erosional (E) surfaces. These surfaces consist of locations where pre-existing lateritic regoliths have been partially eroded. They also include areas where mottle-clay saprolite, saprolite, and plasma zones appear. Thin soil layers on top of eroded surfaces are standard, and lateritic lags and transported sediments frequently overlie these regions.

At higher elevations, the regolith landform is often dominated by transported ferricretes, which define the central geomorphological units (330–350 m). Typically, the transport of materials is induced by iron oxides. This unit forms elongated plateaus and terraces with little to no vegetation on satellite images. This part is also considered under the depositional regime and is mostly encountered on the eastern side, close to the Bagoué River. The depositional regime of the project represents 36.39%, the relict domain is estimated at 24.91%, and erosion is 38.7%. The erosional zone is greatly affected by the excavation work carried out by gold miners.

5. Classification accuracy

5.1. Classification accuracy of SVM

5.1.1. Most optimal SVM classification accuracy

After testing various types of input data and switching MLAs, our most optimal results were obtained from using an SVM classifier and Dataset 7. In this dataset, Landsat-8 data were pan-sharpened using the S1A VH band (Figure 11). This predictive regolith map exhibited an OA of 86.54% with a kappa coefficient of 0.8445 (Table 3).

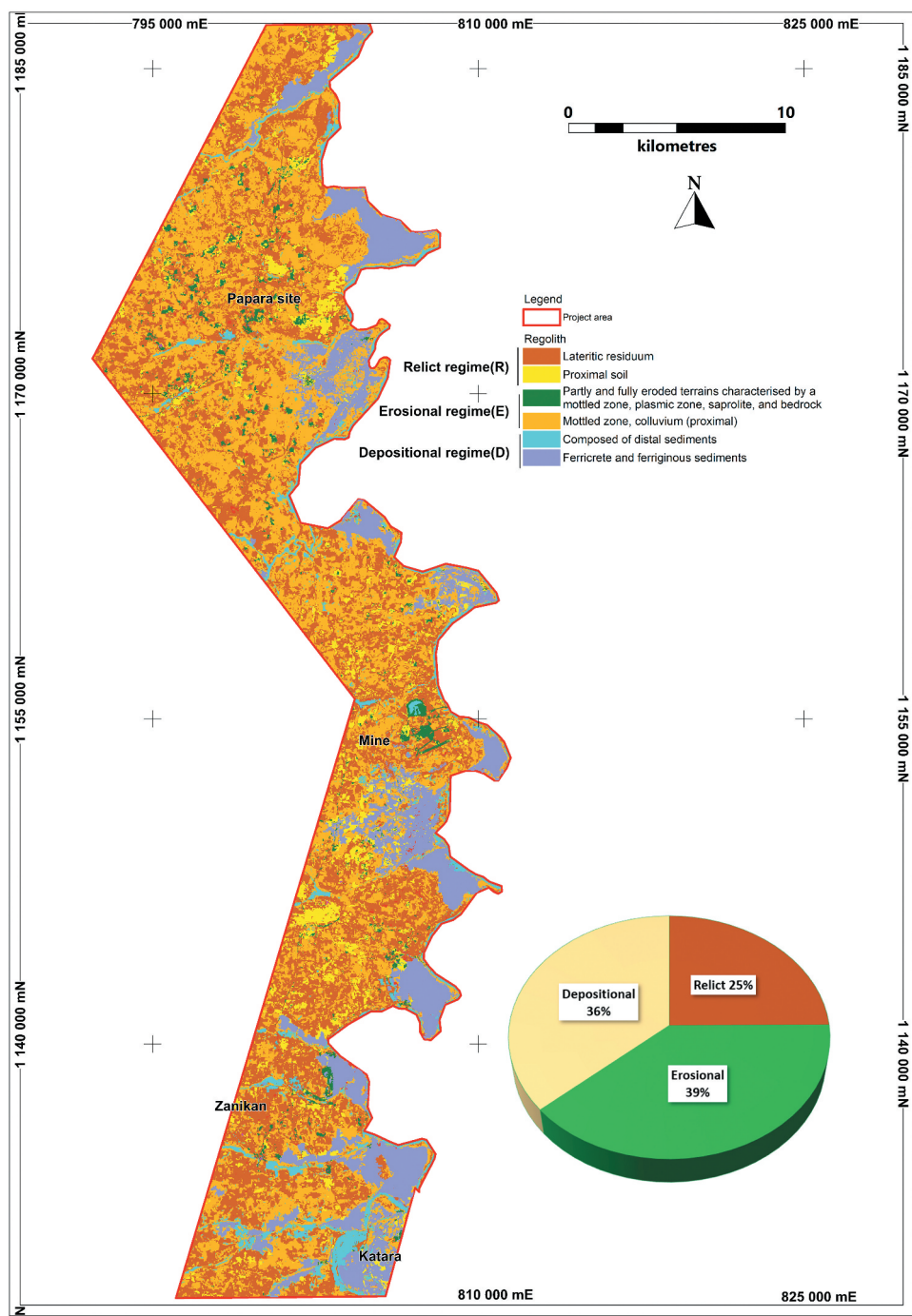


Figure 9. Map of the regolith landform units in the southern part of the Sissingué permit developed using a non-genetic map following the Relict–Erosional–Depositional (RED) scheme.

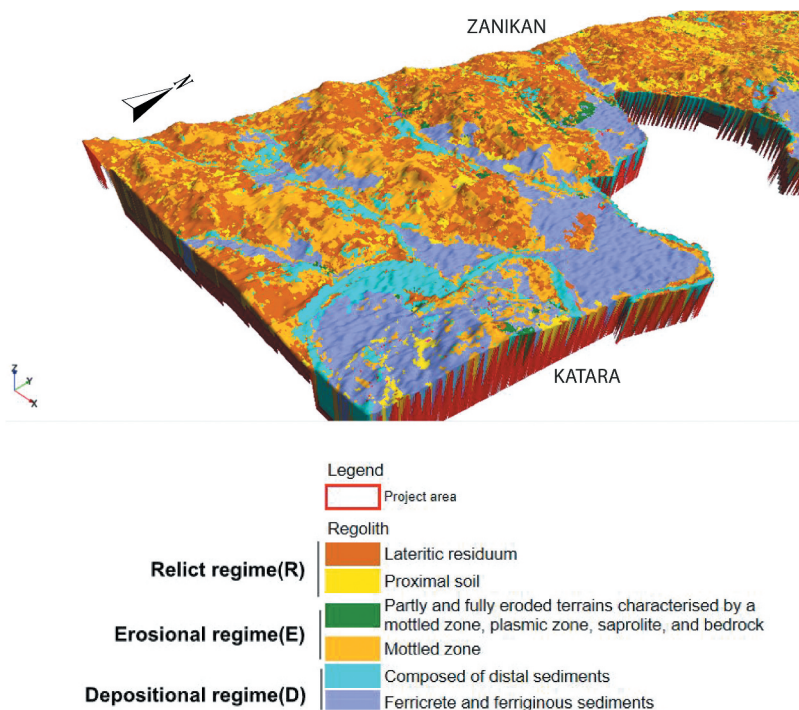


Figure 10. Regolith landform map in 3D, showing all units based on the Relict–Erosional–Depositional (RED) model, with relict areas in elevated areas, whereas the depositional area is in a valley or plain site but with few areas elevated while being transported (elevation exaggerated 10 times).

The most optimal match for this dataset was water containing dirt, which was readily discernible because it contained many fine suspended particles. Gold panning or illegal mining workers constantly use water to wash bulk minerals from the pit to sort the gold by gravity process and this water ends up in the Bagoué River. The PA was 100%, whereas UA reached 70% for this class.

The clear water class was the second-most optimal match, with a PA of 99.83% and UA of 97.47%. Clear water from the river that was not contaminated by the gold panning sites was well differentiated from the other classes, even from the additional water containing dirt from the gold panning site. Both clear and contaminated water types were observed within the depositional area.

Outcrop or sub-outcrop areas exhibited a UA of 71.85%, whereas PA was 92.93%. The outcrop unit was well-identified and generally exposed to illegal gold mining activities. There were also a few disturbed areas in the same place, such as alluvium and laterite fragments, where gold washers collected and washed the lithic material. This explains why a few outcrop pixels were classified as alluvial or laterite (Figure 5(a,b)).

The alluvial unit representing the depositional unit was well mapped, with a PA of 94.22% and a UA of 82.99%. The proximal alluvium or sediment was discernible; however, only the building area was similar because of the alluvium present on the road. This type of regolith is widely used in agriculture.

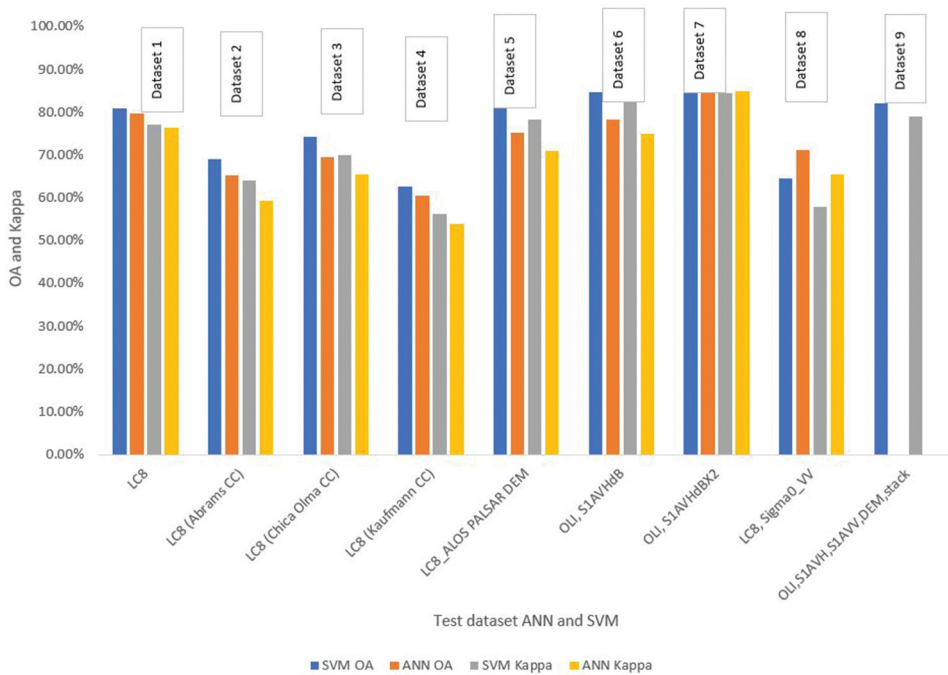


Figure 11. Overview of classification results using six bands of LC8 (Dataset 1), colour composite band ratios obtained from OLI and delineated by Abrams (Dataset 2) and Kaufmann (Dataset 4), Chica–Olma ratios (Dataset 3), LC8 bands added with ALOS PALSAR DEM data (Dataset 5), LC8 bands merged with the S1A Sigma0 VV band (Dataset 8), S1A VH (dB) data (Dataset 7) and S1A VH 0 (dB) ×2 bands (Dataset 9) and pansharpened with S1A Sigma0 VH X2 band (Dataset 6) used as input for processing with SVM and ANN classifier. Abbreviations: LC8, Landsat-8; OLI, operational land imager; ALOS PALSAR, Advanced land Observing satellite Phased Array type L-Band Synthetic Aperture radar; DEM, digital elevation model; S1A, Sentinel-1A; VV, vertical – vertical; VH, vertical – horizontal; SVM, support vector machine; ANN, artificial neural network.

Table 3. Confusion matrix for support vector machine (SVM) classifier using Dataset 7.

Class	Validation										
	OUT	ALL	WA	ALR	LAT	CLWA	BUILD	LLAT	COL	PA %	UA %
OUT	171	31	0	0	34	0	0	0	2	92.93	71.85
ALL	1	717	0	0	3	0	143	0	0	94.22	82.99
WA	0	0	105	45	0	0	0	0	0	100	70
ALR	0	0	0	681	0	0	0	0	0	82.25	100
LAT	12	13	0	1	386	0	18	63	216	79.42	54.44
CLWA	0	0	0	6	0	579	0	9	0	99.83	97.47
BUILD	0	0	0	0	0	1	614	0	0	79.23	99.84
LLAT	0	0	0	0	0	0	0	628	0	89.71	100
COL	0	0	0	95	63	0	0	0	982	81.83	86.14
Total	184	761	105	828	486	580	775	700	1200		
		OA:	86.55%			Kappa:	0.8445				

Abbreviations: OUT, outcrop unit; ALL, alluvium unit; WA, water with dirt material; ALR, recent alluvium (distal); LAT, laterite; CLWA, clear water; BUILD, building and housing; LLAT, low laterite; COL, colluvium; PA, producer accuracy; UA, user accuracy; OA, overall accuracy; Kappa, Kappa coefficient.

Laterite units were mapped into two types, i.e. low laterite areas, literally flat regarding the topography, and mostly transported. This class exhibited a PA of 89.71%, whereas UA reached 100%. This suggests that relatively more pixels were identified as low laterite areas compared with that in the test data. Most misclassified pixels were considered laterite hills (residual or high laterite) because they were similar to low laterite areas and looked identical on satellite imagery.

The laterite hill unit was well-mapped, with a PA of 79.42% and a UA of 54.44%. This unit is similar to the colluvium unit of the area and is mainly composed of laterite fragments and a mottled zone. These ferruginised fragments generally slide down the laterite hill. Low-grade laterite also exhibited the same spectral signature as that in the Landsat-8 images. Using radar bands helped to discriminate this feature in a few places. Building and housing units were mapped to a few SGP areas. This category exhibited a PA of 79.23% and a UA of 99.48%.

Recent alluvial deposits or distal sediments, generally accumulated along the riverside, were classified with a PA of 82.25%, whereas UA reached 100%. This area was typically covered by thick vegetation that grows only on alluvial deposits. A few similarities to the area covered by colluvium on the hill flank with vegetation were observed.

Mapping a colluvial unit with an area mostly covered by regolith material was challenging because it had the same laterite cap. Materials, broken into fragments with lithic angular pieces, were often present on the hill flanks. The unit exhibited a PA of 81.83% and a UA of 86.14%. This is because of their similarities to laterite units. They were often composed of the same material, and only the size and localization were different (Figure 5(c,d)). This class prediction also exhibited a few similarities with the vegetation class because laterite blocks are often covered by scattered vegetation in the colluvial area.

5.1.2. Comparing the accuracy of SVM

The other results of SVM tests using the datasets produced impressive predictions (Figure 11). Using SVM with Landsat-8 data (Dataset 1) provided an OA of 80.98% with a kappa coefficient of 0.77. The generated regolith map was close to reality; however, the laterites were poorly classified, with a UA of 46.34% and a PA of 77.03%. Using colour composite ratio data, such as those of Chica – Olma, Kaufmann, and Adams, average classifications were produced. However, Dataset 7 (Chica – Olma colour composite ratio) was better, with an AO of 74.31% and a kappa of 0.7. Combining DEM data with Landsat-8 data (Dataset 5) yielded a slightly improved OA of 81.22% and a kappa coefficient of 0.78. Using the VV polarization S1A Sigma0 band value coupled with Landsat-8 data (Dataset 8) produced a classification with an OA of 64.43% and a kappa coefficient of 0.58. Thus, there was no improvement compared with that using Landsat-8 data alone or with PALSAR DEM data.

The S1A VH band merged with Landsat-8 images (Dataset 7) provided the most optimal SVM prediction in this study, with an OA of 84.69% and a kappa coefficient of 0.83. Like Dataset 7, the combination of the ALOS PALSAR DEM bands, Landsat-8 data, and S1A VH and VV bands provided reasonable predictions.

Table 4. Artificial neural network (ANN)-based confusion matrix showing the most optimal predicted regolith classes with Dataset 7.

Class	OUT	ALL	WA	Validation			BUILD	LLAT	COL	PA %	UA %
				ALR	LAT	CLWA					
OUT	150	35	9	0	8	0	0	0	0	81.52	74.26
ALL	22	693	0	0	30	0	150	0	0	91.06	77.43
WA	0	0	96	45	0	1	0	0	0	91.43	67.61
ALR	0	0	0	732	0	0	0	0	8	88.41	98.92
LAT	12	33	0	0	344	0	14	82	121	70.78	56.77
CLWA	0	0	0	0	3	578	4	0	0	99.66	98.8
BUILD	0	0	0	0	0	1	607	0	0	78.32	99.84
LLAT	0	0	0	0	3	0	0	618	0	88.29	99.52
COL	0	0	0	51	98	0	0	0	1071	89.25	87.79
Total	184	761	105	828	486	580	775	700	1200		
		OA:		87.01%		Kappa:		0.849			

Abbreviations: OUT, outcrop unit; ALL, alluvium unit; WA, water with dirt material; ALR, recent alluvium (distal); LAT, laterite; CLWA, clear water; BUILD, building and housing; LLAT, low laterite; COL, colluvium; PA, producer accuracy; UA, user accuracy; OA, overall accuracy; Kappa, Kappa coefficient.

5.2. Classification accuracy of ANN

5.2.1. Most optimal predictive accuracy of ANN

The ANN demonstrated an excellent capability for regolith mapping with an OA of 87.01%, whereas the kappa coefficient was 0.849 in this project area with Dataset 7. The results of the ANN were improved by gradually increasing the number of iterations to 4000 iterations. The most optimal input was for the S1A VH band in decibels merged with the Landsat-8 bands (Dataset 7; [Figure 11](#)), which yielded the most promising OA and kappa coefficients ([Table 4](#)).

The most optimal classification unit using the cited input and ANN classifier was the clear-water unit. The PA was 99.66%, and the UA was 98.8%. The area covered by water also represented the distal sediment area. This was almost the same value as that of the SVM classifier for the PA but with a higher UA for the ANN.

Water containing dirt transported from alluvial mining washing sites also exhibited a sound PA of 91.43%, whereas UA was 67.61%. A few pixels from this water were classified as clear water because of similarities in a few areas. All water units were mapped because they were generally located in depositional areas representing distal sediments.

The alluvium representing the proximal sediment unit was well-mapped by the ANN classifier with a PA of 91.6%, whereas UA reached 77.43%. Recent alluvial or distal sediments beside rivers and waterways were also well classified, with a PA of 88.41% and a UA of 98.92%. This unit contained rejected materials from illegal mining and had a few misclassified pixels in the border area between water and vegetation on the riverbed, and a few areas covered by agriculture, suggesting that this area was advantageous to crop farming.

The low laterite area class was also well-mapped, with a PA of 88.29% and a UA of 99.52%. A few pixels were mapped as high laterite because they contained residual laterite fragments cemented with clay and other lithic components. The urban area was also mapped with a PA of 78.32% and a UA of 99.84%. Laterite hills were mapped with a PA of 70.78%, whereas UA reached 56.77%. This unit had an average

score with an ANN classifier. Most misclassified pixels originated from the low laterite area and colluvium classes because they have many similarities. The colluvium unit, mainly composed of laterite and lithic fragments, exhibited a PA of 89.25% and a UA of 87.79%. The laterite units were more similar because their compositions and spectral signatures were similar. The ANN algorithm exhibited mapped outcrop units but with a fine score, and the PA reached 81.52%, whereas the UA was 74.26% (Table 4).

5.2.2. Comparing ANN results

In general, the different performances obtained using ANN followed those obtained using SVM (Figure 11). The test with Landsat-8 data generated useful predictions, with an AO of 79.69% and a kappa coefficient of 0.76. Accuracy did not increase with the use of colour composite images alone, with ratios such as those of Chica – Olma (OA: 69.42%, kappa coefficient: 0.65), Abrams (OA: 65.14%, kappa coefficient: 0.59), and Kaufmann (OA: 60.59%, kappa coefficient: 0.54) making average predictions. The predictions were significantly improved with the combined use of Landsat-8 data and the ALOS PALSAR DEM data (Dataset 5), with an OA of 75.10% and a kappa coefficient of 0.71 (6% and 5% higher, respectively, than test outcomes of Dataset 3, but still lower than that obtained with Dataset 1). The addition of the S1A VV band to the Landsat-8 data (Dataset 8) exerted a positive impact compared with that on the SVM results, with an OA of 71.25% and a kappa coefficient of 0.65. Dataset 6, composed of Landsat-8 data merged with the S1A VH band, produced a reasonable prediction with an OA of 78.35% and a kappa coefficient of 0.75. Dataset 7, as explained above, yielded the finest accuracy in this study, with an OA of 87.01% and a kappa coefficient of 0.85. The last test with Landsat-8 data, ALOS PALSAR DEM data, and the combined VV and VH bands could not be run with the ANN, probably because of the file size and dimensionality of the data, causing the test to crash several times.

6. Prediction validation

Validation data collected in parallel during the field phase confirmed the excellent quality of the predictions from Dataset 7, which was used with the MLA. The OA of the validation data was 78.60% and its kappa coefficient was 0.74% for the ANN on Dataset 7. All units exhibited an average PA of 70%. However, the UA for the laterite class was 55.87%

Table 5. Validation matrix for support vector machine (SVM)-based predictive regolith map.

Class	Validation										
	OUT	LLAT	BUILD	LAT	COL	ALL	CLWA	ALR	WA	PA %	UA %
OUT	598	0	1	1	0	149	0	0	0	88.59	79.84
LLAT	0	425	0	25	0	0	0	5	0	73.28	93.41
BUILD	54	0	359	0	0	0	0	0	0	98.63	86.92
LAT	2	93	0	1595	582	569	0	1	13	94.71	55.87
COL	0	8	0	51	1351	18	0	12	6	69.07	93.43
ALL	21	0	4	12	22	1238	0	0	0	62.72	95.45
CLWA	0	54	0	0	0	0	117	0	4	100	66.86
ALR	0	0	0	0	1	0	0	378	16	95.45	95.7
WA	0	0	0	0	0	0	0	0	270	87.38	100
Total	675	580	364	1684	1956	1974	117	396	309		
				OA :	78.60%	Kappa:	0.74				

Abbreviations: OUT, outcrop unit; LLAT, low laterite; BUILD, building and housing; LAT, laterite; COL, colluvium; ALL, alluvium unit; CLWA, clear water; ALR, recent alluvium (distal); WA, water with dirt material; PA, producer accuracy; UA, user accuracy; OA, overall accuracy; Kappa, Kappa coefficient.

Table 6. Validation matrix for artificial neural network (ANN)-based predictive regolith map.

Class	OUT	BUILD	COL	LAT	Validation		CLWA	ALR	WA	PA %	UA %
					LLAT	ALL					
OUT	473	0	0	0	0	492	0	0	0	70.07	49.02
BUILD	77	356	0	0	0	0	8	0	0	97.8	80.73
COL	0	0	1479	130	14	77	0	9	4	75.61	86.34
LAT	125	0	396	1437	87	427	0	0	0	85.33	58.13
LLAT	0	0	0	107	457	0	0	6	0	78.79	80.18
ALL	0	8	22	10	0	961	0	0	0	48.68	96
CLWA	0	0	0	0	22	0	109	0	0	93.16	83.21
ALR	0	0	56	0	0	0	0	381	16	96.21	84.11
WA	0	0	3	0	0	17	0	0	289	93.53	93.53
Total	675	364	1956	1684	580	1974	117	396	309		
		OA:	73.77%			Kappa:	0.685				

Abbreviations: OUT, outcrop unit; BUILD, building and housing; COL, colluvium; LAT, laterite; LLAT, low laterite; ALL, alluvium unit; CLWA, clear water; ALR, recent alluvium (distal); WA, water with dirt material; PA, producer accuracy; UA, user accuracy; OA, overall accuracy; Kappa, Kappa coefficient.

(Table 5). This can be explained by the presence of laterite fragments in the hillside colluvium. The predictive regolith map produced using the SVM based on field verification and validation data, was consistent with the terrain visited.

The most optimal predictive map generated using the ANN was validated using the validation dataset. These data were used to generate another confusion matrix with an OA of 73.77% and a kappa coefficient of 0.69 (Table 6). However, the outcrop class exhibited a UA of 49.02% and a PA of 70.07%. This implies that relatively more field validation pixels were identified as outcrops compared with those identified as the classified data.

Validation of the regolith landform developed from the predictive map was performed using the non-genetic regolith map overlaid on the predictive map in GIS software. This method has been used by other authors (Anand and Paine 2002; Arhin et al. 2015) to develop regolith landforms that must be differentiated from the regolith predictive map.

7. Prediction verification

The two ML tests (Figure 11) on Dataset 7 produced the most optimal predictions. The SVM predictions exhibited an OA of 86% and a kappa coefficient of 0.84, whereas ANN exhibited a better score with an OA of 87% and a kappa coefficient of 0.84. We examined saprolite from sediments at the Katara illegal mining site (Figure 5(a,b)), which were well mapped by the ANN and SVM. Saprolite was also identified from felsic and sediment in the main pit; however, one pit contained some water with saprolite, which was only mapped by ANN, whereas SVM mapped water and buildings from mine facilities (Figure 12(a–c)).

Alluvial sediments or distal sediment sites were visited on-site; they were present in low-altitude areas, and in flood plains or valleys, where the terrain is often used for agriculture (Figure 5(c,d)). This unit was well-mapped by both MLAs, as shown in the Zanikan site map (Figure 12(d–f)).

Field visits allowed the identification of laterite residuum and low laterite areas. The laterites with voids and round pisoliths are ferricrete, whereas those from the same genetic material and angular lithic material are laterite residuum (Figure 5(e,f)). The ANN generally mapped fewer laterite hills than those mapped by SVM, which mapped relatively more. In

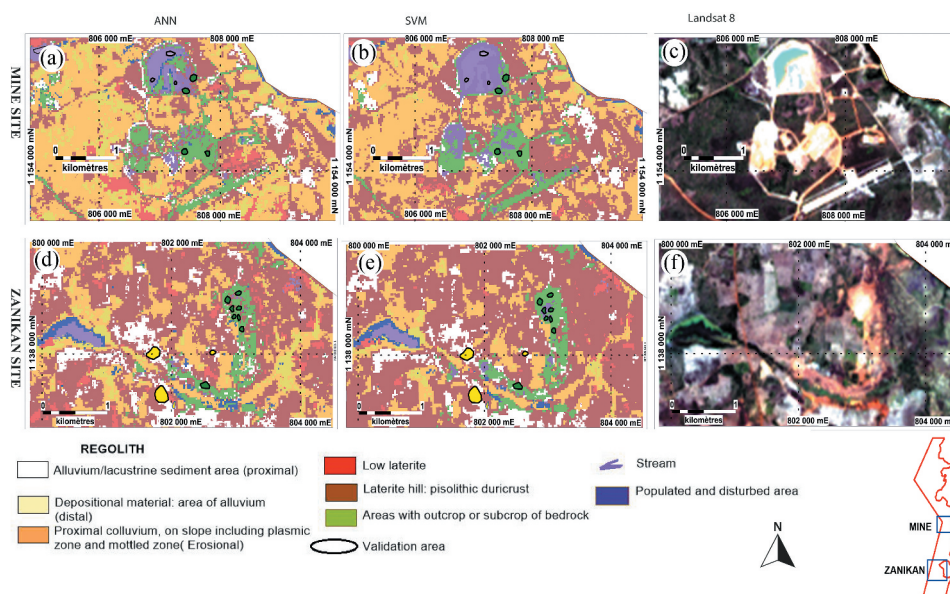


Figure 12. (a) Artificial neural network (ANN) predictive regolith map of the Sissingué Gold Project (SGP) mine showing areas with outcrops on the main pit and the airstrip. (b) Support vector machine (SVM) predictive map showing outcrop in the mine pit and relatively more water compared with that in ANN predictive map. Relatively more laterite duricrusts are surrounding the pit and mine facilities compared with that in the ANN image. (c) Landsat-8 image showing SGP mine. (d) ANN regolith predictive map showing the zanikan site in the southern part of the permit with the presence of the outcrops mapped on-site as quartz veins and sediments. All outcrops from this site were verified. (e) SVM regolith predictive map showing the same zanikan site with validated outcrop overlying the predictive map, but with relatively more laterite compared with that in ANN predictive map. (f) Landsat-8 image showing zanikan illegal mining site.

contrast, ANN detected relatively more areas with low laterite, whereas SVM detected less (Figure 12(d–f)). Colluvial material on the slope of the flank of the hill is shown on-site in a few blocks of laterite, lateritic gravel, and soil (Figure 5(g,h)). The colluvial material was well mapped by both MLAs, with relatively more colluvial material detected by ANN and less by SVM (Figure 12(d–f)).

Both MLAs yielded satisfactory results and can be used to develop regolith landform maps.

8. Prediction assessment

The regolith landform map of the SGP was created using factual data from the field to create the training zones, which generated errors in their edge areas, particularly for small outcrops because the boundaries were not clear in the images with the surrounding units (Metelka et al. 2018). Previous studies have also encountered problems with size when identifying outcrops because of the resolution of Landsat-8 and ASTER images (Cardoso-Fernandes, Teodoro, and Lima 2019; Metelka et al. 2018; Sevin 2014). To overcome this, in the first case, we selected large outcrop and sub-

outcrop areas ranging from 900–1500 m², which allowed better identification by reducing false positives in the outcrop units. However, this also depends on availability in the field.

Vegetation hinders classification based on optical data, as demonstrated in several studies using automatic or semi-automatic detection methods for geological and regolith mapping (Cardoso-Fernandes, Teodoro, and Lima 2019; Cudahy 2016; Köhler et al. 2021; Metelka et al. 2018; Sevin 2014). Our work relied on other studies of radar images to address this vegetation issue. However, data fusion using the Gram–Schmidt Pan Sharpening tool with Landsat-8 data (Table 1) and radar images offers better project classification with medium and dense vegetation in places. This fusion of data, in addition to improving the image resolution, allowed the radar to penetrate specific vegetation. Vegetation in areas with medium cover, such as in our study, is also an aid, particularly for classifying lake sediments along rivers. The vegetation in these areas is thriving and readily identifiable in Landsat-8 images and was also used to identify exotic residues in this study.

The performance and validation of the SVM confirmed that its results remained better compared with those obtained with the ANN at this site. The former is less dependent on the dimensionality of the data and training sample, whereas the ANN is relatively more sensitive to the training sample (Cardoso-Fernandes et al. 2020).

A major difficulty encountered is the use of outcrops. These sites have different spectral signatures depending on their mineralogical composition and radioelement density (Metelka et al. 2018), which are often detrimental when other outcrops have not been recorded, and spectral data can be used to accurately identify the same outcrop in the context of geological mapping (Cracknell and Reading 2014) this is ideal. However, the regolith can be misclassified. In our case, the outcrops were generally highly altered cherts with veins readily identified by the SVM, whereas the ANN did not identify all of them. This is probably due to the level of alteration and because the ANN uses a large amount of spectral information. In contrast, SVM uses patterns, shapes, and textures, and is not affected by training data or dimensionality (Cardoso-Fernandes et al. 2020).

Depending on their origin, laterites and colluviums often contain the same material (Figure 5(g,h)). They are often confused by ML when they are close. These regolith classes were difficult to predict because of their close spectral signature. The same is true for ferricretes, which are often composed of laterites re-cemented with fragments of other laterites, often of unknown origin (Butt and Bristow 2013). The latter group of recycled laterites, found at higher elevations, was challenging to detect using optical remote sensing. Radiometry is a better solution for these unit types as it offers a reasonable resolution (Arhin and Nude 2009; De Boissieu et al. 2017; Metelka et al. 2018).

Artisanal gold mining has extensively reworked the research area, with traces of rejects in rivers and access roads reflecting the spectral signatures of sedimentary rock and fragmentary clay rejects. However, excavations conducted by illegal miners also provide exposure for mapping.

9. Conclusions

This study tested two MLAs with nine different datasets for the class prediction of regolith units. The SVM provided a better prediction when considering the

validation data, OA values and kappa coefficient of the ground truth data. However, for the test data, the kappa coefficients and OA values were slightly higher when using the ANN. The OA values and kappa coefficients of the ANN for the validation data were not better than those of SVM. The most optimal data used for ML were Landsat-8 images enhanced with Gramm – Schmidt image fusion using the VH polarization of S1A images. This allowed for better classification of units, despite medium vegetation on the SGP, while improving the resolution of the final image. However, using the DEM data with the Landsat-8 and S1A images also improved the predictions.

Vegetation was also found to help delineate areas with exotic sediments, as demonstrated in this study. After the prediction, the development of the RLU map required superposition of the field data and elevation contours to manually check and edit the units for attribution and interpretation of the classes that form the RED model.

The map produced at the end of this study showed that the depositional area composed of exotic sediments and other transported units reached 36.39%. The relict regime composed of lateritic residuum and proximal soil, covered 24.91%. The erosional regime composed of outcrop areas (2.32%) and colluvial units (36.38%) covers 38.7%. The methodology used can be replicated on similar grounds to delineate regolith units and develop regolith landform maps. Moreover, ANN and SVM algorithms can assist in developing land-use and geological regolith maps using the same MLA settings. However, at the end of this study, it was difficult to distinguish between residual laterite and transported laterite. The use of high-resolution radiometric images can solve this problem. Developing regolith landform maps should be imperative for terrains in Africa, Australia, and America that are deeply altered by increasingly harsh climates. They could serve as a guide for geochemical sampling work in the early stages of mining projects.

Acknowledgements

This work was supported by the support of CURAT (University Centre for Research and Application in Remote Sensing). This study benefited from the Support of SEMS exploration and three anonymous reviewers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data that support the findings of this study are available from the corresponding author, H.A.D., upon reasonable request.

References

- Aden, I. I. 2020. "Application of Landsat 8 Satellite Images, Hydrothermal Alteration Mapping, Thermal Anomalies and Correlation with XRD Analysis: Case Study of Ne Ghoubbet Area (Djibouti)." 2–8
- Affam, M., and E. Arhin. 2013. "Lack of Exploration Success in Northern Ghana: Problems and Solutions." *Research Journal of Environmental and Earth Sciences* 5 (8): 440–448. <https://doi.org/10.19026/rjees.5.5672>.
- Alasta, A. F. 2011. "Using Remote Sensing Data to Identify Iron Deposits in Central Western Libya." In *International Conference on Emerging Trends in Computer and Image Processing (ICETCIP'2011)*, Bangkok, December 2011. <http://psrcentre.org/>.
- Anand, R., and M. Paine. 2002. "Regolith Geology of the Yilgarn Craton." *Australian Journal of Earth Sciences* 49 (1): 1–1. <https://doi.org/10.1046/j.1440-0952.2002.00912.x>.
- Arhin, E., G. R. T. Jenkin, D. Cunningham, and P. Nude. 2015. "Regolith Mapping of Deeply Weathered Terrain in Savannah Regions of the Birimian Lawra Greenstone Belt, Ghana." *Journal of Geochemical Exploration* 159:194–207. <https://doi.org/10.1016/j.gexplo.2015.09.008>.
- Arhin, E., and P. M. Nude. 2015. "Significance of Regolith Mapping and Its Implication for Gold Exploration in Northern Ghana: A Case Study at Tinga and Kunche." *Geochemistry: Exploration, Environment, Analysis* 9 (1): 63–69. <https://doi.org/10.1144/1467-7873/08-189>.
- Ballo, I., K. A. A. Hein, B. Guindo, L. Sanogo, Y. Ouologuem, G. Daou, and A. Traore. 2009. "The Syama and Tabakoroni Goldfields, Mali." *Ore Geology Reviews* 78:578–585. <https://doi.org/10.1016/j.oregeorev.2015.10.019>.
- Banerjee, K., M. K. Jain, A. T. Jeyaseelan, and S. Panda. 2019. "Landsat 8 OLI Data for Identification of Hydrothermal Alteration Zone in Singhbhum Shear Zone Using Successive Band Depth Difference Technique - A New Image Processing Approach." *Current Science* 116 (10): 1639–1647. <https://doi.org/10.18520/cs/v116/i10/1639-1647>.
- Bodruddoza Mia, M., and Y. Fujimitsu. 2012. "Mapping Hydrothermal Altered Mineral Deposits Using Landsat 7 ETM+ Image in and Around Kuju Volcano, Kyushu, Japan." *Journal of Earth System Science* 121 (4): 1049–1057. <https://doi.org/10.1007/s12040-012-0211-9>.
- Butt, C. R. M. 2016. "The Development of Regolith Exploration Geochemistry in the Tropics and Sub-Tropics." *Ore Geology Reviews* 73 (March): 380–393. <https://doi.org/10.1016/j.oregeorev.2015.08.018>.
- Butt, C. R. M., and A. P. J. Bristow. 2013. "Geomorphology Relief Inversion in the Geomorphological Evolution of Sub-Saharan West Africa." *Geomorphology* 185:16–26. <https://doi.org/10.1016/j.geomorph.2012.11.024>.
- Cánovas-García, F., P. Pesántez-Cobos, and F. Alonso-Sarriá. 2020. "FusionImage: An R Package for Pan-Sharpning Images in Open Source Software." *Transactions in GIS* 24 (5): 1185–1207. <https://doi.org/10.1111/TGIS.12676>.
- Cardoso-Fernandes, J., A. C. Teodoro, and A. Lima. 2019. "Remote Sensing Data in Lithium (Li) Exploration: A New Approach for the Detection of Li-Bearing Pegmatites." *International Journal of Applied Earth Observation and Geoinformation* 76 (April): 10–25. <https://doi.org/10.1016/j.jag.2018.11.001>.
- Cardoso-Fernandes, J., A. C. Teodoro, A. Lima, and E. Roda-Robles. 2020. "Semi-Automatization of Support Vector Machines to Map Lithium (Li) Bearing Pegmatites." *Remote Sensing* 12 (14): 2319. <https://doi.org/10.3390/RS12142319>.
- Chardon, D., J. L. Grimaud, A. Beauvais, and O. Bamba. 2018. "West African Lateritic Pediments: Landform-Regolith Evolution Processes and Mineral Exploration Pitfalls." *Earth Science Review* 179 (April): 124–146. <https://doi.org/10.1016/J.EARSCIREV.2018.02.009>.
- Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37 (1): 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B).
- Cracknell, M. J., and A. M. Reading. 2014. "Geological Mapping Using Remote Sensing Data: A Comparison of Five Machine Learning Algorithms, Their Response to Variations in the Spatial Distribution of Training Data and the Use of Explicit Spatial Information." *Computers & Geosciences* 63:22–33. <https://doi.org/10.1016/j.cageo.2013.10.008>.

- Cudahy, T. 2016. "Mineral Mapping for Exploration: An Australian Journey of Evolving Spectral Sensing Technologies and Industry Collaboration." *Geosciences* 6 (4): 52. <https://doi.org/10.3390/geosciences6040052>.
- De Boissieu, F., B. Sevin, T. Cudahy, M. Mangeas, S. Chevrel, C. Ong, A. Rodger, et al. 2017. "Regolith-Geology Mapping with Support Vector Machine: A Case Study over Weathered Ni-Bearing Peridotites, New Caledonia." <https://doi.org/10.1016/j.jag.2017.05.012>.
- Diaby, H. A. 2013. "APPLICATION DE LA TÈLÉDETECTION ET DES SIG DANS LA CARTOGRAPHIE LITHOSTRUCTURALE ET RÉGOLITHE: CAS DE L'EXPLORATION AURIFÈRE AU BURKINA FASSO." ENVI. 2009. "ENVI Atmospheric Correction Module: QUAC and FLAASH User's Guide." *Module Version*, 44. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:ENVI+Atmospheric+Correction+Module+QUAC+and+FLAASH+user's+guide#0>.
- Feybesse, J. L., and J. P. Milési. 1994. "The Archaean/Proterozoic Contact Zone in West Africa: A Mountain Belt of Décollement Thrusting and Folding on a Continental Margin Related to 2.1 Ga Convergence of Archaean Cratons?" *Precambrian Research* 69 (1–4): 199–227. [https://doi.org/10.1016/0301-9268\(94\)90087-6](https://doi.org/10.1016/0301-9268(94)90087-6).
- Köhler, M., D. Hanelli, S. Schaefer, A. Barth, A. Knobloch, P. Hielscher, J. Cardoso-Fernandes, A. Lima, and A. C. Teodoro. 2021. "Lithium Potential Mapping Using Artificial Neural Networks: A Case Study from Central Portugal." *Minerals* 11 (10): 1046. <https://doi.org/10.3390/min11101046>.
- Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel. 2006. "World Map of the Köppen-Geiger Climate Classification Updated." *Meteorologische Zeitschrift* 15 (3): 259–263. <https://doi.org/10.1127/0941-2948/2006/0130>.
- Kudamnya, E. A., W. T. Andongma, and J. O. Osumaje. 2014. "Hydrothermal Mapping of Maru Schist Belt, North-Western Nigeria Using Remote Sensing Technique." *International Journal of Civil Engineering* 3 (1): 59–66.
- MacLean, M. G., and R. G. Congalton. 2009. "Evaluating Issues in Map Accuracy: A Study of Mapping Benthic Habitats on the Texas Gulf Coast | Request PDF." https://www.researchgate.net/publication/288465683_Evaluating_issues_in_map_accuracy_A_study_of_mapping_benthic_habitats_on_the_Texas_Gulf_Coast.
- Merzah, Z. F., and H. S. Jaber. 2020. "Assessment of Atmospheric Correction Methods for Hyperspectral Remote Sensing Imagery Using Geospatial Techniques." <https://doi.org/10.1088/1757-899X/745/1/012123>.
- Metelka, V., L. Baratoux, M. W. Jessell, A. Barth, J. Ježek, and S. Naba. 2018. "Automated Regolith Landform Mapping Using Airborne Geophysics and Remote Sensing Data, Burkina Faso, West Africa." *Remote Sensing of Environment* 204 (July 2017): 964–978. <https://doi.org/10.1016/j.rse.2017.08.004>.
- Morales-Barquero, L., M. B. Lyons, S. R. Phinn, and C. M. Roelfsema. 2019. "Trends in Remote Sensing Accuracy Assessment Approaches in the Context of Natural Resources." *Remote Sensing* 11 (19): 2305–2316. <https://doi.org/10.3390/rs11192305>.
- Parra-Avila, L. A., E. Belousova, M. L. Fiorentini, L. Baratoux, J. Davis, J. Miller, and T. C. McCuaig. 2016. "Crustal Evolution of the Paleoproterozoic Birimian Terranes of the Baoulé-Mossi Domain, Southern West African Craton: U – Pb and Hf-Isotope Studies of Detrital Zircons." *Precambrian Research* 274:25–60. <https://doi.org/10.1016/j.precamres.2015.09.005>.
- Pournamdari, M., M. Hashim, and A. B. Pour. 2014. "Application of ASTER and Landsat TM Data for Geological Mapping of Esfandagheh Ophiolite Complex, Southern Iran." *Resource Geology* 64 (3): 233–246. <https://doi.org/10.1111/RGE.12038>.
- Prieto-Amparan, J. A., F. Villarreal-Guerrero, M. Martinez-Salvador, C. Manjarrez-Domínguez, E. Santellano-Estrada, and A. Pinedo-Alvarez. 2018. "Atmospheric and Radiometric Correction Algorithms for the Multitemporal Assessment of Grasslands Productivity." *Remote Sensing* 10 (2): 219. <https://doi.org/10.3390/rs10020219>.
- Ramadan, T. M., M. G. Abdelsalam, and R. J. Stem. 2001. "Mapping Gold-Bearing Massive Sulfide Deposits in the Neoproterozoic Allaqi Suture, Southeast Egypt with Landsat TM and SIR=C/X SAR Images." *Photogrammetric Engineering & Remote Sensing* 67 (4): 491–497.
- Research Systems. 2001. "ENVI User's Guide." *Research Systems*. <https://lost-contact.mit.edu/afsen/enea.it/software/rsi/envi/docs/userguid.pdf>.

- Sevin, B. 2014. "CARTOGRAPHIE DU REGOLITHE SUR FORMATION ULTRABASIQUE DE NOUVELLE-CALEDONIE: Localisation Dans l'espace et Le Temps Des Gisements Nickèlifères: Localisation Dans l'espace et Le Temps Des Gisements Nickèlifères."
- "Sissingué Gold Mine, May-June 2019 Investor Site Visit Presentation." 2019. <https://perseusmining.com/wp-content/uploads/2019/06/02110781-1.pdf>.
- Solutions, Exelis Visual Information. 2013. "ENVI." Boulder, Colorado: Exelis Visual Information Solutions. <https://www.nv5geospatialsoftware.com/Support/Maintenance-Detail/ArtMID/13350/ArticleID/16210/ENVI-51-Release-Notes>.
- Tardy, Y. 1997. *Petrology of Laterites and Tropical Soil*, 408. Rotterdam: Oxford and IBH Publishing.
- Traoré, Y. D., L. Siebenaller, S. Salvi, D. Béziat, and M. L. Bouaré. 2016. "Progressive Gold Mineralization Along the Syama Corridor, Southern Mali (West Africa)." *Ore Geology Reviews* 78 (November): 586–598. <https://doi.org/10.1016/j.oregeorev.2015.11.003>.
- Veci, L. 2015. "Orthorectification Tutorial." *Science Toolbox Exploitation Platform*, March: 20. PALSAR Orthorectification Tutorial.pdf. <http://step.esa.int/docs/tutorials/ALOS>.