Received: 16 April 2023 | Revised: 13 October 2023 | Accepted: 30 November 2023

DOI: 10.1002/esp.5759

RESEARCH ARTICLE

ESPL WILEY

Classifying types of gully changes with unoccupied aircraft vehicles 3D multitemporal point clouds for training of satellite data analysis in Northwest Namibia

Miguel Vallejo[1,2] | Carlos Castillo[3] | Vivien Zahs[2] | Olaf Bubenzer[2,4] | Bernhard Höfle[2,4,5]

[1]Department of Land and Spatial Sciences, Namibia University of Science and Technology, Windhoek, Namibia

[2]Institute of Geography, Heidelberg University, Heidelberg, Germany

[3]Department of Rural Engineering, Civil Constructions and Engineering Projects, University of Córdoba, Córdoba, Spain

[4]Heidelberg Center for the Environment (HCE), Heidelberg University, Heidelberg, Germany

[5]Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany

Correspondence
Miguel Vallejo, Department of Land and Spatial Sciences, Namibia University of Science and Technology, 13 Jackson Kaujeua Street, Windhoek 13388, Namibia.
Email: morti@uni-heidelberg.de

Funding information
This work was supported by the DLR TanDEM-X Science Team under Grant DEM_HYDR2024, Namibia University of Science and Technology under Grant ILMI Seed Grant programme and Heidelberg University under the Kurt-Hiehle-Foundation funding pogramme.

Abstract

The development of standardised data acquisition strategies and analytical workflows is crucial to quantify gully changes. In this study, we explore synergies between unoccupied aircraft vehicles (UAV) and satellite remote sensing in order to classify gully morphodynamics. Using Time Series Forest (TSF) and the Sentinel-1 radar backscatter coefficient (σ_o), gully scenarios can be classified into four categories: gully topographical change, no change outside gully, no change inside gully, and non-topographical change. In addition, a Random Forest (RF) classification is performed employing individual features obtained from elevation models and temporally aggregated datasets. Training data are generated from multitemporal UAV-borne photogrammetric point clouds through a manual segmentation of different gullies in Namibia. This information is transferred from point clouds (sub-m) to satellite imagery (10 m) generating training data at Sentinel-1 pixel level. Results indicate that the TSF (on the σ_o Vertical-Vertical polarisation) and RF (on temporally aggregated features) perform best when training and testing areas are located within the same geographical extent. Both approaches yield similar Total Accuracy (TA $\approx$ 79%–80%) and Cohen's Kappa value (Kappa $\approx$ 0.7), but TSF achieves superior Producer Accuracy (PA = 78.5%) and User Accuracy (UA = 84.6%) for the gully topographical change class. Additionally, the utilisation of TSF in Vertical-Vertical polarisation is the most effective method if the testing and training areas are in different geographical locations, allowing gully identification with TA > 80% and Kappa = 0.49. However, this method presents limitations to precisely delineate the change types, as dynamics are rain-driven and therefore are geographically related. In summary, by combining the complementary benefits of UAV-based and satellite-based solutions, this study opens a line of research for the study and classification of surface land dynamics and geomorphological feature extraction in regional extents.

KEYWORDS
change detection, gully monitoring, machine learning, Namibia, radar time series, spatial aggregation, spatial classification

1 | INTRODUCTION

Gullies are dynamic geomorphological forms that accelerate land degradation and lead to morphological and environmental changes in the landscape at the micro and macro geographical scale (Montanarella et al., 2015). Their study has been approached from an increasing number of scientific perspectives, spatial scales and geographic regions (Bennett & Wells, 2019; Castillo & Gómez, 2016). Some remarkable examples of long-term joint efforts are the studies conducted in Australia to map (Shahabi et al., 2019), assess the impact (Hancock & Evans, 2010) and remediate inland gully erosion (Wearne et al., 2018). In China, various initiatives are mainly focused on developing gully control practices in agricultural lands (Guo et al., 2020; Liu et al., 2019). Furthermore, extensive research focuses on the development of empirical models to map gully heads at a global and continental scale (Omran et al., 2022; Sidorchuk, 2021; Vanmaercke et al., 2020). Identifying and measuring gullies in large extents at a given time in 2D has been addressed by various authors (Golosov et al., 2018; Shahabi et al., 2019; Vallejo-Orti et al., 2021). Nevertheless, certain aspects remain understudied and relevant, such as the temporal dimension of gullies and techniques to measure their evolution over long periods, that is, >5 years (Hayas et al., 2017), and at regional scales (Vanmaercke et al., 2021), that is, >1000 ha (Kimura et al., 2009; Morgan, 2005). These approaches would provide geomorphologists, regional planners, and agronomists with useful information to better understand triggering factors, gully-human interactions, and to plan remediation strategies (Vanmaercke et al., 2021) as well as to calibrate and evaluate empirical methods (Omran et al., 2022; Sidorchuk, 2021).

Previous studies show that 3D monitoring of the geomorphic activity of gullies is typically based on pairwise change quantification between two epochs. Terrestrial (Zhong et al., 2021) and airborne (Du et al., 2021) laser scanning as well as UAV-borne photogrammetry-based dense image matching (Nota et al., 2022) are well-established close-range sensing techniques for generating these 3D datasets. Their use is valuable due to the achieved level of detail (sub-cm); however, they are limited to relatively small geographic areas (i.e., areas <5 ha). As gullies are complex systems where numerous types of processes occur (i.e., detachments, laminar erosion, vegetation changes and stable areas), it is essential to classify gully sections by their predominant type of activity. Using 3D close-range sensing techniques enables the detection and quantification of minute surface changes, as well as the extraction of 3D surface information. Consequently, although their application to large geographical areas is costly and time-consuming, they are of great value as potential training and reference data of surface dynamics for satellite-based solutions.

As scalable and transferable methods are desired to cover large areas and longer time periods, satellite observation systems with global coverage, all-weather performance, high-frequent re-visits need to be considered. In addition, the selected sensors need to capture distinguishing features of gullies, such as terrain ruggedness, soil compaction, water balance, and vegetation type and distribution. Synthetic Aperture Radar (SAR) Sentinel-1 (S1) offers adaptability to observe these features and fulfil the aforementioned requirements. As one of SAR direct features, backscatter coefficient (σ_o) is the normalised measure of the reflective strength of a radar target defined as per unit area on the ground. σ_o provides information about soil physical and chemical features and vegetation status. Thus, a spatio-temporal analysis of σ_o reveals insights into land surface changes for much larger geographic areas and offers short intervals between acquisitions (Sentinel-1 SAR User Guide, 2021). The detection of earth surface changes using SAR is widespread, for example, to monitor snow landscapes changes (Snapir et al., 2019), dune dynamics (Rozenstein et al., 2016), land subsidence (Canova et al., 2012; Pu et al., 2022), river (Ahmad & Kim, 2019) and soil moisture dynamics (Ullmann et al., 2023).

Within this context, we investigate how four types of dynamics (gully topographical change, no change outside gully, no change inside gully and non-topographical change) identified in point clouds can be classified in satellite datasets. The Time Series Forest (Deng et al., 2013), and the Random Forest (Ho, 1995) algorithms are applied on radar time series and on 11 single features, respectively. We apply and test our approach in the Kunene Region in Namibia over the course of one and a half years encompassing the rainy seasons of 2019–2020 and 2020–2021. During this period training data are generated, firstly from the same area, and secondly from a different gully network located in the Krumhuk Farm, in central Namibia.

In this line, the primary objectives of this research are (i) to develop an approach for transferring information from very high-resolution 3D UAV data to an operational scale in space and time, based on spaceborne Earth observation; (ii) to study whether the use of the temporal signature of a single radar variable (σ_{oVV} or σ_{oVH}) can replace or improve the use of multivariate analysis based on elevation models and temporally aggregated features; (iii) to explore the geographic transferability of the methods to determine the viability of using a training gully site distinct from the classification target one.

2 | STUDY AREA AND DATASETS

2.1 | Study areas

The first study area is in the Kunene Region in Namibia, 15 km towards the south of Opuwo (Figure 1a,b), covering part of the Okatjandja Kozomenje conservancy. Our second study area is in the Krumhuk Farm, in central Namibia, located approximately 30 km from Windhoek (national capital city) towards the south. One of the major gullies in this farm is selected as alternative training area (Figure 1a,c). Both study sites are representative of large active valley bottom gullies in (semi-) arid areas, typical in Namibia (Mendelsohn et al., 2002), South Africa (le Roux et al., 2022; Olivier et al., 2023) and other regions (Shahabi et al., 2019).

Between November 2019 and March 2021, four UAV data collection campaigns were conducted in the Kunene Region and eight in the Krumhuk Farm. The extents covered during these campaigns are framed by a blue polygon in Figure 1b,c.

A summary of both study area descriptions is presented in Table 1.

As both study sites present similarities in terms of soil and climate, the main differences are identified in the gully morphology. The gully in Kunene is larger, with dimensions of up to 300 m wide and 15 m deep as compared to the 50 and 5 m in width and depth of the gully in Krumhuk (Figures 1d,e and 2). Krumhuk also has denser grass coverage during the rainy season.

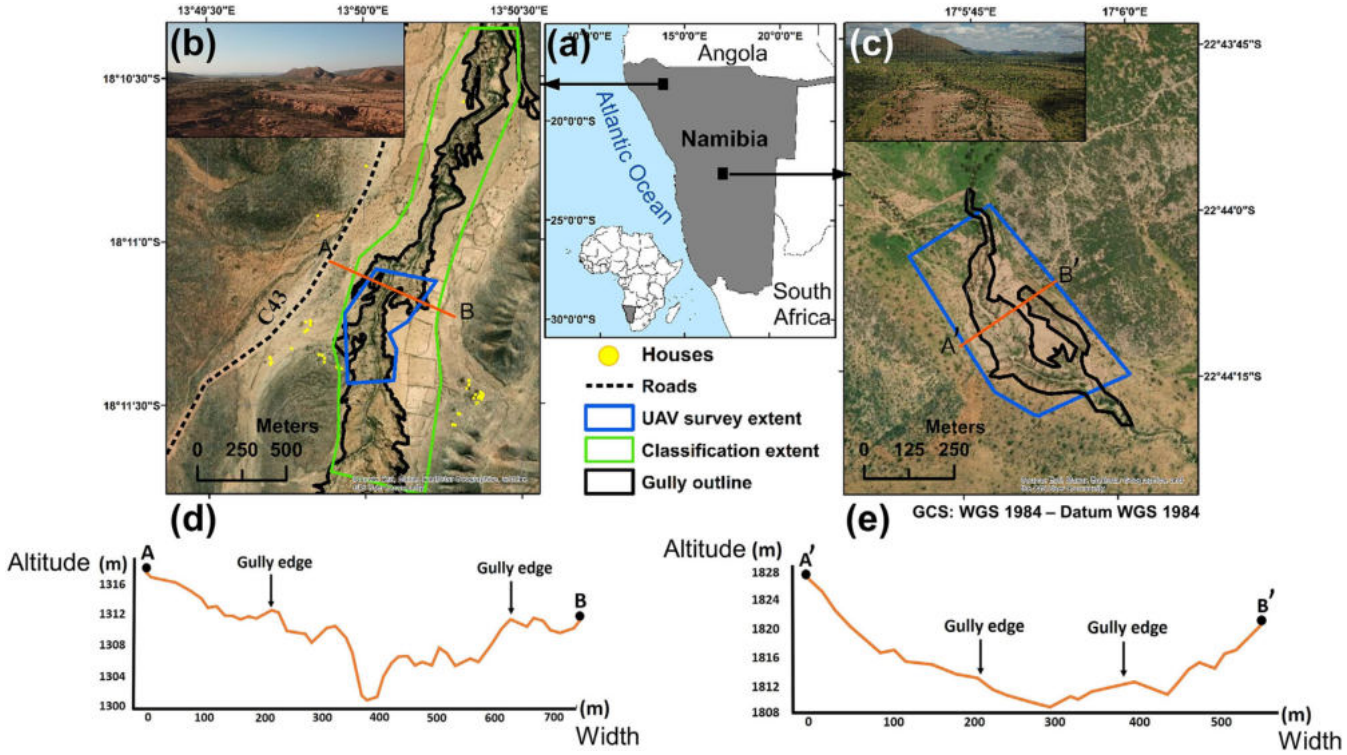

FIGURE 1 (a) General map of Namibia locating the two study areas. (b) Gully in the Kunene region with the target classification area (green polygon), the UAV training/testing area (blue polygon), the gully outline (black line), the main road (dashed black line) and some houses (yellow dots) (Namibia Statistical Agency, 2021). (c) Gully in Krumhuk farm displaying the UAV training area (blue polygon) and the gully outline (black line). Terrain topographic profiles across gullies extracted from Tandem-X DEM (DLR, 2014) in the (d) Kunene region and (e) Krumhuk farm. Base maps source is ESRI, Maxar earthstar Geographics. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Descriptive parameters of the study areas.

	Opuwo	Krumhuk	
Location (WGS 1984)	Lon: 13.83°, Lat: −18.19°	Lon: 17.22°, Lat: −22.71°	
Soil Type	Lithic leptosols and eutric regosols	Regosols, eutric leptosols and lithic leptosols	(Jones et al., 2013)
Altitude (m a.s.l)	800–1400	1753–2335	
Annual Rainfall (mm)	300–350	250–350	
Vegetation	Grasslands and scattered trees	Dense shrubland	(Mendelsohn et al., 2002).

Typically, as seen in the gully site in Kunene Region (Figure 3), the outer limits of the gully are the most active zones, where erosive linear advancement driven by concentrated water flows and local detachments in the most vertical walls are frequent ([1] gully topographical changes). An intermediate zone with gentle slope and sparse vegetation is characterised by a more stable erosion area inside the gully ([3] no change inside gully). Permanent woody vegetation is identified as the primary non-topographic elevation change in the area ([4] non-topographic change) in the main channel's central gully sections. Finally, non-sloped areas in the gully surroundings have not been affected by gully erosion ([2] no change outside gully). These generalised types of changes are used as the foundation to describe the four target classes in subsequent sections.

2.2 | Datasets

The contents of Table 2 outline the characteristics, derivatives and acquisition times of the source datasets utilised in this research.

In Sections 2.2.1–2.2.3, we elaborate on the specifics of data collection, pre-processing and the various datasets preparation for the subsequent analysis.

2.2.1 | Photogrammetric point clouds and derived datasets

The training data for this study is generated through UAV-borne photogrammetry using a DJI Mavic Pro with a model FC220 embedded

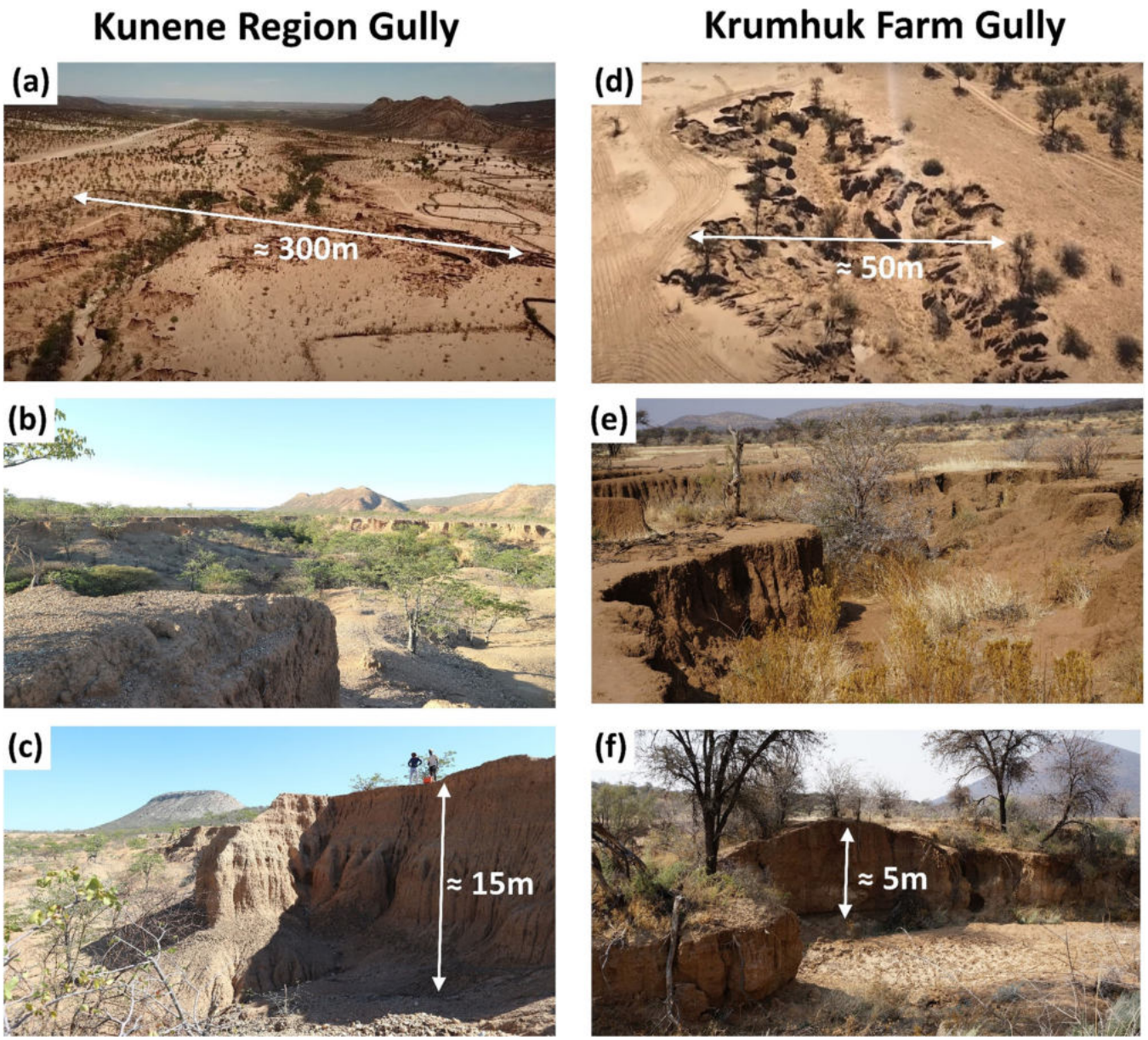

FIGURE 2 Representative descriptive photos of both study sites. (a) General view, (b) interior perspective and (c) close-up of a gully wall in the Kunene region. (d) Overall view, (e) interior perspective and (f) close-up of a gully wall at the Krumhuk farm. [Color figure can be viewed at wileyonlinelibrary.com]

camera (DJI, 2022). Aerial imagery from two campaigns (November 2019 and March 2021) for each study area is selected on the dates indicated in Table 2. The flights are carried out at a constant altitude of 100 m above-ground level (AGL) with 70% front and side overlap between images and 5-mm focal length, producing imagery with 72 dpi resolution and 2.3 cm ground sampling distance (GSD).

For November 2019 and March 2021, 3D point clouds are reconstructed in a dense image-matching procedure using Agisoft Metashape (Agisoft Metashape, 2021), resulting in an average point spacing of 10 cm (std. dev.: 8 cm). Global georeferencing is applied on the point clouds using 10 surveyed ground control points (GCP) with Garmin GPS MAP 64, computing waypoints-averaged fixes (<0.5 m accuracy), yielding average horizontal georeferencing residuals of 2.5 cm between the GCP and the point clouds. The multitemporal point clouds are aligned by minimising the 3D distance between these point clouds in stable areas outside the gully using an iterative closest point (ICP) algorithm (Besl & McKay, 1992), detecting a mean registration error (root-mean-square error) of 0.1 m. Subsequently, M3C2 (linear multi-scale) distance and significant change are derived (Lague et al., 2013). We used an ICP and M3C2 algorithm implementation in Cloud Compare (Cloud Compare, 2022). M3C2 computes distances between point clouds based on a local normal vector along which a cylinder is oriented. This cylinder captures points from each point cloud, which are averaged (separately for each point cloud) and compared to derive a linear multi-scale distance (DiFrancesco et al., 2020; Zahs et al., 2022). The registration error (derived from the ICP) and the local roughness of each point cloud determines the level of detection (LoD), which is a measure of uncertainty associated to each change, as defined in Lague et al. (2013). Although LoD is spatially variable, it presents a mean = 0.70 m and std. dev = 0.19 m for the entire point cloud, resultant of the M3C2 operation. As a final derivate, the significant change (*SC*) parameter indicates whether the M3C2 distance computed for a certain point corresponds to a real change (*M3C2 distance* > LoD) or not (*M3C2 distance* ≤ LoD). Thus,

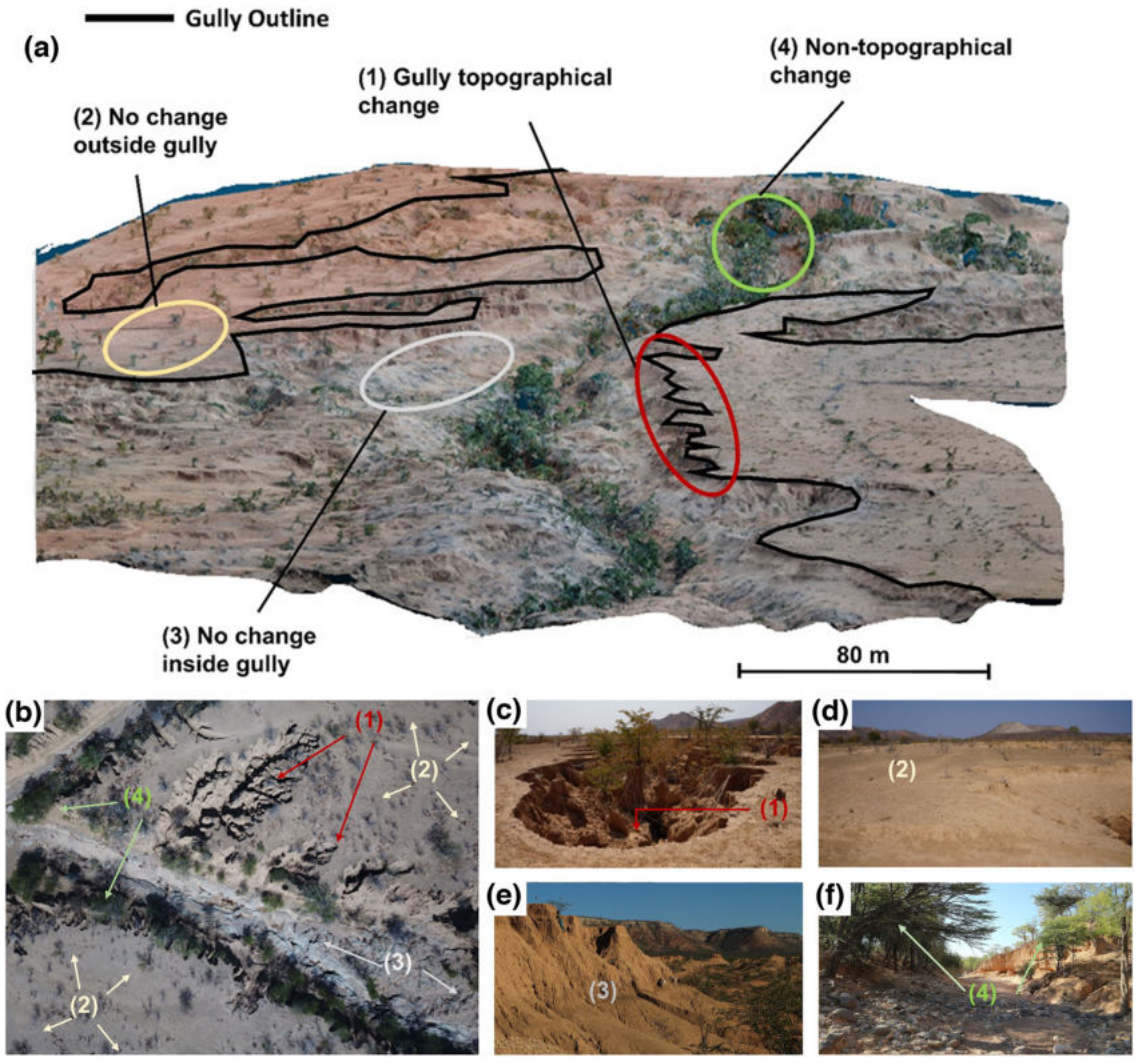

FIGURE 3 (a) 3D point cloud of the gully site in the Kunene region with the approximated gully outline (in black line) and the four generalised identified changing scenarios ([1] gully topographical change, [2] no change outside the gully, [3] no change inside the gully and [4] non-topographical change). (b) Aerial drone image locating the different change types, displayed in detail in subfigures (c)–(f) for each class of change (1, 2, 3 and 4), respectively. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Description of the main data sources. The columns display, from left to right, the name of the original dataset, products derivates, ground sampling distance (GSD) and acquisition dates (grouped by years appearing in bold, with dates in the format month.day).

Dataset	Derivates	GSD	Acquisition dates (year: month.day)
Sentinel-1 GRD	Backscatter coefficients ($\sigma_{o\ VV}$, $\sigma_{o\ VH}$)	10 m	**2019:** 07.07, 07.19, 07.31, 08.08, 09.05, 09.17, 09.29, 10.10, 10.23, 11.04, 11.16, 11.28, 12.10.12.22 **2020:** 01.03, 01.15, 01.27, 02.08, 02.20, 03.03, 03.15, 03.27, 04.08, 04.20, 05.02, 05.15, 05.26, 06.07, 06.19, 07.01, 07.13, 08.06, 08.18, 09.11, 09.23, 10.05, 10.17, 10.29, 11.10, 11.22, 12.04, 12.16, 12.28 **2021:** 01.21, 02.14, 02.26, 03.10, 03.22, 04.03, 04.15, 04.27
Sentinel-1 SLC	Backscatter mechanism (*SM*)	10 m	**2019**.11.04, **2021**.04.03
Sentinel-2 MSIL2A	Normalized Difference Vegetation Index (*NDVI*)	10 m	**2019**.11.09, **2021**.03.03
TanDEM-X HRTI-3 Global DEM	Terrain Ruggedness Index (*TRI*), Slope (*S*)	12 m	**2015**
UAV-borne photogrammetry	M3C2 distance (*M3C2*) Significant change (*SC*) Slope (*S*) Terrain Ruggedness Index (*TRI*) Green leaf Index (*GLI*)	0.1 m	**2019**.11, **2021**.03

the present configuration differentiates statistically significant changes ($SC = 1$), for example, produced by gullying activity, from less significant ($SC = 0$) smaller-scale processes, like laminar and rill erosion (smaller than 0.7 m).

As a derivate of the point clouds, RGB (Red, Green, Blue) orthoimages and DEMs are also generated and processed in R (R, 2021) 'raster' package to create Terrain Ruggedness Index (*TRI*), Slope (*S*) and Green Leaf Index (*GLI*) (Louhaichi et al., 2001). *TRI* is the resultant square root of the squared and averaged elevation difference between the centre pixel and its eight adjacent pixels of a moving kernel (Riley et al., 1999). It represents local surface spatial variability. The *GLI* uses the RGB channels of an optical image to extract information of the vegetation status, ranging from non-living (negative GLI) to green vegetation (positive GLI), according to Expression (1).

$$GLI = \frac{2G - R - B}{2G + R + B} \quad (1)$$

2.2.2 | Radar backscatter time series

S1-GRD (Ground Range Detected) products (ESA, 2021a) are used to derive σ_{oVV} and σ_{oVH} for the full time series between July 2019 and April 2021. To generate σ_o from S1-GRD imagery, pre-processing operations are carried out in ESA SNAP software (ESA, 2021c), including orbit correction, radiometric and geometric calibration. Global TanDEM-X DEM (DLR, 2014) is used to apply geometric corrections to S1-GRD. These operations are repeated for each epoch (relative orbit 29, frame 1106 ascending) generating time series of σ_{oVV} and σ_{oVH} images. These images time series (σ_{oVV} and σ_{oVH}) are cropped to the extent of our study area and projected to UTM Zone 33 South, WGS 1984 ellipsoid (EPSG:32733).

Recent studies on the geolocation accuracy of SAR mission products (Gisinger et al., 2020; Small & Schubert, 2022) indicate that S1 registers absolute location errors (ALE) ranging from sub-m to approximately 3 m, which fits the requirements for our analysis.

2.2.3 | Single and temporally aggregated features

Utilising the σ_{oVV} and σ_{oVH} time series, the rate between the first and last values of the series ($\Delta\sigma_{oVV}$ and $\Delta\sigma_{oVH}$), and the standard deviation for the entire series (SD [σ_{oVV}] and SD [σ_{oVH}]) are computed as temporally aggregated values for each pixel.

Three classes reflecting dominant backscatter mechanism (volume, double-bounce and surface backscatter) are derived from S1-SLC (Single Look Complex) products (ESA, 2021a) conducting a H-Alpha Wishart dual polarisation unsupervised classification (Lee et al., 1999) in ESA SNAP software. This results in the dominant scattering mechanism for Nov 2019 (SM_{2019}) and March 2021 (SM_{2021}), as well as the computed pixel-based binary change $|\Delta SM|$.

TRI and S are also computed from Tandem-X DEM in R (R, 2021) 'raster' package.

Sentinel-2 (S2) (ESA, 2021b) datasets for November 2019 and March 2021 are used to produce the Normalized Difference Vegetation Index *(NDVI)* (Rouse et al., 1973), to derive vegetation change ($\Delta NDVI$) and average status μ (*NDVI*).

The characteristic of each dataset is presented in Table 3:

3 | METHODS

The proposed approach aims to classify types of activity at satellite spatial resolution in a gully zone using point clouds as training data (Figure 4). Aligned with the dynamics depicted in Figure 3, the four target classes are: Class 1 (gully topographical change; C1), Class 2 (no change outside gully; C2), Class 3 (no change inside gully; C3) and Class 4 (non-topographical change; C4).

These classes are manually segmented as polygons in the point clouds by experts in gully erosion. They are subsequently interpolated to the whole point cloud using a Random Forest (Ho, 1995) and upscaled to S1 as training pixels. To classify gully change at satellite pixel level, a time series classifier, Time Series Forest (TSF) is evaluated and compared to a Random Forest (RF) approach using temporally aggregated features (see Table 3) as a proxy of type of change. The developed workflow is presented in detail in Figure 5, and the different procedures are explained in Sections 3.1–3.4.

3.1 | Generation of UAV reference data

The process of manually identifying the four classes involves the expert-based observation of differences between two distinct RGB point clouds. This identification is facilitated by the M3C2 distance (*M3C2*) (Lague et al., 2013) and associated significant change (*SC*) between November 2019 and March 2021. Polygons for each class are manually segmented. $SC = 1$ determines changing areas, which are separated into gully topographical change (C1) or non-topographical change (C4) with support of RGB information. No changing sites ($SC = 0$) can be separated between those inside (C3) or outside the gully (C2). In Figure 6a–d we show framed areas enclosing examples of points of each class. Figure 6e presents a profile line showing the typical respective topographic changes, with neutral change in C2 and C3, negative change in C1 and positive in C4.

To derive a wall-to-wall reference data, the manually labelled UAV training data is used to conduct the RF classification applied to the whole point cloud resulting from the M3C2 distance operation. The same number of sampled points is selected for each segmented class, split 80% for training and 20% for testing. Five explanatory variables are used and extracted from the UAV-derived data point clouds: (1) *M3C2*, (2) *SC*, (3) *S*, (4) *TRI* and (5) *GLI*. The RF classification generates a point cloud (UAV-predicted reference data) where each point is assigned to a class with a prediction probability.

3.2 | Generation of satellite pixel training data

To solve the resolution mismatch between UAV derived point cloud and satellite (S1) product, we transferred the information from the point clouds (0.1 m) to the spatial pixel resolution and extent of S1 (10 m). Figure 7 depicts the component of the methodology dealing with the information resampling from point clouds to satellite pixels.

TABLE 3 Summary of the characteristics of the 11 features utilised for random Forest classification at the pixel level. The columns present, from left to right, the variable name, the variable unit, brief description, source product and an indication of whether they are temporally aggregated or one-time features.

Variable	Units	Description	Source product	Temporally aggregated
(12) $\Delta\sigma_{oVV}$ (13) $\Delta\sigma_{oVH}$	dB	Rate of σ_o between November 2019 and March 2021 for VV and VH polarizations	Sentinel-1 GRD	YES
(14) *SD* (σ_{oVV}) (15) *SD* (σ_{oVH})	dB	The standard deviation for the time series of VV and VH polarizations	Sentinel-1 GRD	YES
(16) SM_{2019} (17) SM_{2021}	3 nominal classes: volume, double-bounce and surface.	The dominant scattering mechanism for Nov 2019 and March 2021	Sentinel-1 SLC	NO
(18) $\lvert\Delta SM\rvert$	0- No Class Change 1- Class Change	A binary dataset representing change or no change in the dominant scattering mechanism between Nov 2019 and March 2021.	Sentinel-1 SLC	YES
(19) *TRI*	Metres (m)	Topographic Roughness Index in 2015	TanDEM-X HRTI-3	NO
(20) *Slope*	Sexagesimal degrees (°)	Terrain Slope (°) in 2015	TanDEM-X HRTI-3	NO
(21) *ΔNDVI*	No units (−1 to 1)	Vegetation change expressed in Rate of NDVI index between November 2019 and March 2021.	Sentinel-2 MSIL2A	YES
(22) *μ (NDVI)*	No units (−1 to 1)	The mean for NDVI for the time series between November 2019 and March 2021.	Sentinel-2 MSIL2A	YES

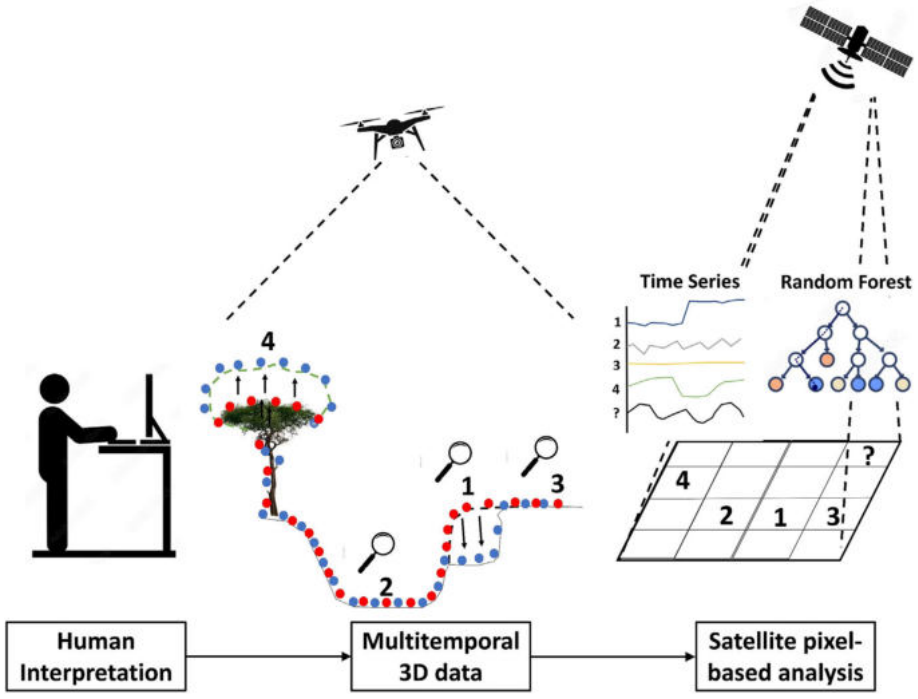

FIGURE 4 Conceptual explanation for the proposed approach. Types of changes ([1] gully topographical change, [2] no change outside the gully, [3] no change inside the gully and [4] non-topographical change) are identified by humans (human interpretation) on the original and on the already analysed multitemporal point clouds (multitemporal 3D data) and transferred to Sentinel-1 pixel scale to apply pixel-based classifications (satellite pixel-based analysis). The source of the icons used in this figure is stock.adobe.com. [Color figure can be viewed at wileyonlinelibrary.com]

For this purpose, zonal statistics are implemented for each S1 pixel coverage through the calculation of a purity weighted index (q_c). q_c ranges from 0 to 1 and indicates the degree of dominance (high $q_c = 1$ or low $q_c = 0$ dominance) of each class c. As presented in Expression (2), p_c is the number of points of a certain class within a pixel, $\overline{p_c}$ is the average probability of these points of belonging to this class c resultant of the RF classification, p_T is the total number of points inside the same pixel and $\overline{p_T}$ is the average probability of all points of belonging to their assigned class. Therefore, the following expression is calculated four times for each pixel, generating q_1, q_2, q_3 and q_4 indices.

$$q_c = \frac{p_c \cdot \overline{p_c}}{p_T \cdot \overline{p_T}} \quad (2)$$

Thus, a pixel that hypothetically contains only points of C1, will obtain $q_1 = 1$, regardless of the individual probability associated to each point.

Lastly, to ensure sufficient point coverage within each pixel for further analysis, a pixel must encompass a quantity of points surpassing half of the median number of points found in all pixels. Consequently, pixels failing to meet this criterion are excluded, leading to gaps in the dataset, as illustrated in Figure 8e.

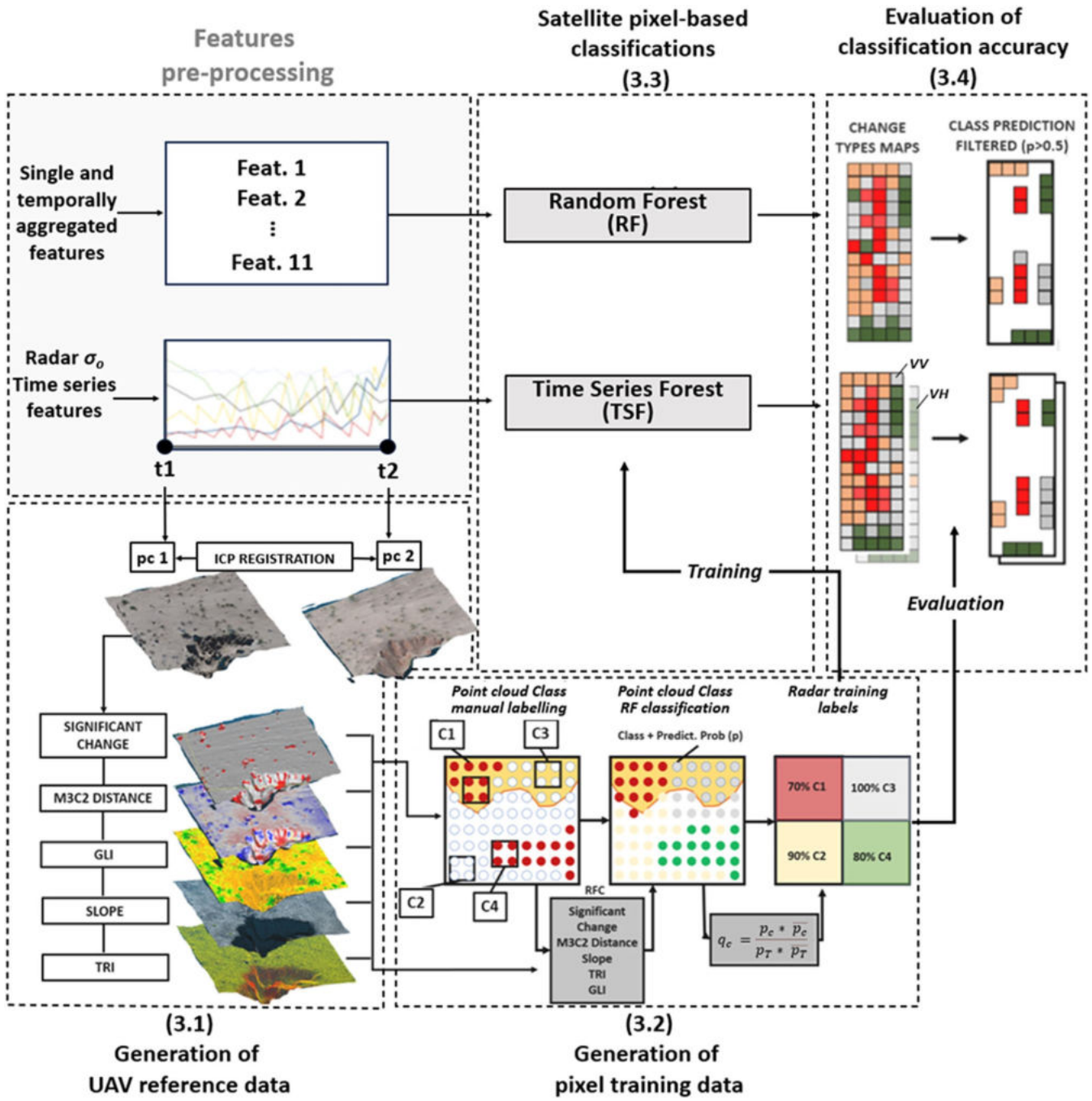

FIGURE 5 Description of the methodology. Each frame corresponds to the methodological block, beginning with the pre-processing of the time series and temporally aggregated features, and followed by the (3.1) generation of UAV reference data, (3.2) generation of pixel training, (3.3) satellite-based classifications generating independent results for time series Forest ($\sigma_{vv,}$ and σ_{vh}) and random Forest and (3.4) evaluation of classification accuracy. [Color figure can be viewed at wileyonlinelibrary.com]

3.3 | Satellite pixel-based classifications

We conduct a new classification extrapolated to the entire study area based on S1-σ_o time series and TSF algorithm. Features such as the mean, standard deviation, and slope of σ_o time series are extracted in different moving windows. These features are used to configure the TSF classifier, with the total number of features resultant from the product of the number of simple features multiplied by the predefined number of windows (Deng et al., 2013).

The number of training samples is set to the least represented class, selecting those pixels with highest q_c in the rest of the classes. σ_o time series (50-time epochs) for the Vertical-Vertical (VV) and Vertical-Horizontal (VH) polarisations are utilised as explanatory features. This leads to two independent classifications of the TSF algorithm (Figure 9b,c) where, as per standard procedure, 80% of the samples are used for training the classifier and 20% are reserved for testing the results. Additionally, different training/testing proportions are computed to observe the performance of the algorithms with different training datasets sizes (Figure 9e). The TSF is defined with 100 trees and 200 moving windows, as results become optimal and stable beyond this threshold (Figure 9f).

The same training data is used to fit a model based on RF (constructed with 100 trees) with 11 features (listed in the Table 3)

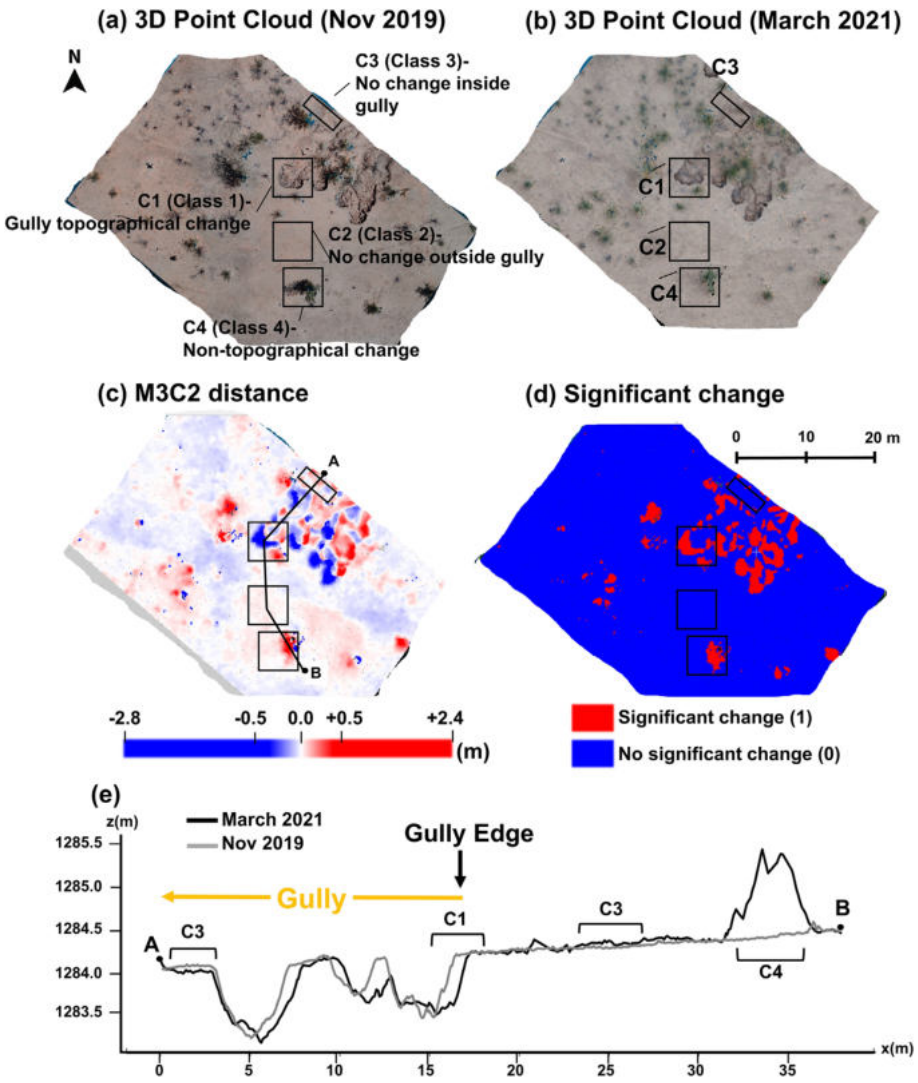

FIGURE 6 Azimuthal view of point clouds of a gully changing site captured in (a) November 2019 and (b) March 2021. Computed (c) M3C2 distance and (d) significant change between (a) and (b) with frames pointing out to each type of change class. (e) Elevation profiles for November 2019 (grey) and March 2021 (black) along a line crossing each predefined class. [Color figure can be viewed at wileyonlinelibrary.com]

derived from variables that can represent change in the terrain and/or distinguish different types of terrain (Figure 9d).

3.4 | Evaluation of classification accuracy

As previously mentioned, we assess the accuracy of the classifications conducted at the point cloud and satellite pixel levels by splitting our reference data into training and testing datasets at an 80% to 20% proportion. For point clouds, the reference data includes manually tagged points as described in Section 3.1. On the other hand, for pixel-based classification, we utilise the resultant pixels outlined in Section 3.2 for training and validation.

The performance of classifications is assessed using overall metrics such as Total Accuracy (TA) and Cohen's Kappa (Kappa), along with specific metrics for individual classes like Producer Accuracy (PA) and User Accuracy (UA). These metrics are calculated based on the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

The TA expresses the percentage of well-classified pixels among all classified pixels. Kappa coefficient is a statistic that is used to measure inter-rate reliability for qualitative (categorical) items (Cohen, 1960).

The PA represents the probability of a reference pixel (on the ground) being accurately classified. It is calculated by dividing the number of correctly classified pixels of a specific class by the total number of reference pixels belonging to that class (Congalton, 1991).

The UA is the probability that a classified pixel on the map accurately represents its corresponding category on the ground. It is calculated by dividing the number of correctly classified pixels of a specific class by the total number of classified pixels for that class on the map (Story & Congalton, 1986).

Additionally, possible sources of errors are also investigated aiming at evaluating the limitations of the methodology through the application of a statistical MANOVA test (Stahle, 1990) for comparing multivariate sample means.

4 | RESULTS

4.1 | UAV training data generation

The RF classification applied on the UAV point cloud (Figure 8d) achieves a TA = 99% and Kappa = 0.98, with UA and PA for each class above 98% (Table 4). As the objective is to create exhaustive continuous data in the point cloud using the RF classifier as an interpolator, external testing datasets are not used to validate the accuracy. The most outstanding features for this classification are TRI and GLI, both presenting RF feature importance of approximately 25%.

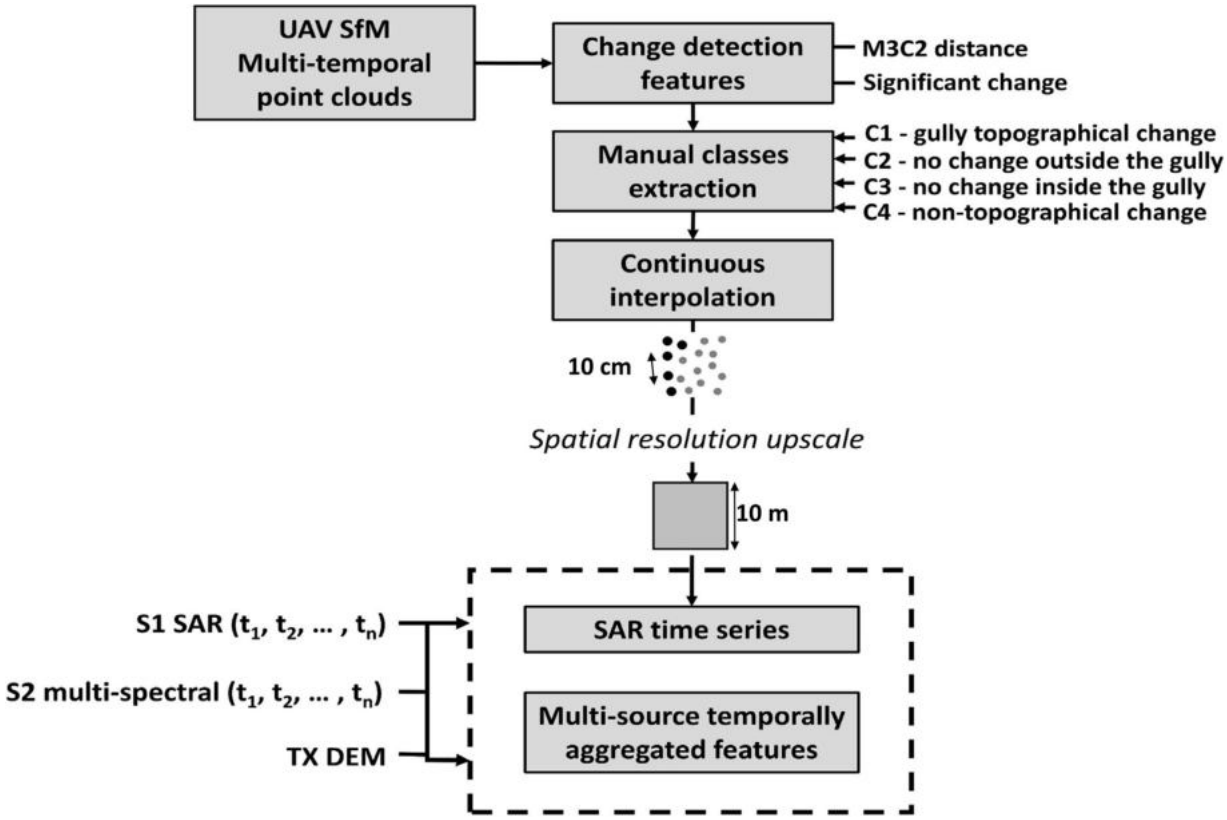

FIGURE 7 Graphical representation of the resampling process of the four classes ([1] gully topographical change, [2] no change outside the gully, [3] no change inside the gully and [4] non-topographical change) denoting gully change types from point clouds to sentinel pixel resolutions.

Transferring this information from point cloud level to S1 satellite level (Figure 8d,e) generates 1321 pixels (88, 413, 740 and 80 pixels for C1, C2, C3 and C4 respectively), together with their corresponding purity weighted index (q_c).

4.2 | Satellite remote sensing-based classifications

Results for the pixel classifications and their corresponding accuracy reports are presented in Figure 9b–d and Table 5.

The resulting accuracy values are similar for TSF (σ_{oVV}), TSF (σ_{oVH}) and RF when the proportion between training and testing data is 80/20, with TA close to 80% and Cohen's Kappa values around 0.7. They all classify changing scenarios with UA and PA always > 50%.

It is noteworthy that for TSF, C1 is classified with UA > 80% and PA > 78% in both polarisations, but the σ_{oVV} dataset poses superior capacity to classify C2, C3 and C4 than σ_{oVH}. In terms of feature importance, mean (35%), standard deviation (33%) and slope (32%) of the different changing windows are well balanced in both polarisations.

In the RF classifier utilising 11 individual features as shown in Figure 9d, the performance is slightly superior to the best results obtained using the entire time series discussed earlier. This yields TA = 80% and Kappa = 0.73. This procedure successfully separates C2 from the others (UA and PA > 90%), and it also possesses the capacity to separate the C1 with PA = 56% and UA = 81%. Sorted list of each feature importance (*TRI* [15.8%], μ (*NDVI*) [14.5%], Δ*NDVI* [14.4%], *S* [11.7%], SM_{2019} [9.43%], SM_{2021} [8.25%], SD (σ_{oVV}) [6.5%], $\Delta\sigma_{oVH}$ [6.4%], $\Delta\sigma_{oVV}$ [6.3%], SD (σ_{oVH}) [5.6%] and Δ*SM* [0.9%]) shows that most contributing features are topographic features (*TRI* and *S*) and the features reflecting modifications in the photosynthetic activity (μ [*NDVI*] and Δ*NDVI*).

Moreover, the experiment comparing the TSF and RF performances evaluating different portions of training data from 0.1/0.9 to 0.9/0.1 reveals that the TA ranges from 60% to 80% for RF, whereas the TSF-based solutions present a significant decrease in performance (TA) when training data is below 50%.

4.3 | Geographic transferability

Here we present classification outcomes for the same testing area in the Kunene Region as in Section 4.2. However, the training process is conducted in a different gully situated within the Krumhuk Farm (Figure 10 and Table 6).

As observed in Figure 10c–e and Table 6, classes denoting change (C1 and C4) are not geographically transferred in an efficient manner due to the high level of false positives and true negative errors in the classification, with values of UA and PA below 50%, both for TSF and RF (excluding PA for C4). Grouping C1, C3 and C4 into a 'Gully area' class to work on a binary classification approach (with C2 as non-gully) demonstrates that our approach can be used to identify the presence of gullies even when training the classifier in remote and different gully areas. In this case, TSF (σ_{oVV}) presents the best accuracy results, with a TA = 80% and Kappa value close to 0.5.

4.4 | Type of change class description

In order to investigate error sources, the origin of False Positives (FP) and True Negatives (TN) conducted in the TSF (σ_{oVV}) classification for the Kunene Region (Figure 9b), we examine the points-

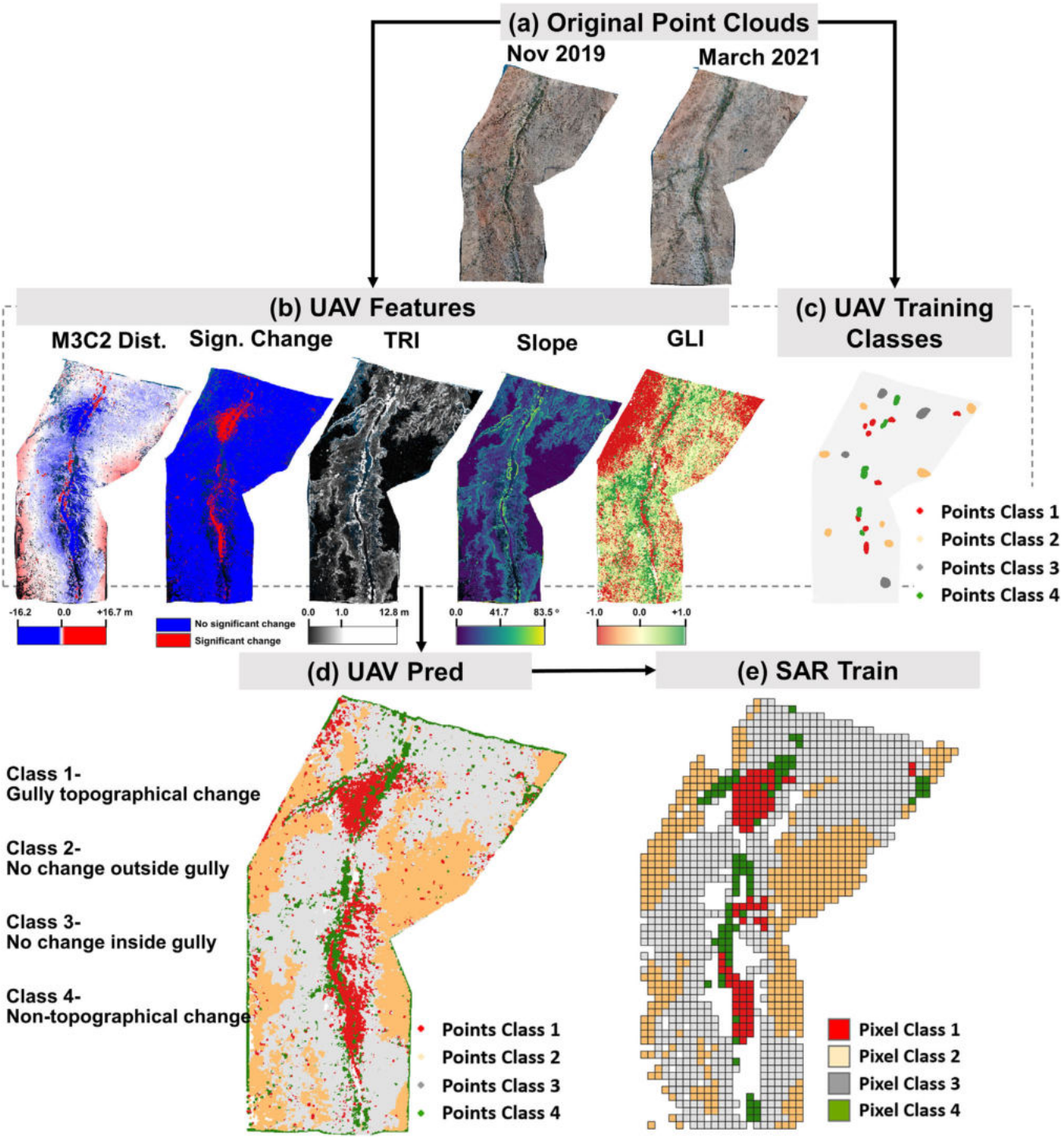

FIGURE 8 (a) Original point clouds for November 2019 and March 2021 in the Kunene region. (b) UAV features generated for the original point clouds. (c) Training classes segmented manually from the point clouds. (d) Result of the random Forest classification on the whole point cloud generating continuous type of gully change information. (e) Training data at satellite pixels resolution. [Color figure can be viewed at wileyonlinelibrary.com]

inside-pixels statistics generated during the data generalisation and classification process. To this end, a comparative analysis is performed at point cloud level to quantify the degree of mixture of each class and the basic statistics of each explanatory variable. Table 7 presents each class statistics for the initial training classes defined manually in the 3D point clouds (UAV Training), the RF results at the point cloud level (UAV Pred), the training classes generated at the S1 pixel level (SAR Train), the final prediction derived from TSF (σ_{oVV}) dataset (TSF SAR Pred), and these results once pixels with > 50% probability of the resultant class are selected (TSF SAR Pred > 50%).

C1 (gully topographical change) is characterised by a negative *M3C2* and high significant change (*SC* > 0.9), *TRI* and *S* above average, and with the absence of vegetation as indicated by low *GLI*. It is also understood that the first two parameters (*SC* and *M3C2*) serve to differentiate C1 (gully topographical change) from C3 (no change inside gully).

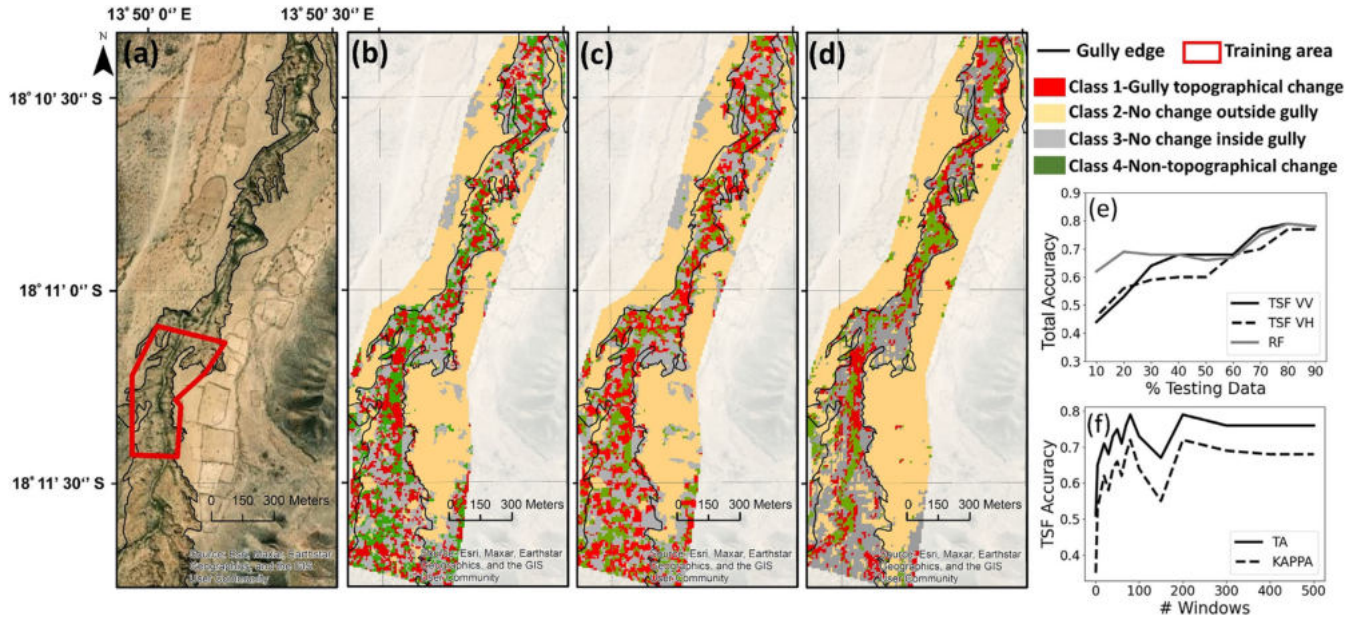

FIGURE 9 (a) Study area in the Kunene region with a red polygon demarcating the training area. Map of classification results for time series Forest-TSF applied on (b) σ_{oVV} and (c) σ_{oVH} time series. (d) Map of classification results for random Forest-RF on temporally aggregated features. (e) Comparison of TA for TSF (σ_{oVV}), TSF (σ_{oVH}) and RF performance for different training/testing data proportions. (f) TSF accuracy measures for different number of moving windows. Base maps source is ESRI, Maxar earthstar Geographics. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Accuracy results of the random Forest classification at point cloud level, as presented graphically in Figure 8d. User accuracy (UA%) and producer accuracy (PA%) are reported for each class (C1, C2, C3 and C4), and Total accuracy (TA%) and Cohen's kappa values are included as overall accuracy measures.

	C1		C2		C3		C4			
	UA	PA	UA	PA	UA	PA	UA	PA	TA	KAPPA
UAV Pred	99.4	98.7	99.9	99.6	98.3	99.3	99.2	99.2	99.2	0.98

TABLE 5 Accuracy results for the classifications conducted using time series Forest (σ_{oVV} and σ_{oVH}) and random Forest on temporally aggregated features. User accuracy (UA%) and producer accuracy (PA%) are reported for each class (C1 [gully topographical change], C2 [no change outside gully], C3 [no change inside gully] and C4 [non-topographical change]). Total accuracy (TA%) and Cohen's kappa values are included as overall accuracy measures.

	C1		C2		C3		C4			
	UA	PA	UA	PA	UA	PA	UA	PA	TA	KAPPA
TSF (σ_{oVV})	84.6	78.5	81.6	92.8	84.2	72.7	66.6	76.9	79.3	0.72
TSF (σ_{oVH})	81.2	81.2	93.7	78.9	72.2	76.4	58.3	70.0	77.4	0.69
RF	81.0	56.3	100.0	90.1	69.3	91.0	68.2	81.0	80.1	0.73

C2 (no change outside gully), located outside the gully zone, is essentially characterised by the absence of any physical change (with *M3C2* close to 0), and low *S*, *GLI* and *TRI* values.

Lastly, C4 (non-topographical change) presents positive change (*M3C2* > 2 m), but relatively insignificant (*SC* ≈ 0.5) and highly spatially variable ($\sigma_{M3C2} \approx 3.8$). These changes in C4 are derived from vegetation growth, as evidenced by a GLI value that exceeds that of those registered for other classes, as well as high *TRI* and *S* values resulting from the great elevation heterogeneity caused by dense woody vegetation.

Analysing the central and dispersion values change for each class following each data transfer can reveal how points-inside-pixels statistics evolve. To examine these variations, a MANOVA test is computed for each data transformation:

1. UAV Training → UAV Pred;
2. UAV Pred → SAR Train;
3. SAR Train → TSF SAR Pred;
4. TSF SAR Pred → TSF SAR Pred > 0.5.

This involves comparing the feature values of points (derived from the point clouds) categorised under each class after each transformation to the original manually segmented classes (UAV Training). These feature value points serve as the reference for MANOVA Null Hypothesis. The higher the F_{MANOVA} value, the larger the change in the multivariate dataset.

In Figure 11b, F_{MANOVA} test statistic presents a peak in the transition between SAR Train to TSF SAR Pred (i.e., SAR Train → TSF SAR Pred) for C1 and C4, showing that the largest data modification occurs in this stage, but with the F_{MANOVA} value stable in the two the previous transformations (i.e., UAV Training → UAV Pred and UAV Pred → SAR Train). With simple filtering of the prediction value (TSF_{pred} > 0.5) of the dominant class, however, the F_{MANOVA} test statistic returns to values similar to the initial transformation (UAV Training → UAV Pred), or even lower, as observed in C1 (Figure 11b). Observing the density functions in Figure 11a, it is evident that the *GLI*, *TRI* and *SC* variables exhibit minimal evolution in C2 and C3. This is indicated by the density curves of these variables, which remain very similar for the different data stages (UAV Training, UAV Pred, SAR Train, TSF SAR Pred, TSF SAR Pred > 0.5). In contrast, the C1 experiments large transformations in *M3C2* and *SC* variables, and to a lesser extent *TRI* and *S*. The C4 is the most affected as all features present statistical changes, as confirmed by the MANOVA test.

Table 8 compares the C1's True Positives (TP) with False Negatives (FN) and True Negatives (TN), as the C1 is of particular interest to us. First, while the well-classified pixels (TP) contain 84.2% points of C1 from the point clouds, this percentage decreases to 54.3% in the TN, and only 5.3% in the FP pixels, with C3 points contaminating C1 predominantly (with 85%). This is explained by the small proportion of topographically changing zones (C1 in the point cloud) in relation to the stable gully zones (C3 in the point cloud). Hence, the changing erosive processes occur at a very localised scale, mostly surrounded by non-changing points, but still inside the gully (C3 points) area.

This fact can be extrapolated for the other classes, as illustrated and confirmed by the histograms and photos included in Figure 12.

As presented in Figure 12a–d, most errors in the classification (FP + TN) are derived from pixels with a high degree of mixture among multiple classes (at point level) in their composition, whereas the purest pixels are generally well classified (TP), except for C3. These types of errors are frequent and somehow expected, since the study area is heterogeneous and the natural forms that represent the four predefined classes are spatially very mixed, coexisting on the spatial scale of a S1 pixel. This is especially evident for C1, C3 and C4 inside the gully borders. This fact can be clearly observed in the photos of Figure 12. Figure 12e represents the transition zone between C2 and C3. Figure 12f is an active gully zone (C1) with sparse vegetation (C4). Figure 12g shows active gully site (C1) close to zones without change outside the gully (C2), and Figure 12h presents the main channel with erosive activity (C1), surrounded by slopes with little change and vegetation (C4). Hence, when the level of mixture is reduced and a specific change dominates, the likelihood of accurately predicting this type of change increases.

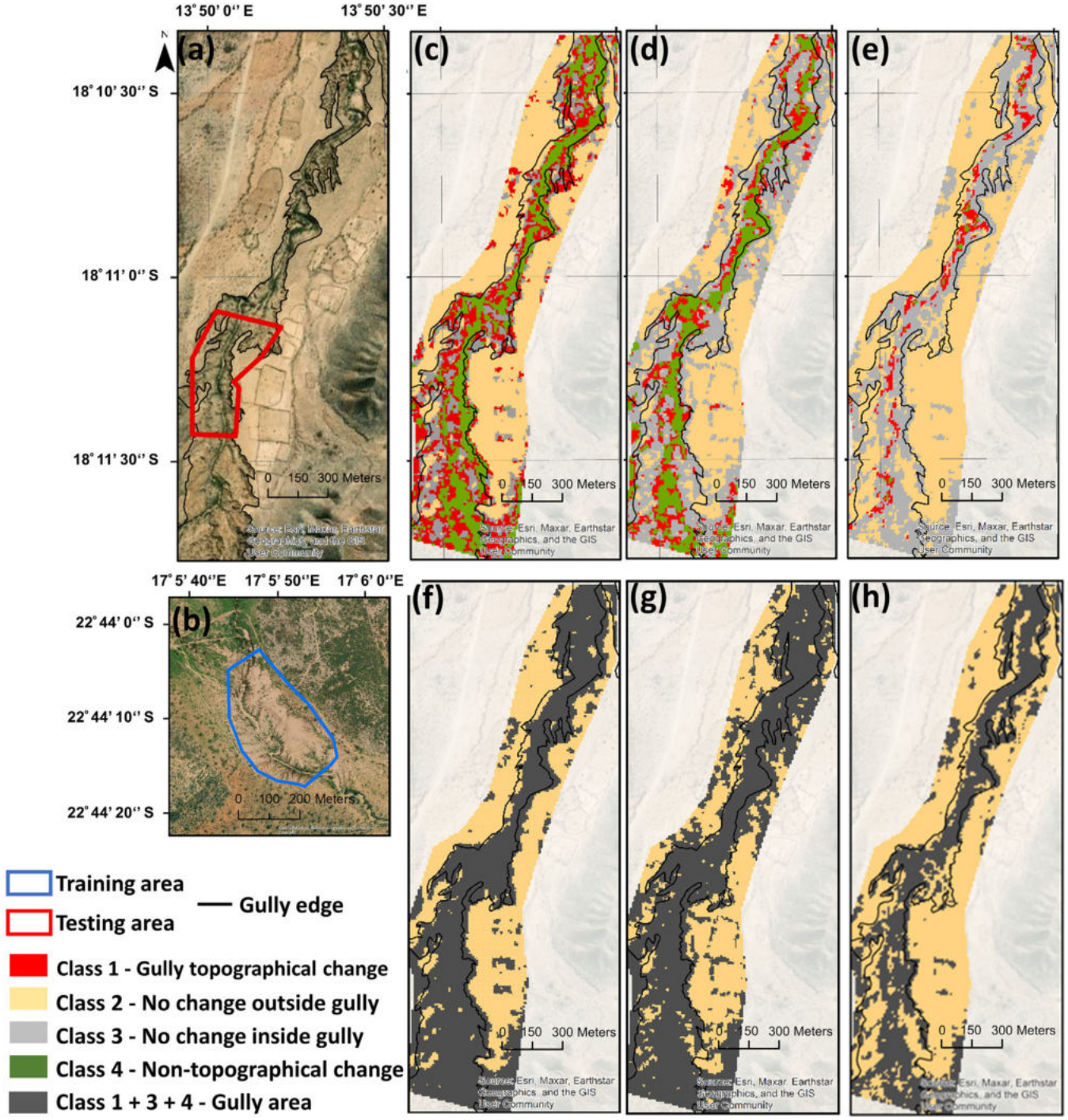

FIGURE 10 (a) Gully site in the Kunene region used for testing (red polygon) when the training was conducted in Krumhuk farm. (b) Gully site in the Krumhuk used for training (blue polygon). Maps of classification results of four classes separating training and testing areas for (c) time series Forest-TSF (σ_{oVV}), (d) TSF (σ_{oVH}) and (e) random Forest-RF. Maps of binary classification results (gully affected and no gully affected) separating training and testing areas for (f) TSF (σ_{oVV}), (g) TSF (σ_{oVH}) and (h) RF. The aggregation of classes 1, 3 and 4 is displayed with dark grey colour. Base maps source is ESRI, Maxar earthstar Geographics. [Color figure can be viewed at wileyonlinelibrary.com]

5 | DISCUSSION AND LIMITATIONS

5.1 | Discussion

In order to explore the potential of machine learning coupled with time series to characterise gully dynamics, we compared satellite pixel-based classifications of TSF on S1-σ_o time series with RF on 11 temporally aggregated and one-time features (derived from S1, S2 and TX).

5.1.1 | Radar time series versus aggregated features

Classification accuracy shows TA and Kappa around 80% and 0.7, respectively, for both the TSF classification (σ_{oVV} and σ_{oVH}) and RF. UA and PA to classify gully topographic changes (C1) are observed to be more accurate working with TSF than RF. However, the latter is less sensitive to modifications in the proportion of training to testing data. These results represent an advance in the zonal characterisation and monitoring of gullies (Vanmaercke et al., 2021), surpassing the

TABLE 6 Accuracy results (UA, PA and TA in %) for the classifications conducted using TSF (σ_{oVV} and σ_{oVH}), and RF on temporally aggregated features with training data generated in Krumhuk farm, and the testing in the Kunene region. The above block of the table presents the results for the four classes (C1 [gully topographical change], C2 [no change outside gully], C3 [no change inside gully] and C4 [non-topographical change]), while the below block presents the results aggregating classes 1,3, and 4 (C1 + C3 + C4) to work in a binary (gully and non-gully) classification.

	C1		C2		C3		C4			
	UA	PA	UA	PA	UA	PA	UA	PA	TA	KAPPA
TSF (σ_{oVV})	33.5	10.7	48.8	82	30.5	63.3	26.6	4.4	36	0.15
TSF (σ_{oVH})	22.2	10.7	33.8	79.2	47.2	54.8	48.5	10.7	40.7	0.11
RF	27.2	27.2	74.5	48.4	56	69.6	1.2	100	56.6	0.26
			C2			C1 + C3 + C4				
			UA	PA		UA	PA		TA	KAPPA
TSF (σ_{oVV})			48.8	82.0		95.3	81.1		81.3	0.49
TSF (σ_{oVH})			33.8	79.2		96.0	76.4		76.8	0.35
RF			74.5	48.4		63.8	84.6		67.2	0.31

mapping of gullies at a specific time using RF (Vallejo-Orti et al., 2021) and adding value to the quantification of changes to the gully perimeter in large periods (Shruthi et al., 2015).

Additionally, our method allows the estimation of terrain change volume by utilising the average vertical topographic deviation per pixel (derived from the M3C2 distance in their enclosed points), as reported in Table 7. This facilitates extrapolations to large areas. For instance, in Figure 9a, there are 2084 pixels classified as C1 in the entire target classification area (approximately 140 ha), with an average topographic change of −0.2 m per pixel. This results in a volume of 21.6 m^3 per pixel and a total of 45,014 m^3 for the entire study area.

Our method, based on remote measurements, applied annually can also be used to evaluate and calibrate empirical gully evolution (Sidorchuk, 2021) and simulation models (Omran et al., 2022) involving water streams. Thus, erosion rates predicted by these models can be assessed at the catchment and sub-catchment scale, facilitating the calibration of coefficients and sensitive parameters such as rainfall, stream velocity and discharge.

5.1.2 | Geographical transferability

As observed, σ_o time series provide a more accurate representation of the dynamics inside and outside the gully, compared to temporally aggregated features. This is particularly relevant when the gullies used for training and validation differ in their location, vegetation structure and rainfall distribution. To separate gully and non-gully zones, even when training and classification are conducted in geographically distinct areas, the use of time series proves to be an effective method. In this instance, TSF applied on σ_{oVV} overperforms TSF applied to σ_{oVH} and RF.

These results are in line with previous research on storm-derived change detection, whith σ_{oVV} identified as the most sensitive polarisation for monitoring land degradation in sparsely vegetated zones (Cerbelaud et al., 2021). Additionally, working on time series, there is room for improvement in exploring the combination of both polarisation signatures (σ_{oVV}, σ_{oVH}) applying multivariable time series algorithm (Ruiz et al., 2021) or generating cross-polarised time series for soil moisture studies (Ouellette et al., 2017).

5.1.3 | Spatial resampling from point clouds to satellite

As expected, pixels with less mixture tend to be more accurately classified. This is due to the large coexistence of different dynamics primarily within the gully borders (i.e., topographic changes, vegetation changes and stable sites), which is the primary source of classification error. This fact indicates that we arrived at the limit of the spatial resolution of S1 remote sensing for small gully change patterns, suggesting the applicability of this study to large gully changes.

Our results also indicate that in our target classification areas, topographic changes (Class 1) are closely linked to the presence of green vegetation (Class 4). This is because ephemeral water flows are associated with soil wetness that favours permanent vegetation, while simultaneously acting as the agent of gully erosion.

The availability and usage of SAR multitemporal imagery with higher spatial resolution (i.e., 1 m) would lead to significant uncertainty reduction in the classes' generalisation and resampling. To ensure better results, working with those areas where dynamics are in their purest states is critical to discriminate pixels where dynamics coexist. In this regard, filtering the result in the classification pixels with >50% probability leads to significant improvement in the classifications. Due to the correlation between pixel class probability, MANOVA test statistic (F_{MANOVA}) and accuracy, the MANOVA test applied at pixel level can also reduce error propagation.

5.2 | Limitations

Main limitations of this study arise from the utilisation of diverse datasets with differing spatial resolutions and horizontal alignment. Consequently, a certain level of uncertainty must be acknowledged due to the geolocation accuracy of the generated point clouds and S1. The absolute horizontal geolocation error is approximately 0.5 m for the computed point clouds and 1 m to 3 for S1 (Small & Schubert, 2022). Thus, sufficient spatial overlap for the given spatial pixel resolution of 10 × 10 m is guaranteed for our spatial analysis and it is also sufficient for time series analysis (Small & Schubert, 2022).

TABLE 7 Summary statistics for each dataset at the different stages of the methodology (UAV training, UAV Pred, SAR train, TSF SAR Pred, TSF SAR Pred >50%). The labels P1–P4 indicate the number of points (from point clouds) of each class (C1–C4). The last five columns present the mean (μ) and the standard deviation (σ) of each point cloud original feature *M3C2*, *SC*, *TRI*, *S* and *GLI*.

		P1	P2	P3	P4		*M3C2*	*SC*	*TRI*	*S*	*GLI*
UAV Training	**C1**	3287				μ	−0.913	0.921	0.314	17.902	−0.014
						σ	1.460	0.268	0.320	14.371	0.027
	C2		7347			μ	0.276	0.001	0.032	2.032	−0.014
						σ	0.158	0.016	0.024	1.513	0.008
	C3			7855		μ	−0.050	0.008	0.350	21.35	−0.004
						σ	0.430	0.089	0.212	10.976	0.013
	C4				6290	μ	2.008	0.537	1.683	49.593	0.126
						σ	3.792	0.498	1.239	20.641	0.102
UAV Pred	**C1**	48,345				μ	−0.920	0.807	0.346	21.239	−0.014
						σ	0.848	0.394	0.209	11.699	0.022
	C2		204,503			μ	0.212	0.000	0.034	2.174	−0.012
						σ	0.172	0.000	0.016	1.140	0.009
	C3			372,727		μ	−0.052	0.000	0.370	22.857	−0.005
						σ	0.155	0.002	0.193	10.747	0.008
	C4				50,324	μ	0.966	0.326	2.011	56.442	0.066
						σ	2.652	0.468	1.301	18.219	0.089
SAR Train	**C1**	31,592	10	4026	3451	μ	−0.728	0.747	0.455	24.045	−0.003
						σ	1.008	0.434	0.523	14.880	0.037
	C2	1296	178,468	11,314	1764	μ	0.196	0.007	0.076	3.620	−0.011
						σ	0.311	0.082	0.345	7.956	0.011
	C3	6793	13,363	329,307	11,451	μ	−0.039	0.020	0.410	23.305	−0.004
						σ	0.433	0.140	0.419	12.808	0.017
	C4	4340	499	6064	26,597	μ	0.831	0.377	1.523	47.197	0.055
						σ	2.799	0.484	1.186	22.156	0.095
TSF SAR Pred	**C1**	33,159	13,179	91,658	8337	μ	−0.216	0.210	0.431	23.147	−0.004
						σ	0.745	0.407	0.508	14.973	0.026
	C2	2846	156,009	45,688	3961	μ	0.188	0.009	0.144	7.194	−0.011
						σ	0.319	0.097	0.408	12.447	0.012
	C3	5482	29,802	156,943	6888	μ	−0.011	0.026	0.379	20.690	−0.004
						σ	0.500	0.160	0.514	14.286	0.020
	C4	8168	18,043	87,986	32,123	μ	0.186	0.119	0.667	27.336	0.012
						σ	1.510	0.324	0.889	20.247	0.057
TSF SAR Pred >50%	**C1**	29,718	1366	20,159	3281	μ	−0.525	0.495	0.431	23.754	−0.004
						σ	0.820	0.499	0.471	14.158	0.029
	C2	1270	126,678	25,289	2263	μ	0.206	0.009	0.125	6.300	−0.012
						σ	0.325	0.098	0.378	11.552	0.012
	C3	2952	7914	79,184	2593	μ	−0.045	0.026	0.378	21.178	−0.003
						σ	0.434	0.161	0.487	13.209	0.015
	C4	480	2493	23,100	24,345	μ	0.473	0.235	1.074	37.203	0.033
						σ	2.197	0.424	1.112	22.549	0.080

As for the point clouds *LoD*, which is approximately 0.7 m in average, it enables the possibility to detect changes larger than 70 cm, while statistically differentiating them from smaller topographical changes below this threshold. Therefore, the type of topographic changes targeted in this study are mainly those derived from large detachments and linear concentrated flows, while laminar erosion and other micro topography modifications produced by animals' paths are neglected.

Additionally, expert-based manual labelling of the point cloud change types is challenging and time-consuming. Therefore, one of the main limitations of this study is the availability of continuous wall-to-wall training data at point cloud level. The adopted solution using RF to apply classes to the whole point cloud achieved extremely high accuracy values, which should be interpreted with caution, as the geographical transfer of the classifier cannot be guaranteed. For future applications, locally trained classifiers may be replaced by more

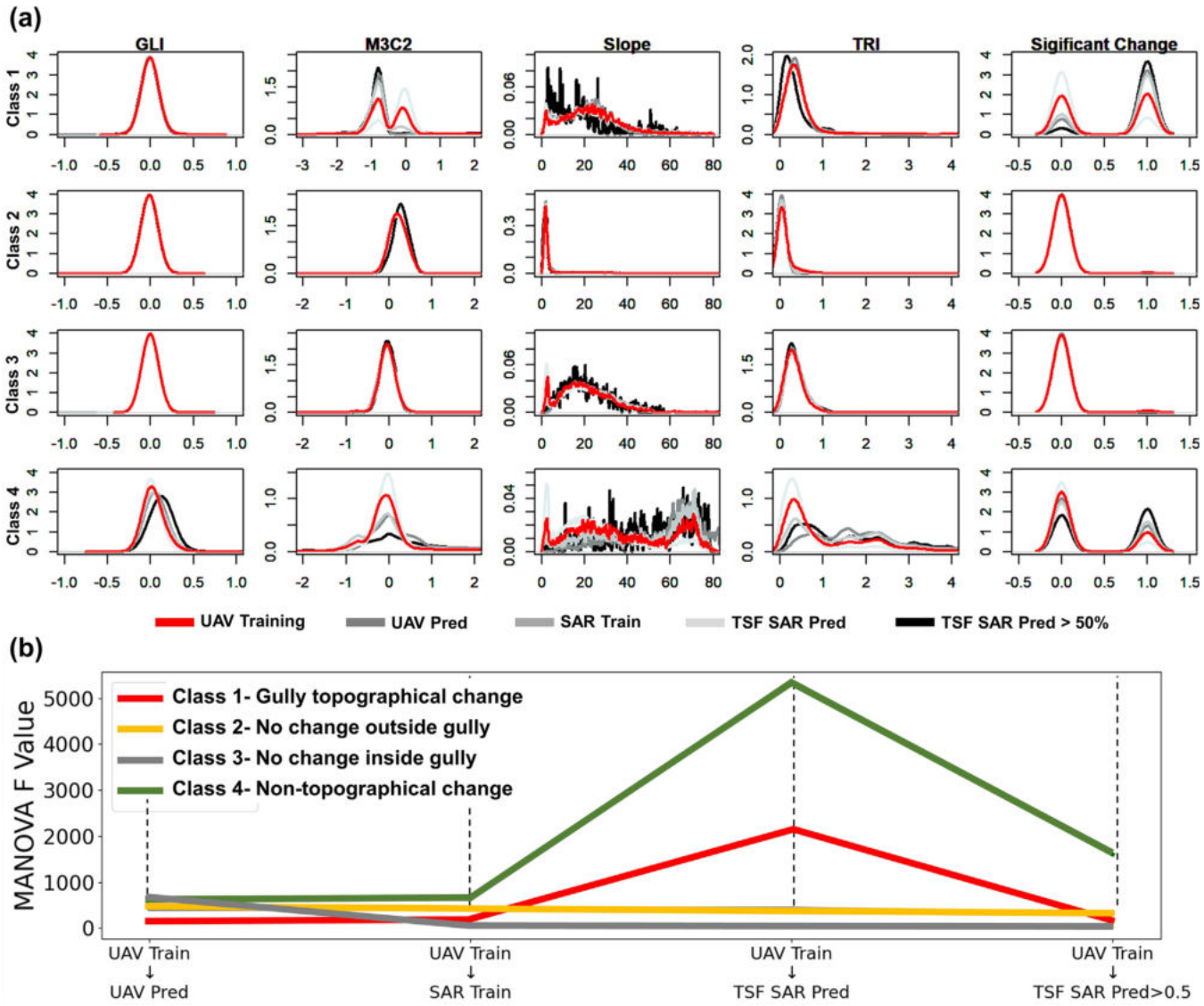

FIGURE 11 (a) Density curves of explanatory variables separated by stage of the dataset organised per class. Each subplot includes multiple density curves for UAV training (black line), UAV Pred (dark grey line), SAR train (grey line), TSF SAR Pred (light grey line) and TSF SAR Pred > 0.5 (red line) of each point cloud explanatory feature (*GLI*, *M3C2*, *S*, *TRI* and *SC*) organised per columns. (b) Values of MANOVA *F* test statistic to quantify the similarity between each transformed dataset (UAV Pred, SAR train, TSF SAR Pred and TSF SAR Pred > 0.5) and the originally segmented classes (UAV training). [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 8 Summary of the points-inside-pixels statistics for the different classification results for the Class 1, comparing the true positive (TP) with true negative (TN) pixels and the confusion between classes 1 and 3 (C1-C3). The columns points and points % present the count and the percentage of points from the interpolated UAV Pred contained by each Class 1 training pixel (px.).

	TP (74 px.)		FP (47 px.)		TN (1 px.)		Confusion C1–C3 (45 px.)	
	Points	Points %	Points	Points %	Points	Points %	Points	Points %
C1	28,200	84.2	998	5.3	209	54.3	995	5.5
C2	0	0.0	1,206	6.4	0	0.0	400	2.2
C3	2840	8.5	16,085	85.3	77	20.0	16,019	89.1
C4	2444	7.3	567	3.0	99	25.7	566	3.1

flexible and generalised methods reducing local overfitting (Yu et al., 2021). Additionally, these methods can address potential underfitting problems derived from classical machine learning when they are used for generalisation. This would save time and enable scaling of our approach, such as potentially for global applications (Vanmaercke et al., 2021). Micro-mapping collaborative approaches (Herfort et al., 2018) can help to create reliable segmented classes and reduce uncertainties and issues related to class generalisation.

As previously outlined, our approach can classify gullies when the training and testing datasets are geographically distant. However, it

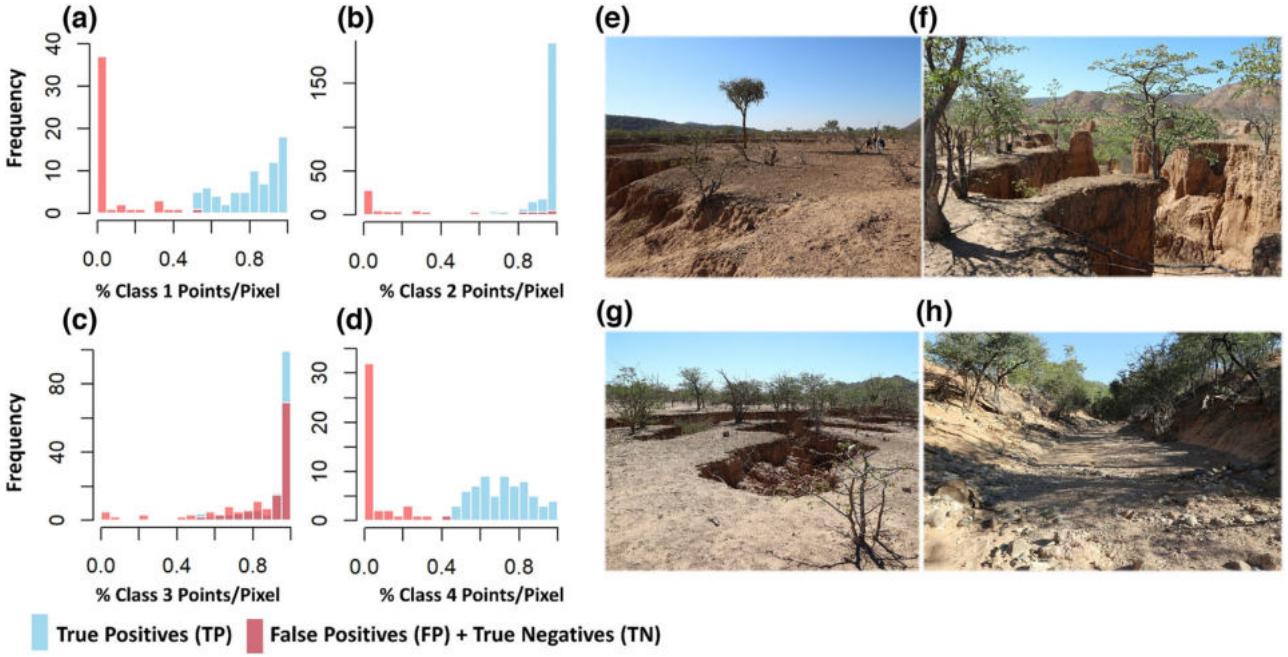

FIGURE 12 (a–d) Histograms organised per classes displaying in the vertical axis the count of true positive (TP) classified pixels (in blue) and false negatives plus true negatives (FN + TN) pixels (in pink) versus the percentage of points per pixel of the corresponding class in the horizontal axis. Photos showing transition zones between (e) Class 2 and Class 3, (f) Class 1 and Class 4, (g) Class 1 and Class 2 and (h) Classes 1, 3 and 4 to illustrate typical mixture among classes leading to lower classification accuracies at pixel level. [Color figure can be viewed at wileyonlinelibrary.com]

encounters challenges in accurately classifying gully topographic changes, possibly due to the different temporal rainfall distributions, which serve as the main driver of terrain changes. To ensure similar accuracies (i.e., TA = 80%) for the classification of change types, it is recommended to restrict the classification to the same region where the training is conducted, thus erosive rain events would be more aligned in time. This issue could also be resolved by generating TSF features based only on changing periods and correlating them to rain events.

Regarding the geographical transferability of the methodology, it is assumed that similar classification results can be achieved when the training and classification is conducted on permanent large valley bottom gullies in (semi-) arid areas. In contrast, urban gullies, or ephemeral agricultural gullies would fall outside the scope of this work.

6 | CONCLUSION

This research explores the combination between close-range UAV and spaceborne remote sensing as a novel application for classifying gully activity types. This study contributes to addressing the existing research gap in gully monitoring and characterisation over large areas. To date, few approaches have attempted to monitor gully changes with spaceborne sensors. This is primarily due to the significant amount of manual labour required to generate ground reference data for temporal changes, particularly on large scales (Vanmaercke et al., 2021). In order to characterise the dynamics of gullies, we collected and processed 3D multitemporal point clouds to generate training data. The data was transferred from point clouds to spaceborne remote sensing images pixels entailing a necessary generalisation resampling from a sub-m to 10-m spatial resolution. As an innovation in gully research, erosion processes and geomorphological features are studied by means of time series of radar data, introducing the use of a time series classifier, Time Series Forest (TSF), to separate types of changes.

Employing this method, TA > 80% are achieved to classify types of gully changes when the training is conducted within the geographic region of the target classification. This method also serves to differentiate between gully and no gully zones when the training data is conducted in a different geographical setting.

However, it is observed that the resolution mismatch between the very high-resolution point clouds with Sentinel-1, and temporal misalignment of gully erosion events in different areas are two issues limiting the accuracy of the results and the geographic transferability of the training data.

In general, our approach represents an improvement in terms of the type of information that can be extracted from gullies, complementing the line of separating affected from unaffected areas in a binary mapping task. In this regard, annual systematic application of our method in large gully affected catchment areas can serve as a decision-making tool for the planning and assessment of gully restoration campaigns at a large scale (regional level). For instance, to quantify the effectiveness of soil-fixing plantations, or temporary zonal restrictions on cattle.

Finally, this paper suggests three directions for extending our approach to characterise large gully networks in sparsely populated areas in the future: (i) expanding the UAV data collection campaigns to other areas in order to investigate the relation between gully characteristics, local settings and geographic transferability of the training data, (ii) applying additional tests to optimise the aggregation from UAV points to satellite pixels and (iii) explore multivariate time series analysis to combine different radar polarisation and optical derivates as proxies for gully dynamics.

ACKNOWLEDGEMENTS

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. There are no relevant financial or non-financial competing interests to report.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author [M. V] upon reasonable request.

ORCID

Miguel Vallejo https://orcid.org/0000-0002-8464-772X
Carlos Castillo https://orcid.org/0000-0001-5848-0332

REFERENCES

Agisoft Metashape (2021). (https://www.agisoft.com/). Version 1.8.1. 2021.

Ahmad, W. & Kim, D. (2019) Estimation of flow in various sizes of streams using the Sentinel-1 synthetic aperture radar (SAR) data in Han River basin, Korea. *International Journal of Applied Earth Observation and Geoinformation*, 83, 101930. Available from: https://doi.org/10.1016/j.jag.2019.101930

Bennett, S.J. & Wells, R.R. (2019) Gully erosion processes, disciplinary fragmentation, and technological innovation. *Earth Surface Processes and Landforms*, 44(1), 46–53. Available from: https://doi.org/10.1002/esp.4522

Besl, P.J. & McKay, N.D. (1992) A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256. Available from: https://doi.org/10.1109/34.121791

Canova, F., Tolomei, C., Salvi, S., Toscani, G. & Seno, S. (2012) Land subsidence along the Ionian coast of SE Sicily (Italy), detection and analysis via Small baseline subset (SBAS) multitemporal differential SAR interferometry. *Earth Surface Processes and Landforms*, 37(3), 273–286. Available from: https://doi.org/10.1002/esp.2238

Castillo, C. & Gómez, J.A. (2016) A century of gully erosion research: urgency, complexity and study approaches. *Earth-Science Reviews*, 160, 300–319. Available from: https://doi.org/10.1016/j.earscirev.2016.07.009

Cerbelaud, A., Roupioz, L., Blanchet, G., Breil, P. & Briottet, X. (2021) A repeatable change detection approach to map extreme storm-related damages caused by intense surface runoff based on optical and SAR remote sensing: evidence from three case studies in the south of France. *ISPRS Journal of Photogrammetry and Remote Sensing*, 182, 153–175. Available from: https://doi.org/10.1016/j.isprsjprs.2021.10.013

Cloud Compare. (https://www.danielgm.net/cc/). Version 2.11.2. 2022.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. Available from: https://doi.org/10.1177/001316446002000104 hdl:1942/28116. S2CID 15926286.

Congalton, R. (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. Available from: https://doi.org/10.1016/0034-4257(91)90048-B

Deng, H., Runger, G., Tuv, E. & Vladimir, M. (2013) A time series forest for classification and feature extraction. *Inf. Sci. (Ny).*, 239, 142–153. Available from: https://doi.org/10.1016/j.ins.2013.02.030

DiFrancesco, P.M., Bonneau, D. & Hutchinson, D.J. (2020) The implications of M3C2 projection diameter on 3D semi-automated Rockfall extraction from sequential terrestrial laser scanning point clouds. *Remote Sensing*, 12(11), 1885. Available from: https://doi.org/10.3390/rs12111885

DJI. (2022). *Mavic pro - product information*. https://www.dji.com/mavic/info (accessed 10.21.22).

DLR (Deutsches Zentrum für Luft- und Raumfahrt). (2014). *TanDEM-X ground segment: announcement of opportunity, Oberpfaffenhofen*. Germany: DLR (Deutsches Zentrum für Luft- und Raumfahrt), Rep. TD-PDPL-0032, May.

Du, Z., Zheng, G., Shen, G. & Moskal, L.M. (2021) Characterizing spatio-temporal variations of forest canopy gaps using aerial laser scanning data. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102588. Available from: https://doi.org/10.1016/j.jag.2021.102588

ESA. (2021a). *Copernicus Sentinel 1 data, (2019–2021)*. Retrieved from: ASF DAAC (https://search.asf.alaska.edu/), processed by ESA.

ESA. (2021b). *Copernicus Sentinel 2 data, (2019–2021)*. Retrieved from: Copernicus Open Access Hub [https://scihub.copernicus.eu/], processed by ESA.

ESA. (2021c). *SNAP* (https://step.esa.int/main/download/snap-download/). Version 8.0. 2021.

Gisinger, C., Schubert, A., Breit, H., Garthwaite, M., Balss, U., Willberg, M., et al. (2020) In-depth verification of Sentinel-1 and TerraSAR-X geolocation accuracy using the Australian corner reflector Array. *IEEE Transactions on Geoscience and Remote Sensing*, 59, 2(2), 1154–1181. Available from: https://doi.org/10.1109/TGRS.2019.2961248

Golosov, V., Yermolaev, O., Rysin, I., Vanmaercke, M., Medvedeva, R. & Zaytseva, M. (2018) Mapping and spatial-temporal assessment of gully density in the middle Volga region, Russia. *Earth Surface Processes and Landforms*, 43(13), 2818–2834. Available from: https://doi.org/10.1002/esp.4435

Guo, M., Wang, W., Wang, T., Wang, W. & Kang, H. (2020) Impacts of different vegetation restoration options on gully head soil resistance and soil erosion in loess tablelands. *Earth Surface Processes and Landforms*, 45(4), 1038–1050. Available from: https://doi.org/10.1002/esp.4798

Hancock, G.R. & Evans, K.G. (2010) Gully, channel and hillslope erosion - an assessment for a traditionally managed catchment. *Earth Surface Processes and Landforms*, 35(12), 1468–1479. Available from: https://doi.org/10.1002/esp.2043

Hayas, A., Vanwalleghem, T., Laguna, A., Penã, A. & Giráldez, J.V. (2017) Reconstructing long-term gully dynamics in Mediterranean agricultural areas. *Hydrology and Earth System Sciences*, 21(1), 235–249. Available from: https://doi.org/10.5194/hess-21-235-2017

Herfort, B., Höfle, B. & Klonner, C. (2018) ISPRS journal of photogrammetry and remote sensing 3D micro-mapping: towards assessing the quality of crowdsourcing to support 3D point cloud analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 137, 73–83. Available from: https://doi.org/10.1016/j.isprsjprs.2018.01.009

Ho, T.K. (1995) *Random decision forests*, Vol. 1. Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, pp. 278–282. Montreal, QC, Canada.

Jones, A., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Dewitte, O., et al. (2013) *Soil atlas of Africa*. Luxembourg: Publications Office of the European Union.

Kimura, S.D., Hatano, R. & Okazaki, M. (2009) Characteristics and issues related to regional-scale modeling of nitrogen flows. *Soil Science & Plant Nutrition*, 55(1), 1–12. Available from: https://doi.org/10.1111/j.1747-0765.2008.00353.x

Lague, D., Brodu, N. & Leroux, J. (2013) Accurate 3D comparison of complex topography with terrestrial laser scanner: application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10–26. Available from: https://doi.org/10.1016/j.isprsjprs.2013.04.009

le Roux, J., Morake, L., van der Waal, B., Leigh Anderson, R. & Hedding, D.W. (2022) Intra-gully mapping of the largest documented gully network in South Africa using UAV photogrammetry: implications for restoration strategies. *Progress in Physical Geography*, 46(5), 772–789. Available from: https://doi.org/10.1177/03091333221101057

Lee, J.S., Grunes, M.R., Ainsworth, T., Du, L.J., Schuler, D. & Cloude, S. (1999) Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing 1999*, 37, 2249–2258. Available from: https://doi.org/10.1109/36.789621

Liu, X., Li, H., Zhang, S., Cruse, R.M. & Zhang, X. (2019) Gully erosion control practices in Northeast China: a review. *Sustain*, 11(18), 5065. Available from: https://doi.org/10.3390/su11185065

Louhaichi, M., Borman, M.M. & Johnson, D.E. (2001) Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16(1), 65–70. Available from: https://doi.org/10.1080/10106040108542184

Mendelsohn, J., Jarvis, A., Roberts, C. & Robertson, T. (2002) *The atlas of Namibia: a Potrait of the land and its people*. Cape Town: David Philip Publishers; New Africa Books (Pty) Ltd.

Montanarella, L. et al. (2015) Global soil status, processes and trends. In: *Status of the World's soil resources*, pp. 100–144, Food and Agriculture Organization of the United Nations, Rome. ISBN: 978-92-5-109004-6.

Morgan, R.P.C. (2005) *Soil erosion and conservation*, 3rd edition. Oxford: Blackwell Publishing.

Namibia Statistical Agency. (2021). (https://digitalnamibia.nsa.org.na/). *Digital Namibia*. Windhoek (Namibia): Namibia Statistical Agency. Access March 2021.

Nota, E.W., Nijland, W. & de Haas, T. (2022) Improving UAV-SfM time-series accuracy by co-alignment and contributions of ground control or RTK positioning. *International Journal of Applied Earth Observation and Geoinformation*, 109, 102772. Available from: https://doi.org/10.1016/j.jag.2022.102772

Olivier, G., Van De Wiel, M.J. & De Clercq, W.P. (2023) Intersecting views of gully erosion in South Africa. *Earth Surface Processes and Landforms*, 48(1), 119–142. Available from: https://doi.org/10.1002/esp.5525

Omran, A., Schröder, D., Sommer, C., Hochschild, V. & Märker, M. (2022) A GIS-based simulation and visualization tool for the assessment of gully erosion processes. *Journal of Spatial Science*, 1–18. Available from: https://doi.org/10.1080/14498596.2022.2133020

Ouellette, J.D., Johnson, J.T., Balenzano, A., Mattia, F., Satalino, G., Kim, S.B., et al. (2017) A time-series approach to estimating soil moisture from vegetated surfaces using L-band radar backscatter. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6), 3186–3193. Available from: https://doi.org/10.1109/TGRS.2017.2663768

Pu, C., Xu, Q., Zhao, K., Chen, W., Wang, X., Li, H., et al. (2022) Spatiotemporal evolution and surface response of land subsidence over a large-scale land creation area on the Chinese loess plateau. *International Journal of Applied Earth Observation and Geoinformation*, 111, 102835. Available from:: https://doi.org/10.1016/j.jag.2022.102835

R. (2021). (https://www.r-project.org/). Version 4.2.0. 2021

Riley, S.J., DeGloria, S.D. & Elliot, R. (1999) A terrain ruggedness index. *International Journal of Science*, 5, 23–27.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS (Earth Resources Technology Satellite). Proceedings of 3rd Earth Resources Technology Satellite Symposium, Greenbelt, 10-14 December, SP-351, 309–317.

Rozenstein, O., Siegal, Z., Blumberg, D.G. & Adamowski, J. (2016) Investigating the backscatter contrast anomaly in synthetic aperture radar (SAR) imagery of the dunes along the Israel–Egypt border. *International Journal of Applied Earth Observation and Geoinformation*, 46, 13–21. Available from: https://doi.org/10.1016/j.jag.2015.11.008

Ruiz, A.P., Flynn, M., Large, J., Middlehurst, M. & Bagnall, A. (2021) The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* Springer US, 35(2), 401–449. Available from: https://doi.org/10.1007/s10618-020-00727-3

Sentinel-1 SAR User Guide. (2021). [https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar], accessed February 2021.

Shahabi, H., Jarihani, B., Piralilou, S.T., Chittleborough, D., Avand, M. & Ghorbanzadeh, O. (2019) A semi-automated object-based gully networks detection using different machine learning models: a case study of bowen catchment, Queensland, Australia. *Sensors (Switzerland)*, 19(22), 1–21. Available from: https://doi.org/10.3390/s19224893

Shruthi, R.B.V., Kerle, N., Jetten, V., Abdellah, L. & Machmach, I. (2015) Quantifying temporal changes in gully erosion areas with object-oriented analysis. *Catena*, 128, 262–277. Available from: https://doi.org/10.1016/j.catena.2014.01.010

Sidorchuk, A. (2021) Models of gully erosion by water. *Watermark*, 13, 3293. Available from: https://doi.org/10.3390/w13223293

Small, D. & Schubert, A. (2022) *Guide to Sentinel-1 geocoding. Measurements/products/policy*. Zurich (Switzerland): University of Zurich 1.12, ref: UZH-S1-GC-AD.

Snapir, B., Momblanch, A., Jain, S.K., Waine, T.W. & Holman, I.P. (2019) A method for monthly mapping of wet and dry snow using Sentinel-1 and MODIS: application to a Himalayan river basin. *International Journal of Applied Earth Observation and Geoinformation*, 74, 222–230. Available from: https://doi.org/10.1016/j.jag.2018.09.011

Stahle, L. (1990) Multivariate analysis of variance (MANOVA). *Chemometrics and Intelligent Laboratory Systems*, 9(2), 127–141. Available from: https://doi.org/10.1016/0169-7439(90)80094-M

Story, M. & Congalton, R. (1986) Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3), 397–399.

Ullmann, T., Jagdhuber, T., Hoffmeister, D., May, S.M., Baumhauer, R. & Bubenzer, O. (2023) Exploring Sentinel-1 backscatter time series over the Atacama Desert (Chile) for seasonal dynamics of surface soil moisture. *Remote Sensing of Environment*, 285, 113413. Available from: https://doi.org/10.1016/j.rse.2022.113413

Vallejo-Orti, M., Winiwarter, L., Corral-Pazos-De-Provens, E., Williams, J.G., Bubenzer, O. & Höfle, B. (2021) Use of TanDEM-X and sentinel products to derive gully activity maps in Kunene region (Namibia) based on automatic iterative random Forest approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 14, 607–623. Available from: https://doi.org/10.1109/JSTARS.2020.3040284

Vanmaercke, M., Chen, Y., Haregeweyn, N., De Geeter, S., Campforts, B., Heyndrickx, W., et al. (2020) Predicting gully densities at sub-continental scales: a case study for the horn of Africa. *Earth Surface Processes and Landforms*, 45(15), 3763–3779. Available from: https://doi.org/10.1002/esp.4999

Vanmaercke, M., Panagos, P., Vanwalleghem, T., Hayas, A., Foerster, S., Borrelli, P., et al. (2021) Measuring, modelling and managing gully erosion at large scales: a state of the art. *Earth-Science Reviews*, 218, 103637. Available from: https://doi.org/10.1016/j.earscirev.2021.103637

Wearne, L., Telfer, D., Windridge, D. & Behzadnia, S. (2018) Innovative solutions to alluvial gully remediation: a case study from the great barrier reef catchments. *9th Aust. Stream Manag. Conf.*, 554–566.

Yu, H., Wu, Y., Niu, L., Chai, Y., Feng, Q., Wang, W., et al. (2021) A method to avoid spatial overfitting in estimation of grassland above-ground biomass on the Tibetan plateau. *Ecological Indicators*, 125, 107450. Available from: https://doi.org/10.1016/j.ecolind.2021.107450

Zahs, V., Winiwarter, L., Anders, K., Williams, J.G., Rutzinger, M. & Höfle, B. (2022) Correspondence-driven plane-based M3C2 for lower uncertainty in 3D topographic change quantification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, 541–559. Available from: https://doi.org/10.1016/j.isprsjprs.2021.11.018

Zhong, W., Zhang, T., Chen, J., Shang, J., Wang, S., Mu, C., et al. (2021) Seasonal deformation monitoring over thermokarst landforms using terrestrial laser scanning in northeastern Qinghai-Tibetan plateau. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102501. Available from: https://doi.org/10.1016/j.jag.2021.102501

How to cite this article: Vallejo, M., Castillo, C., Zahs, V., Bubenzer, O. & Höfle, B. (2024) Classifying types of gully changes with unoccupied aircraft vehicles 3D multitemporal point clouds for training of satellite data analysis in Northwest Namibia. *Earth Surface Processes and Landforms*, 49(3), 1135–1155. Available from: https://doi.org/10.1002/esp.5759