

# **Automated pattern recognition to support geological mapping and exploration target generation - A case study from southern Namibia**

Detlef Eberle<sup>1,2,7</sup>, David Hutchins<sup>3</sup>, Sonali Das<sup>4,5</sup>, Anandamayee Majumdar<sup>6</sup>, Hendrik Paasche<sup>2</sup>

<sup>1</sup> geotec Rohstoffe GmbH, Friedrichstr. 95, 10117 Berlin, Germany

<sup>2</sup> UFZ - Helmholtz Centre for Environmental Research, Department of Monitoring and Exploration Technologies, Permoserstr. 15, 04318 Leipzig, Germany

<sup>3</sup> Geological Survey of Namibia, 1 Aviation Rd., Windhoek, Namibia

<sup>4</sup> Spatial Planning and Systems, CSIR Built Environment, P.O.Box 395,

Pretoria 0001, South Africa

<sup>5</sup> Department of Statistics, Nelson Mandela Metropolitan University, P.O.Box 7700,

Port Elizabeth 6031, South Africa

<sup>6</sup> to be filled

<sup>7</sup> Corresponding author: telephone: +49 157 31573236; +264 81 7781480

fax: +49 341 235 1939

Email: [d.eberle@rohstoffe.org](mailto:d.eberle@rohstoffe.org)

Keywords: Cluster analysis; discriminant analysis; lineament detection; mapping; southern Namibia, Namaqua Metamorphic Belt.

Manuscript submitted for publication in *Journal of African Earth Science*, November, 2014

## **Abstract**

This paper demonstrates a methodology for the automatic joint interpretation of high resolution airborne geophysical and space-borne remote sensing data to support geologic mapping in a largely automated, fast and objective manner. At the request of the Geological Survey of Namibia (GSN), part of the Gordonias Subprovince of the Namaqua Metamorphic Complex situated in southern Namibia was selected for this study.

All data were gridded, with a spacing of adjacent data points of only 200 m. The data points were coincident for all data sets. Published criteria were used to characterize the airborne magnetic data and to establish a set of attributes suitable for the recognition of linear features and the tectonic pattern of the study area. This multi-attribute analysis of the airborne magnetic data provided the magnetic lineament pattern of the study area and a structural map.

To obtain a geological map, the high resolution airborne gamma-ray data were integrated with selected Landsat band data using unsupervised fuzzy partitioning clustering. The outcome of this unsupervised clustering is a classified (zonal) map which in terms of spatial resolution is superior to any geological mapping. To achieve a (pseudo-) geologic map, the classified zones are assigned geological/geophysical parameters as observed from the study area, e.g.

lithology, rock properties and age, geophysical attributes etc. This information is obtained from the examination of archived geological reports, borehole logs and any kind of existing geological/geophysical maps.

To assess the quality of unsupervised fuzzy clustering, stepwise linear discriminant analysis was used. Comparing the classification results obtained from clustering and discriminant analysis, only a small percentage (8%) of the samples was misclassified. Furthermore, a comparison of the aposterior probability of class assignment with the trustworthiness values provided by fuzzy clustering indicates only slight differences. These observed differences are

due to the exponential class probability term which tends to deliver either fairly high or low probability values.

The methodology and results presented here demonstrate that automated pattern recognition can support geological mapping and mineral exploration target generation of large study areas. This methodology is considered well suited to a number of African countries whose large territories have recently been covered by high resolution airborne geophysical data, but where geologic mapping is poor, incomplete or outdated.

## **1. Introduction**

The Geological Survey of Namibia (GSN), a directorate within the Ministry of Mines and Energy, embarked on a high resolution airborne geophysical campaign two decades ago with the aim of achieving complete national coverage of both magnetic and gamma-ray (radiometric) data. To date, apart from very small gaps, the country is completely covered by a number of high density airborne geophysical data suites. In addition, nearly complete coverage of Landsat 7 satellite imagery is existent.

The main objective of acquiring such data has been to support geologic mapping and stimulate mineral exploration in the country, add value by updating existing geological maps and to highlight areas with increased mineral potential. Conventional geological mapping using aerial photo and satellite image interpretation coupled with field visits is often handicapped by several factors, such as (a) dense vegetation or extensive sand cover, (b) extensive size of the study area; (c) time and budget constraints; (d) mapper's skills. A number of African countries with large surface areas, such as Mauretania, Nigeria or Namibia, have acquired high resolution airborne geophysical data sets in the past through national and international funding for support of geological mapping and mineral target area generation. Despite being in possession of complete airborne geophysical data sets, hardly any of these countries have complete geologic map coverage. This paper suggests a methodology to generate geologic maps from high resolution airborne geophysical data, integrated with satellite imagery or any other related data complete from the same area.

Our paper will show how automated linear feature detection and objective classification methods can accelerate the process of extracting information and evaluating the airborne geophysical and remote sensing data suites to provide geological maps. It is also apparent that in view of the size of the country coupled with the huge amount of existing airborne

geophysical and remote sensing data that a novel approach is required. Furthermore, high quality geophysical and remote sensing data can be recorded and processed rapidly using various moving platforms, whilst the process of interpreting and evaluating such data is by far not as fast as data acquisition. To overcome the ever increasing gap between data acquisition and the interpretation of these data, our paper demonstrates a recent methodology (Paasche and Eberle, 2009; Paasche and Eberle, 2011) to extract structural and lithology information in a largely automated and objective manner. At the end of this process, a representative geological map is obtained revealing various sets of tectonic elements as well as a number of lithological units identified by their diversified physical properties.

The method uses fuzzy clustering methods (e.g. Höppner et al., 1999) and allows integrating high resolution airborne geophysical data with other data sets, e.g. remote sensing, regional geochemistry or hyper-spectral imaging. Prior to clustering, all data sets need to be gridded uniformly so that a multivariate sample vector can be created at each grid point. Unsupervised fuzzy cluster analysis classifies the samples into a number of groups without prior training and assigns a fuzzy class membership to every sample. After clustering linear stepwise discriminant analysis is used to assess the performance and trustworthiness of fuzzy clustering achieved at every grid point (Paasche and Eberle, 2011). This new approach facilitates high resolution/high quality map compilation of extensive areas with downsized input of time and effort.

Multivariate statistical classification methods have been used in geophysical data interpretation for a few decades. First suggestions to use decision theoretical and multivariate statistical techniques for the integrated interpretation of magnetic and gravity data were published as early as 1972 by Harff (1972). Since the advent of powerful computing capabilities, case studies of crisp clustering have successfully been carried out over small areas selected for mineral exploration, hydrogeological or engineering geology (e.g. Eberle,

1993; Schetselaar, 2002; Tronicke et al., 2004; Debeglia et al.; 2006, Teixeira et al., 2006; Eberle et al., 2012). More recently, fuzzy cluster analyses have become increasingly popular, since they offer a fuzzy, but quantitative, numerical assessment of the classification trustworthiness (e.g. Paasche et al., 2006, Paasche and Eberle, 2009, Paasche and Eberle, 2011). In this paper, we will discuss the maps resulting from both automated recognition of linear tectonic features using a combination of published ridge detection algorithms on high resolution airborne magnetic data and consecutive fuzzy clustering of high resolution airborne gamma-ray data and satellite imagery from part of the Karas Region situated in southern Namibia. All grid data were co-located with a grid spacing of 200 m which easily enabled the creation of a multivariate sample vector at each grid point. In order to prevent constraints of the computing capacity of a standard laptop computer, we did not use the original grid data, which were 50 m spaced, but rather created a new grid with 200 m spacing.

## **2. Study Area**

The study area extends between latitudes S 26°30' and S 28°30' and longitudes E 17°00' and E 19° 20' in south-central Namibia, approximately rectangular in shape with a width of 100 km and approximate length of 200 km in the SW-NE direction (Figure 1). The village of Grünau is situated approximately in the centre of the study area, and the Karasburg - Grünau - Seeheim road separates the area into a northeastern and southwestern part, which coincidentally are different in terms of their morphology. To the northeast of this road, the morphology is characterized by tectonic graben and horst features with partially steep escarpments, whilst around and southwest of Grünau the landscape is fairly flat and monotonous becoming more and more arid closer to the South Africa - Namibia border which runs along the banks of the Orange river (Figure 1).

Geologically, the study area extends across the SE-NW trending Gordonia Subprovince of the Namaqua-Natal Metamorphic Belt between its Southern Front Zone in the southwest and the Namaqua Front in the northeast (Cornell et al., 2006; Miller, 2008). The Gordonia Subprovince is a high-metamorphic grade, deeply exhumed, intensely foliated terrane with a granulite and charnockite core and lower grade marginal zones. Pre-tectonic rocks within this subprovince have apparent ages of about 2000 Ma (Miller, 2008). High grade metamorphism is strongest in the Klein-Karas and Grünau areas, which gradually weakens when approaching the Pofadder-Marshall Rocks lineament and shear zone in the southwest and Excelsior-Lord Hill lineament and shear zone in the northeast (cf. Figure 1, Becker et al., 2006). The Namaqua tectogenesis took place in the time interval from 1300 to 1100 Ma.

Large parts of the study area are, however, overlain by younger rock of the pan-African Nama Group and the Jurassic-Tertiary Karoo Supergroup (Geological Survey SWA/Namibia, 1977). These are mostly sandstone, shale and quartzite of the Nama Group and tillite, silt-, mud-, limestone of the Karoo Supergroup as well as extensive post-Karoo dolerite sills (Schneider, 2008).

The tectonic stress pattern within the study area is primarily reflected by three sets of dykes with different age. The pre-Nama Gannakouriep dyke swarm trends SW-NE to SSW-NNE and is strongly reflected by the magnetic data (Figure 2a). According to geologic mapping (Geological Survey SWA/Namibia, 1977) the occurrence of Gannakouriep dykes is most frequent in the vicinity of the Klein-Karas and Grünau settlements. The Gannakouriep Suite represents a mafic dyke swarm which cuts across the basement rock and was emplaced over a prolonged period of time. The dyke swarm extends more than 300 km in southern Namibia

and the Northern Cape Province of South Africa. The width of individual dykes is typically between 5 and 30 m, but their length can be as much as 100 km (Minaar and Eberle, 2013).

Lamprophyre and Bostonite dykes are associated with the first intrusion phase of the post-Nama Kuboos-Bremen Suite while acid and alkaline dykes are younger and associated with the second phase (Schreuder and Genis, 1977). The acid dykes are represented by rhyolite and quartz porphyry and the alkaline dykes by bostonite and porphyritic trachyte. These dykes often strike W-E and can be observed to cut across the Gannakouriep dykes (Minaar and Eberle, 2013).

The Kuboos-Bremen Suite comprises a number of individual intrusive complexes which are situated along a SW-NE trending line from the west coast of the Northern Cape Province of South Africa to farm Bremen in southern Namibia (Smithies and Marsh, 1996). The suite is acidic and consists almost entirely of foyaitite, syenite and alkali granite. It is post-orogenic, undeformed and the result of Pan-African magmatism (Smithies, 1992).

Jurassic-Cretaceous dolerite sills are extensively developed in the study area. Based on outcrop observations and borehole data, their thickness varies between 50 and 150 m.

Intrusion of the dolerite magma was accompanied by faulting and gentle folding (Minaar and Eberle, 2013).



### 3. Database

Airborne magnetic and gamma-ray data were acquired over the study area during the early years of the new millennium by companies from Australia and South Africa. Flight line spacing was 200 m over individual survey blocks, each of them partially covering the study area. Flight line direction was N-S. State-of-the-art sensors for magnetic and gamma-ray data were mounted aboard fixed-wing aircraft. Magnetic data were recorded every tenth of a second and the integration time of the airborne gamma-ray detector was 1 second. Assuming an average flight velocity of < 250 km/h, magnetic readings were taken every seven metres, and gamma ray counts were accumulated over consecutive distance intervals of 70 m along flight track. Data recording was digital.

After comprehensive processing, all airborne data were subsequently projected onto a regular grid with 50 m spacing and referenced to UTM Zone 33 S. Figures 2a – 2c display the airborne geophysical data collected over the study area. The International Geomagnetic Reference Field (IGRF) corrected total magnetic intensity  $\Delta T$  is shown in Figure 2a.

Numerous S-N to SSW-NNE trending magnetic dikes are obvious. These are pre-Nama, late Proterozoic mafic dikes, known as the Gannakouriep dike swarm, post-Nama late Palaeozoic trachyte and rhyolite dikes, and Jurassic-Cretaceous post-Karoo dolerite dikes. In this study, the magnetic data and their derivatives are used to reveal the tectonic stress pattern existing in the study area (see Section 4.2).

The surface distributions of the natural radioelements potassium ( $^{40}\text{K}$ ) and thorium (Th) are displayed by Figures 2b and 2c. This conversion of the decay count rates recorded in the potassium and thorium windows was possible because the airborne gamma-ray systems were calibrated using the Hentjies Bay Test Range prior to surveying (IAEA, 2010). The total count and uranium data were not used for this study as the total count may be considered a linear

combination of the data collected in the three radioelement windows. Moreover, in terms of decay statistics, the uranium channel data are known to be the poorest in quality.

With regard to the Landsat (LS) data, previous investigations (e.g. Eberle and Paasche, 2012) have shown that Landsat bands 1 - 5 and 7 are highly correlated and do not provide any additional information which would not be reflected by bands 1 and 6 on a regional scale.

Thus, LS bands 1 and 6, with wavelengths centred at 0.5  $\mu\text{m}$  and 11  $\mu\text{m}$ , respectively, (Figures 2d and 2e) were chosen for this study. These two Landsat bands and the 40K and Th radioelement data were integrated to obtain the lithology pattern of the study area (see Section 5). These four data sets are similar in that their data sources almost completely originate from the surface and certainly not deeper than from 50 cm below surface.

The grid points are shared by all four data sets presented in Figures 2b – 2e. It is therefore possible to establish the data matrix with four columns representing the four variables (40K, Th, LS 1 and LS 6) and  $n$  rows where  $n$  is the total number of grid points with known positions  $(x,y)$ . Each row represents a multivariate sample vector. The data matrix was subsequently submitted as input to fuzzy clustering for extracting information relevant to mapping the surface lithology of the study area.

## **4. Recognition of linear features and their pattern**

### **4.1 Methodology**

Over the last decades, numerous approaches have been developed to identify linear features, such as ridges or edges, in potential field data. In most cases, a single attribute has been derived from the underlying data, e.g., based on local curvature or slope from mapped data, highlighting linear features of a distinct spatial frequency. In this paper, different classical attributes relying on slope or curvature measures are integrated in a largely automated manner

to delineate the edges of prominent linear features hidden in the mapped total magnetic intensity  $\Delta T$  (Figure 2a). The general processing flow is sketched in Figure 3 and commences with deriving five different attributes (Section 4.1.1). Consecutively, the resulting attribute maps are stepwise integrated to provide a single map displaying the location and extent of structural linear features. This merging of different attributes is considered to be superior to the single-attribute approach. Well known morphology and breadth-first search algorithms are used to further enhance significant linear features, i.e. a certain length and significance value must be provided by the analysed attributes. The properties of these attributes, their integration and the selective enhancement of only relevant linear features are discussed in the following.

This multi-attribute approach was applied to the first horizontal derivative of the mapped total magnetic intensity  $\Delta T$  bearing in mind that all attributes highlight ridges rather than edges.

Thus, edges present in the magnetic data are automatically recognized.

#### **4.1.1 Attributes**

As a result of a desktop study five attributes were chosen. These are the

- Blakely-Simpson Index (BSI; Blakely and Simpson, 1986)
- Modified D8 Algorithm (O'Callaghan and Mark, 1984)
- Discrete Laplacian (Reuter et al., 2009)
- Mean Curvature (Gray, 1997)
- D8 Algorithm as originally published by O'Callaghan and Mark, 1984.

##### **4.1.1.1 Blakely-Simpson Index (BSI)**

The Blakely-Simpson algorithm analyses the local curvature by comparing a sample,  $g_{ij}$ , of a data grid to its spatial neighbours (Figure 4a). If,  $g_{ij}$ , is larger than two diametrically

opposed spatial neighbours, its BSI is increased by one. Thus, the BSI will be a natural number in the range from zero to four. The BSI is calculated at all possible grid points and, as applied to the first horizontal derivative, it will locate edges present in the total magnetic intensity data. All grid points with BSI=0 are discarded and, due to the low dynamic range of the BSI, a significance assessment of the remaining grid points with BSI > 0 is not considered valuable.

#### **4.1.1.2 Modified D8 Algorithm**

The second attribute is developed from the D8 algorithm (O'Callaghan and Mark, 1984) which enables the location of ridges in grid data. Assuming a sample,  $g_{i,j}$ , the slope to its spatial neighbours will be determined (Figure 4b). If,  $g_{i,j}$ , is a local high, all calculated slopes would be negative. Out of the neighbours of,  $g_{i,j}$ , with negative slope, only the sample with the minimum slope is retained for further consideration. The slope from this sample to all its spatial neighbours is then calculated and once again only the spatial neighbour with minimum slope out of those with negative slope will be retained. This process is repeated until a sample with only positive or zero slopes is located (point '0' in Figure 4b). The number of analysed samples until meeting the final sample with non-negative slopes will be ascribed to,  $g_{i,j}$ .

To simplify, this approach could be described as analogue to tracking the length of the surface runoff of a water drop that falling on sample,  $g_{i,j}$ , where the data grid is considered as a topographic model. The runoff paths of all samples,  $g_{i,j}$ , are set to zero, once they receive influx from at least another sample (Figure 4b). To have the same (low) dynamic range as that for the BSI, all values exceeding four are set back equal to four. Similarly to the BSI, all grid points with a modified D8 attribute of zero can be discarded, as they do not indicate ridge locations of the horizontal derivative.

#### **4.1.1.3 Discrete Laplacian**

The discrete Laplacian (e.g. Reuter et al., 2009) provides a higher dynamic range and more continuous information than the BSI and modified D8 attributes as the output is not confined to discrete natural numbers. The Laplace operator as applied to the grid data is shown in Figure 4c where for every sample,  $g_{i,j}$ , the sum of differences to its spatial neighbours is calculated. The operator is, however, disadvantageous in that detected features are usually blurred and numerous artefacts are produced complicating the interpretation of the discrete Laplacian. Since the operator is applied on the inverse first horizontal  $\Delta T$  derivative, all negative values are set to zero.

#### **4.1.1.4 Mean Curvature**

The mean curvature (e.g. Gray, 1997) is used to highlight ridge positions rather than edges on a surface described by data samples. Mean curvature is understood as the average of the two principal curvature values of a sample,  $g_{i,j}$ , i.e. the minimal and maximal curvature values (see K1 and K2 in Figure 4d). This attribute offers a high dynamic range, but may produce features which appear blurred. Here it is calculated based on the inverse first horizontal  $\Delta T$  derivative. Since there is interest only in ridges, all negative values are again set to zero.

#### **4.1.1.5 D8 Algorithm**

Compared to the quite local information provided by the previously considered attributes, the D8 algorithm in its original version can be used to enhance long-wavelength structures, with maximum values for samples, for example, situated in an extended channel. Once again, using the analogy of precipitation surface runoff, the process can be described as quantifying the catchment at sample position,  $g_{i,j}$ , (Figure 4e). Assuming that a single water drop lands on

each grid point and counting all water drops arriving at grid point,  $g_{i,j}$ , during their surface runoff, then this total count number can be ascribed to,  $g_{i,j}$ . Note that this attribute will result in differently sized catchments for every grid point depending on the surface topography of the analysed data set. This attribute was calculated using again the inverse first horizontal  $\Delta T$  derivative data set.

#### **4.1.2 Attribute Integration**

The integration commences using the attributes with low dynamic range, namely BSI and modified D8. This integration is achieved by creating a new binary map with zeros at all positions where either the BSI or the modified D8 attribute equals zero as these positions can be discarded as places of potential linear features. A skeleton (e.g. Kong and Rosenfeld, 1996) of the binary map using a morphology algorithm (e.g. Dougherty, 1992) is then computed to ensure that each detected potential feature has a width of only one sample. To achieve a weighted skeleton map, the elements of the binary skeleton are multiplied with the BSI and modified D8 attributes. The resulting weighted skeleton map contains values out of the set  $\{1, 2, 3, 4, 6, 8, 12, 16\}$  at positions where edges are present in the  $\Delta T$  data set, with higher numbers generally indicating a higher detection significance.

Each sample of the weighted skeleton map is now multiplied individually by the high dynamic attributes, namely the discrete Laplacian, the mean curvature, and the D8 attributes. Once again it is obvious that the resultant three maps display zero where either the BSI or the modified D8 attributes are zero. Scaling the detection significance of each feature is then made possible using the integrated information of the weighted skeleton map (BSI and modified D8 attributes) and one of the three high-dynamic attributes. The normalization of the resultant three maps to a minimum-maximum range between 0 and 1 enables stacking these three maps while ensuring relatively equal contributions of the discrete Laplacian, the mean curvature and the D8 attributes. The properties of the resultant stacked map are determined by

the different characteristics of the considered attributes, for example, the high dynamic range of some attributes enables assessing the detection significance of the features at different wavelengths. Conversely, features at grid points where at least one of the low dynamic attributes is zero, are discarded. This final stacked map, however, still contains a number of features with very limited lengths or low detection significance which requires further processing to highlight meaningful features and diminish features which appear dubious.

#### **4.1.3 Enhancement of meaningful features**

To highlight meaningful features all detected features represented by only a single sample,  $g_{i,j} > 0$ , which is surrounded by zero samples, are removed from the map using a morphology algorithm (e.g. Lam et al., 1992). A list of all the remaining features is then computed demarcating their start/end positions applying a morphology algorithm (e.g. Olsen et al., 2011) on a binary version of the final stacked map. Subsequently a breadth-first search (e.g. Cormen et al., 2001) is performed to determine a feature dependent measure. This measure corresponds to the sum of the stacked detection significance values of the samples composing an individual linear feature.

Simply the detection significance value of each sample point is summed from the beginning or start point to the end of the feature. Once the end point (or the end points if the feature splits into several branches) is reached, the summed detection significance is multiplied by the length (number of samples) of the feature under consideration. Thus features with numerous sample points, i.e. extending over a long distance, are highlighted. At the same time, features with a rather limited number of samples but with very high detection significance values are also enhanced. It follows that rather short features with low detection significance values receive diminished importance. To ensure that every feature is analysed only once and also for reason of computational efficiency, all end points are removed from the list of start/end points once they are reached.

To visualize the result of the final linear feature enhancement, the transparency of the map symbols are scaled according to their relevant detection significance value. An upper and lower empirically determined threshold is used for scaling the colour saturation of the detected features. In this case, lower and upper thresholds of 0.8 % and 5 % of the maximal detection significance were chosen. Thus, all samples with detection significance of 5 % of the maximum value or greater will be fully opaque/black, and samples with detection significance of less than 0.8 % of the maximal value will be fully transparent/white.

Since the start/end points of each feature are known, any detected linear feature may be described as a vector with a step length from sample to sample and a directional angle. Thus, the detected linear feature map can easily be overlain on any other map type, e.g. geology, remote sensing, geophysics, of different discretization without resampling the linear feature map.

## **4.2 Application**

As a depth-related analysis was required, upward magnetic field continuations for 1000 m and 5000 m above sensor height were calculated and converted to their first horizontal derivative. The linear feature detection was then carried out following the processing flow as sketched in Figure 3, using (a) the difference field of the  $\Delta T$  data acquired at flight height and the 1000 m upward continuation response, (b) the difference between 1000 m and 5000 m upward continuation responses, and (c) the response of the 5000 m upward continuation, respectively.

Figure 5 displays the five attribute maps which refer to the difference field between the  $\Delta T$  data acquired at flight height and their 1000 m continuation response, thus illustrating the first step of the multi-attribute based linear feature detection process. The integration of the five attribute (Figures 5a – 5e) and the subsequent application of the linear feature enhancement procedure resulted in the final linear feature map for shallow magnetic sources as shown in



Figure 6a. Similarly, the final maps for lineaments with deeper-situated magnetic sources (between 1000 m and 5000 m, and > 5000 m) are given in Figures 6b and c, respectively. In the right column of Figure 6, the detected lineament patterns are overlain on the coloured image of the anomalies  $\Delta T$  of the total magnetic intensity at sensor height for ease of comparison.

### **4.3 Interpretation**

The geological equivalent of a linear feature detected in a magnetic data set can be varied. In the southern African context dolerite dikes represent the most frequently occurring geological equivalent or model. Geological faults along which magnetic rock properties are different will also produce extended linear features. In general, any edge of a geological body across which rock magnetization abruptly changes direction and/or magnitude results in a linear feature which is detectable using the newly developed, automatic multi-attribute process.

The lineament pattern shown in Figures 6a and 6b is dominated by the SSW-NNE to SW-NE trending Gannakouriep dike swarm. The occurrence of the dikes is highest in the //Karas magnetic trough where the magnetic field intensity is low on a regional scale (Eberle et al., 2002, 2005). The trough is oriented SE-NW extending more than 200 km with an average width of 100 km. It characterizes the zone where high grade Namaqua metamorphics have been mapped (Table 7-7, p.7-47 in Miller, 2008), possibly indicating that high grade Namaqua metamorphic rock is either a carrier of a strong remanent magnetic component or is largely depleted in magnetic minerals.

The Gannakouriep dike swarm was reactivated during the Jurassic/Early Tertiary giving rise to numerous post-Karoo dolerite dikes which are concentrated in the Grünau and Klein //Karas areas (Geological Survey SWA/Namibia, 1977). The post-Karoo dikes have a similar

orientation to the pre-Nama Gannakouriep dike swarm and it is hardly possible to distinguish them using the maps shown in Figures 6a and 6b.

Towards the marginal zones of the Gordonia Subprovince in the southwest and northeast of the study area, the number of individual dikes decreases. The remaining magnetic lineaments swing into the SE-NW strike of the //Karas trough as they approach the Excelsior Lord Hill Lineament and Shear Zone in the northeast, and to the Pofadder-Marshall Rocks Lineament and Shear Zone in the southwest. This is coincident with the magnetic trend marginal to the Gordonia Subprovince which also generally follows a SE-NW orientation.

It is also important to note that the E-W trending Bostonite, trachyte and rhyolite dikes are largely non-magnetic as any such linear features have not been detected and are not reflected by the maps shown in Figure 6. Assessing the lineament map displayed in Figure 6c, it is noted that nearly all linear features are fairly extended and roughly orientated either east-west or south-north. This might indicate that the tectonic stress pattern at depths >5000 m is different from the near-surface pattern where the Gannakouriep and Kuboos-Bremen directions are prominent.

## **5. Cluster analysis for geologic mapping**

### **5.1 Methodology**

Unsupervised fuzzy clustering of large, high resolution airborne geophysical data suites combined with satellite and/or hyper-spectral imagery to compile (pseudo-) lithology maps has recently been described for automated, objective geologic mapping in detail (Paasche and Eberle, 2009, 2011). Eberle and Paasche (2012) have also presented a mineral exploration case study from the Northern Cape Province of South Africa where airborne geophysical and regional geochemistry data sets together with Landsat imagery were clustered and the clustering result was subsequently used for geologic mapping and target area generation. In this study, the method as discussed in detail by Paasche and Eberle (2011) was applied on selected areas covered by the excellent Namibian high resolution airborne magnetic and gamma-ray data sets and also by Landsat 7 satellite imagery. The Grünau area is one out of four areas chosen within the //Karas region of Namibia.

Prior to running any kind of cluster analysis, it is expedient to examine a few statistical properties of the data with regard to their interrelationship and distribution. Abundance and high correlation amongst the data sets should be avoided as increase of information is not necessarily acquired by adding a data set which is highly correlated. Following Paasche and Eberle (2011), a fuzzy clustering algorithm was used employing a Mahalanobis distance measure for calculating the distances from the cluster centres to each sample. A non-diagonal global covariance matrix was inserted in the Mahalanobis distance measure to relax restrictive conditions on the shape of the histogrammic distribution of each data set, e.g. as required by algorithms using simple Euclidean distance measures (cf. Paasche and Eberle, 2011).

To start the clustering process, the number of clusters (groups) into which the data are to be assembled must be selected. It is a matter of experience to determine the optimum number of

clusters, often done in the context of what is known about the data suites. Alternatively, repeated running the clustering process by consecutively increasing the number of clusters by one enables the optimum number to be determined by either analysis of mathematical criteria (e.g. Paasche and Eberle, 2009) or by empirical assessment of the clustering results. For example, a three-cluster case is not appropriate when trying to support geological mapping of an area as large as 20 000 km<sup>2</sup> containing several major lithological units. Conversely, to capture the heterogeneity of the data suite it is sometimes necessary to discard small optimum group numbers as determined by the mathematical criteria and, instead, to go for a clustering result with a greater number of clusters (groups).

In this study, the objective of clustering is to produce a high resolution (pseudo-) geological map over a large area which is covered by high resolution airborne geophysical data and satellite imagery. Following the flow chart shown in Figure 7, the airborne 40K- and Th-surface distribution data were integrated with the Landsat 7 band 1 and band 6 data using unsupervised fuzzy clustering to obtain a zonal (classified) map. The result of the fuzzy cluster analysis was defuzzified (e.g. Paasche and Eberle, 2011) and the final zonal (classified) map is displayed by Figure 8a. Eight clusters (groups) were considered sufficient to describe the variability of surface geology within the study area and the zones are indicated by different colours. The colour intensity is chosen as a linear measure proportional to the trustworthiness (Paasche and Eberle, 2011) of the assignment of a sample. For example, a sample assigned to a cluster where it has a low degree of membership will be characterized by low colour intensity, whilst a sample ascribed to a cluster with a high degree of membership will have high colour saturation.

The zonal (classified) map (Figure 8a) can now contribute towards the geological context. As soon as the zonal (classified) map is overlain by a linear feature map provided by multi-attribute analysis of the magnetic data a high resolution geological map is obtained. In this

case, the tectonic pattern of the magnetic sources buried at depths from 1000 m to 5000 m (Figure 6b) was selected to draft the high resolution, value-added geological map shown in Figure 8b. This map can be further improved by ascribing geological/geophysical parameters and units known from the study area to each cluster provided by the unsupervised fuzzy clustering process. This is achieved by examining and exploiting archived geological reports, borehole data, existent map materials, and by scrutinizing what is known from previous earth science investigations.

## **5.2 Geologic Interpretation**

As this cluster analysis is based on data with no or very small depth of penetration (< 50 cm), an immediate comparison with the mapped surface geology (Genis and Schalk, 1984; Geological Survey SWA/Namibia, 1977; Geological Survey Namibia 1997, 1999a, 1999b; Minaar and Eberle, 2013) is possible. Based on these maps and reports, a number of geological/geophysical attributes could be ascribed to the eight zones (clusters) obtained from advanced fuzzy c-means clustering. These are described in Table 1.

It can be learnt from Table 1 that the members of the basal Nama Group (clusters 1 and 5) are characterized by low K- and Th-concentrations, which is also supported by Figures 2c and 2d depicting the spatial K- and Th-distributions. In contrast to the Nama Group, most syntectonic Namaqua granitoids (clusters 6 and 7) are characterized by high K- and Th-concentrations (> 5%) and strong reflectance in Landsat band 1. On the other hand, pre-tectonic Namaqua granitoids and gneiss (cluster 4) occurring around Grünau and to the north in the ‘Corridor’ are not as rich in K and Th as the syntectonic varieties. Reflectance of the thermal Landsat band 6 is high over areas covered by sediments of the Dwyka and Ecca groups (cluster 2).

These are mainly sediments of the Prince-Albert Formation, in the southeast and northwest corners of the study area.

Despite the above, it is difficult to distinguish between late Nama sediments and Karoo-age sediments of the Ecca Group which both fall within cluster 3 (Table 1). Similarly, cluster 8 appears to comprise both late Nama sediments and post-Karoo dolerite sills. However, these sills may be recognized by their eye-catching sharp borders circumscribing radioelement lows (Figures 2b and 2c). The small number of data sets used for clustering might have limited the power of differentiation and delineation of sediments with similar mineral composition, but different age.

### **5.3 Zonation quality**

The variable colour saturation reflects the trustworthiness of the assignment of a sample to a certain cluster which enables spatially variable pseudo-probabilistic assessment of the zonation quality. The degree of membership determined for a sample ascribed to a certain cluster is linearly linked to its trustworthiness, and can therefore be considered a direct fuzzy measure of clustering quality inherently provided by the fuzzy cluster algorithm.

More than two decades ago, attempts were undertaken to evaluate the aposterior classification probability of crisp clustering results. For example, Eberle (1993) employed a crisp cluster algorithm to produce a zoned map based on integrated analyses of various helicopter-borne geophysical data sets. In contrast to fuzzy clustering, crisp cluster analysis does not provide a numerical measure of the degree of cluster membership, but rather in a binary sense, i.e. a sample is a member of a specific cluster (1), or a sample is not a member of this cluster (0). In an attempt to obtain a numerical quality measure to assess the crisp cluster result, Eberle (1993) applied stepwise linear discriminant analysis to obtain an aposterior probability matrix with  $n$  rows and the number of columns equal to the number of the clusters (groups, classes).

The aposterior classification probability values are equivalent to the trustworthiness in that they sum to unity for every sample.

The result of the fuzzy cluster analysis shown in Figure 8a was entered as a training data set to run a discriminant analysis (e.g. Fraley and Raftery, 2012) using a Mahalanobis distance measure with stratified covariance matrix by class (group) and equal probability for each class. The classified map resulting from discriminant analysis is shown in Figure 8c and resembles the cluster map of Figure 8a pretty closely with regard to classification, notwithstanding, differences in colour saturation evident in some areas of the map. These differences in colour saturation are indicative of varying membership and probability values of each sample. Differences may also occur where samples are assigned to different affiliations after consecutive fuzzy cluster and discriminant analysis. Figure 9a depicts all the sample points where the assignments by fuzzy clustering and discriminant analysis are different and represents the difference between Figures 8a and 8c known as misclassification. In the present case, approximately 8 % of the samples were not ascribed to the same cluster/class. This suggests that aposterior probability and trustworthiness are not identical parameters nor are they related by linear transformations. This also becomes evident when calculating the aposterior probability, which requires a recurrent exponential term.

To study the properties of probability and trustworthiness, the cluster 2 sample membership (Figure 9b) is compared with the class 2 probability (Figure 9c). It becomes evident that the sample membership values are more gradational between 0 and 1 than the probability values which tend to be either close to 1 (high probability) or to 0 (low probability). When assessing the maps with regard to classification objectives, the human eye may feel more comfortable with Figure 9b.

The differences of assignment occur most frequently in areas with low trustworthiness or aposterior probability values. In geologic mapping, however, these areas can be of special

interest as they may highlight the transition between adjacent geologic units or fault zones which may warrant further ground investigations.

Fuzzy clustering and discriminant analysis generate very similar results thus providing the opportunity of randomly clustering a subset of the data to learn discriminant rules based on the cluster results and by using the clustered subset as a training data set. In so doing, it was realized that zonation results were highly robust when tested using random training data sets comprising only 1 % of the total number of samples. This approach may enable the clustering/classifying of much larger data sets than analysed in this paper (ca. 500 000 samples) in a numerically efficient manner.

In their applications, Paasche et al. (2006), Paasche and Eberle (2010), Hachmöller and Paasche (2013) consider fuzzy cluster analysis in their applications not only as a classification technique, but rather as an imaging tool which enables the integration of multiple, spatially continuous data sets. The spatial heterogeneity of the multiple data is described by a dimensionless structural fuzzy membership matrix, which delineates the spatial heterogeneity of all input maps, and an attribute matrix which provides the mean values of the underlying data types for each cluster. The analysis of the membership distribution, however, provides valuable guidance to establish a sampling scheme with a smaller number of sampling points, but which are optimized with regard to their spatial position. Such a sampling scheme may be used to acquire additional target measurements for subsequent interpolation using fuzzy membership information. A very similar technique could also be adopted to aposterior classification probability values, but the smoother membership distribution is considered more suitable when using the zonation information to interpolate sparse measurements, such as chemical abundance sampling. This makes fuzzy cluster analysis a technique superior to the combination of crisp clustering and aposterior probability calculation.



## 5. Conclusions

This study demonstrates a method for largely automated, fast and objective interpretation of airborne geophysical and space-borne remote sensing data to provide a geological map of a large area, 200 km by 100 km in size. This support to geological mapping consists of two elements, namely a structural interpretation of the airborne geophysical data and a determination of the spatial distribution and occurrence of lithological units.

- Based on integrated multi-attribute analysis a method to detect linear features from airborne magnetic data was designed using a number of published attributes (Figure 3). At the completion of this stepwise, largely automated process the tectonic pattern of the study area was revealed (Figure 6).
- To obtain the spatial distribution and occurrence of lithological units, unsupervised fuzzy clustering of high resolution multi-method data (airborne geophysics, satellite imagery) proved to be highly accurate and reliable when compared to conventional geological mapping results. The method is fully computerized, rapid and objective, but when it comes to ascribing the multi-method zonal pattern to geological/geophysical parameters, cognisance of the known geology of the study area is helpful to compile the geology map (Table 1 and Figure 7).
- To assess the quality of the fuzzy clustering results, discriminant analysis was used. The discriminant system was trained using the clustering results. Comparing the cluster and discriminant analysis results reveals that only a small percentage of 8% was ascribed to another class/group by discriminant analysis than by previous fuzzy clustering (misclassification).
- Classification by discriminant analysis also proved to be robust, even when the training data set was only a small portion, e.g. 1%, of the complete multivariate data set. When using a smaller, randomly chosen, subset to develop classification rules,

fairly good and reliable discrimination results were obtained with a success rate of 90%. This should be kept in mind when very large multivariate data suites have to be clustered.

- Lastly a comparison between the membership (trustworthiness) provided by clustering and the aposterior probability values resulting from discriminant analysis shows that these two parameters have different characteristics. The histogram of the membership (trustworthiness) is smoother and therefore considered more suitable when the interpolation of sparse data points is required.

The methodology and results presented here demonstrate that automated pattern recognition can support geological mapping and mineral exploration of large study areas. This methodology is considered well suited to a number of African countries whose large territories have recently been covered by high resolution airborne geophysical surveys, but where geological mapping is poor, incomplete or outdated.

## **6. Acknowledgement**

The authors acknowledge the support received from a number of authorities and individuals in Namibia, South Africa and Germany. We are thankful to the Geological Survey of Namibia (GSN) and its Geophysics Division which was the driving force behind the //Karas Interpretation Project. U. Schreiber (GSN) thankfully provided materials for setting up Figure 1.

The practical aspects of using cluster algorithms for geologic mapping and mineral target generation were examined and laid through bi-lateral research co-operation between the South African Council for Geoscience (CGS) and the University of Potsdam-Germany.

The authors are deeply obliged to the South African National Research Foundation (NRF) and the International Bureau (IB) of the Ministry of Education and Research (BMBF) of Germany that provided the funds for travel and fruitful interaction of researchers from both countries over a period of four years (BMBF: SUA08/15; NRF: UID 69441, 81579).

## 7. References

Becker, T., Schreiber, U., Kampunzu, A.B., Armstrong, R., 2006. Mesoproterozoic rocks of Namibia and their plate tectonic setting. *Journal of African Earth Sciences* 46, 112 – 140.

Blakely, R. J., Simpson, R.W., 1986. Approximating edges of source bodies from magnetic or gravity anomalies. *Geophysics* 51, 1494 – 1498.

Cormen, T., Leiserson, C., Rivest, R.L., Stein, C., 2001. *Introduction to Algorithms*. MIT Press, 2<sup>nd</sup> edition.

Cornell, D.H., Thomas, R.J., Moen, H.F.G., Reid, D.L., Moore, J.M., Gibson, R.L., 2006. The Namaqua-Natal Province, in: Johnson, M.R., Anhaeusser, C.R., Thomas, R.J. (Eds.), *The Geology of South Africa*. Geological Society of South Africa/Council for Geoscience, Pretoria, pp. 319 - 379.

Debeglia, N., Martelet, G., Perrin, J., Truffert, C., Ledru, P., Tourlière, B., 2006. Semi-automated structural analysis of high resolution magnetic and gamma-ray spectrometry airborne surveys. *Journal of Applied Geophysics* 58, 13-28.

Dougherty, E.R., 1992. *An Introduction to Morphological Image Processing*, Optical Engineering Press, ISBN 0-8194-0845-X, pp. 161.

Eberle, D., 1993. Geologic mapping based upon multivariate statistical analysis of airborne geophysical data. - International Institute for Aerospace Survey & Earth Sciences (ITC), *Journal* 1993 – 2, Delft, pp. 173 – 178.

Eberle, D., Andritzky, G., Hutchins, D., Wackerle, R., 2002. The regional magnetic data set of Namibia: Compilation, contributions to crustal studies and support to natural resource management. *South African Journal of Geology* 105, 361-380.

Eberle, D., Andritzky, G., Ostwald, J., Wackerle, R., 2005, Domains and tectonic inventory of the regional airborne magnetic data set (south of latitude 24S) as a guide to areas with future mineral potential in southern Namibia, in: *Geological Survey Bulletin* 4, 53 – 80. Ministry of Mines and Energy, Windhoek, Namibia.

Eberle, D., Daudi, E.X., Muiuane, E.A., Nyabeze, P., Pontavida, A.M., 2012. Crisp clustering of airborne geophysical data from the Alto Ligonha pegmatite field, northeastern Mozambique, to predict zones of increased rare earth element potential. *Journal of African Earth Sciences* 62, 26-34, doi: 10.1016 / j.jafrearsci.2011.08.003

Eberle, D., Paasche, H., 2012. Integrated data analysis for mineral exploration: A case study of clustering satellite imagery, airborne gamma-ray and regional geochemical data suites, *Geophysics* 77, 4, B167-B176. doi: 10.1190/GEO2011-0063.1

Fraley, C., Raftery, A.E., 2012. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* 97, 458, 611-631.

Genis, G., Schalk, K.E.L., 1984. The Geology of Area 2618: Keetmanshoop, Explanation of Sheet 2618, Scale 1: 250 000, Geological Survey of SWA/Namibia, Department of Economic Affairs, Windhoek.

Geological Survey SWA/Namibia, 1977. Geological Map Sheet 2718 Grünau, 1:250 000 Map Series, Ministry of Mines and Energy, Windhoek.

Geological Survey of Namibia, 1997. Geological Map Sheet 2818 Warmbad, 1:250 000  
Geological Series, Ministry of Mines and Energy, Windhoek.

Geological Survey of Namibia, 1999a. Geological Map Sheet 2716 Ai-Ais, 1:250 000  
Geological Series, Ministry of Mines and Energy, Windhoek.

Geological Survey of Namibia, 1999b. Geological Map Sheet 2816 Oranjemund, 1:250 000  
Geological Series, Ministry of Mines and Energy, Windhoek.

Gray, A., 1997. The Gaussian and Mean Curvatures, in: *Modern Differential Geometry of Curves and Surfaces with Mathematica*, 2nd ed., Boca Raton, FL: CRC Press, pp. 373-380.

Hachmöller, B., Paasche, H., 2013. Integration of surface-based tomographic models for zonation and multimodel guided extrapolation of sparsely known petrophysical parameters. *Geophysics* 78, 4, EN43 – EN53, doi: 10.1190/GEO2012-0417.1

Harff, J., 1972. Komplexinterpretation geologischer Informationen mit Klassifizierungsmethoden, in: Peschel, G., (Ed.), *Beiträge zur Theorie der Komplexinterpretation geowissenschaftlicher Informationen*, ZGI Berlin.

Hoepfner, F., Klawonn, F., Kruse, Runkler, T., 1999. *Fuzzy Cluster Analysis: Methods for classification, data analysis, and image recognition*. John Wiley and Sons.

International Atomic Energy Agency (IAEA), 2010. Radioelement Mapping. IAEA Nuclear Energy Series, No. NF-T-1.3, Wien, p. 81.

Kong, T. Y., Rosenfeld, A., 1996. *Topological Algorithms for Digital Image Processing*, Elsevier Science, pp. 291.

Lam, L., Lee, S.W., Suen, C.Y, 1992. Thinning Methodologies - A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, 9.

Miller, R. McG., 2008. Namaqua Metamorphic Complex, in: *The Geology of Namibia*, vol. 1, Archean to Mesoproterozoic, pp. 7-1 – 7-55, ISBN 978-0-86976-731-3, Ministry of Mines and Energy, Windhoek.

Minaar, H., Eberle, D., 2013. An integrated interpretation of high resolution airborne geophysical data of the Karas region, Namibia. Unpublished report commissioned by Council for Geoscience, Pretoria, South Africa, pp. 62 p., 34 figs., 1 app.

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* 28, 3, 323-344.

Olsen, M.A., Hartung, D., Busch, C., Larsen, R., 2011. Convolution approach for feature detection in topological skeletons obtained from vascular patterns. *IEEE Symposium Series on Computational Intelligence* 163-167, doi 10.1109/CIBIM.2011.5949223

Paasche, H., Tronicke, J., Holliger, K., Green, A.G., Maurer, H., 2006. Integration of diverse physical-property models: Subsurface zonation and geophysical parameter estimation based on fuzzy c-means cluster analysis. *Geophysics* 71, 3, H33-H44, doi: 10.1190/1.2192927

Paasche, H., Eberle, D., 2009 Rapid integration of large airborne data suites using a fuzzy partitioning cluster algorithm. *Exploration Geophysics* 40, 277-287. doi: 10.1071/EG08028

Paasche, H., Eberle, D., 2010. Regional geochemical mapping constrained by airborne radiometric data and satellite imagery. *Proceedings of the EGM 2010 International Workshop*, Capri, Italy, pp. 4.

- Paasche, H., Eberle, D., 2011. Automated compilation of pseudo-geology maps from geophysical data sets: A comparison of Gustafson-Kessel and fuzzy c-means cluster algorithms. *Exploration Geophysics* 42, 275-285, doi: 10.1071/EG11014
- Reuter, M., Biasotti, S., Giorgi, D., Patane, G., Spagnuolo, M., 2009. Discrete Laplace-Beltrami operators for shape analysis and segmentation. *Computers & Graphics* 33, 3, 381–390.
- Schetselaar, E.M., 2002. Petrogenetic interpretation from gamma-ray spectrometry and geological data: The Arch Lake zoned peraluminous granite intrusion, Western Canadian Shield. *Exploration Geophysics* 33, 35-43, doi: 10.1071/EG02035
- Schneider, G., 2008, *The Roadside Geology of Namibia*, 2<sup>nd</sup> ed., Bornträger, Stuttgart, Sections 9.11 and 9.12, pp. 189-195.
- Schreuder, C.P., Genis, G, 1977. Explanation to Sheet 2718 Grünau, 1:250,000 Geological Map Series. Government Printer, Pretoria, and Geological Survey Namibia, Windhoek.
- Smithies, R.H., 1992. The geochemical evolution of three alkaline complexes in the Kuboos-Bremen igneous province, southern Namibia. Ph.D. thesis, Rhodes University, Grahamstown, pp. 197, <http://eprints.ru.ac.za/4563/1/SMITHIES-PhD-TR92-68.pdf>
- Smithies, R.H., Marsh J.S., 1996. Alkaline rocks in the Kuboos-Bremen Igneous Province, southern Namibia: The Grootpenseiland and Marinkas Kwela Complexes. *Communications of the Geological Survey of Namibia* 11, 13-20.



Teixeira, A., Silva, A.M., Pires, A.C.B., Morães, R.A.V., Souza Filho, C.R., 2006. Integração e análise de dados aerogeofísicos por meio de técnicas de processamento digital de imagens e classificação não supervisionada: O exemplo do Greenstone Belt Rio das Velhas, Quadrilátero Ferrífero, MG, Revista Brasileira de Geofísica 24, 559-572.

Tronicke, J., Holliger, K., Barrash, W., Knoll, M.D., 2004, Multivariate analysis of crosshole georadar velocity and attenuation tomograms for aquifer zonation, Water Resources Research 40, W01519, doi: 10.1029/2003WR002031

**Table 1. Geologic/geophysical attributes of eight zones obtained after fuzzy clustering of the K and Th surface distributions and Landsat bands 1 and 6.**

<b>Zone (Cluster)</b>	<b>Geologic/Geophysical Attributes</b>
1	Schwarzrand Subgroup: low in K (< 0.5%), Basal Nama Group
2	Dwyka Group and Prince Albert Fm. (strong reflector at 0.45-0.52 $\mu\text{m}$ , only western slopes of Klein Karas Berge), Ecça Group, Karoo Supergroup
3	(a) Stockdale Fm., Lower Fish River Subgroup, Nama Group (b) Aussenkjer Fm., Ecça Group, Karoo Supergroup (only in the southeast of the study area)
4	Namaqua granitoids: not as rich in K as Zones 6 and 7, at an average 3%; pre-tectonic?
5	Kuibis Subgroup: low in K (< 0.5%), Basal Nama Group
6	Namaqua granitoids: rich in K (> 5%); outcropping?; syntectonic?
7	Namaqua granitoids: rich in K (> 5%), covered by thin gravel and soil, syntectonic?)
8	Nababis Fm., Upper Fish River Subgroup, Nama Group; also post-Karoo dolerite sills in the extreme northwest and southeast corners of the study area

## Captions of Figures

**1:** Geological sketch map of the Grūnau study area extending across the Gordonia Subprovince of the Namaqua-Natal Metamorphic Belt. The study area is circumscribed by the dashed red lines.

**2:** Airborne geophysical data and satellite imagery of the study area projected on a grid with 200 m node spacing. (a) Anomalies  $\Delta T$  of the total magnetic intensity, (b) Potassium surface concentration, (c) Thorium surface concentration, (d) Landsat band 1, (e) Landsat band 6.

**3:** Flowchart of the semi-automated processing of multiple attributes to detect and locate linear features in the airborne magnetic data set.

**4:** Schematic illustration of the attributes derived from the total magnetic intensity data and used for the detection and location of linear features hidden in the airborne magnetic data.

(a) Blakely-Simpson Index (BSI), (b) Modified D8 algorithm, (c) Discrete Laplacian, (d) Mean curvature, (e) D8 algorithm.

**5:** (a) Blakely-Simpson index, (b) Modified D8 algorithm, (c) Discrete Laplacian, (d) Mean curvature and (e) D8 algorithm as applied on the difference field between the  $\Delta T$  data as acquired at flight height and the computed response at 1000 m above sensor height.

**6:** Results of the semi-automated linear feature detection process applied on (a) the difference between the  $\Delta T$  data as acquired at flight height and the computed response at 1000 m above sensor height, (b) the difference between the computed upward continuation responses of 1000 m and 5000 m, respectively, and (c) the computed upward continuation of 5000 m. The left column shows the detected linear features in grey to black with a saturation scaled to the detection significance. In the right column, the detected patterns are overlain on top of the corresponding airborne magnetic data.

**7:** Flow chart depicting the integration of the data sets shown in Figures 2b - 2e. Applying fuzzy cluster analysis and adding existing geological/geophysical evidence provides the zonal (classified) map and subsequently the value-added geological map.

**8:** (a) Zonal (classified) maps obtained from fuzzy cluster analysis, (b) Value-added geological map obtained from fuzzy cluster analysis and the multi-attribute linear feature detection process (cf. Figure 6b); (c) Zonal (classified) map obtained from discriminant analysis using the clustered map shown in (a) for training.

**9:** (a) Difference of clustering/classification observed between the maps shown in Figures 8a and 8c. Approximately 8 % of the samples are ascribed to different clusters/classes, i.e. misclassified. (b) Cluster membership and (c) aposterior classification probability for cluster/class 2 as shown in Figure 8.