De-Hazing Landsat Thematic Mapper Images

J. Lavreau
Laboratoire de télédétection aéro-spatiale, Musée royal de l'Afrique centrale, 3080 Tervuren, Belgium

ABSTRACT: An algorithm using no external parameters is proposed for removing the effect of aerosols from Landsat TM images. It uses the haze or fourth tasseled cap parameter as a measure of aerosol density for each pixel and computes corrected TM values according to that parameter.

INTRODUCTION

PARTICULATES PRESENT IN THE ATMOSPHERE (aerosols) constitute a major problem in remote sensing, not only because they mask the objects being imaged but also because they alter their spectral signature. They are common at all latitudes, but smoke from bush fires and a persistent atmospheric veil during the dry season are occasional but major problems encountered in tropical countries. The methodology and the applications presented are therefore derived for the latter environment.

The atmosphere affects the detected radiance of objects at ground level either on a global basis, i.e., by affecting all ground pixels (IFOVs) in a similar way, or on a local basis, i.e., when the effect varies from one ground pixel to another.

The global effects find their origin in small-scale variations in the nature of the atmosphere in the area of interest. Correction of such effects on digital imagery can be performed through external or internal procedures. The former methods use known or inferred atmospheric parameters (precipitable water, horizontal visibility, season, climate, etc.) subsequently computed through a suitable algorithm in turn applied to an adequate atmospheric model, together with calibration data of the signal recorded by the satellite and conditions of illumination at the time of image acquisition. The latter methods rely on information provided by the image itself and are based on the fact that data recorded in the near infra-red are differently affected by atmospheric effects than those recorded in the visible range.

The method described hereafter is based on internal procedures. It is intended to provide a correction of local effects, whenever due to haze, smoke plumes, clouds, or other aerosol particles locally present between the ground pixel and the sensor and whose spectral signature alters that of the pixels at the ground. We will refer to these aerosols as haze. Two parameters must be computed for the removal of their effects: first, the amount of haze present, and second, the relationship between the former and the actual DN values recorded in the different bands.

QUANTIFYING THE HAZE

Haze (*lato sensu*) is due to particles with a diameter typically in the range 0.3- to 10-μm (with extremes down to 0.01 and up to 100 μm), those larger than 0.06 μm having an effect on the optical state of the atmosphere. Their concentration decreases rapidly with height: about 60 percent are found within 1 km of ground level (Chahine, 1983). As it will be seen below, we are not going to deal with a particular type of aerosol. According to experience gathered so far, the proposed algorithm is effective on some type of clouds or parts of clouds, as well as on smoke plumes and on haze *sensu stricto* as long as the DN value recorded in the spectral band is below saturation. Its effectiveness on larger particles (e.g., windborne dust) has not yet been tested.

Not many methods are available to estimate the amount of haze. Although the extent of an area affected by haze can readily be mapped on the images through appropriate density slicing of one or another spectral band of the visible (in some cases, the infrared as well), it is not possible to determine directly what concentration of particulate or gaseous matter the observed plume represent (Pettyjohn, 1980).

Some discrimination methods are based on information provided by Landsat Thematic Matter (TM) thermal band 6: haze and clouds are indeed readily distinguished from ground pixels as cold objects, but smoke or windborn dust are not. Haze or smoke will thus be classified as very distinct features although they give similar effects. Band 6 can, however, be used to discriminate clouds from haze and smoke by combining with the "haze" parameter described below (Rice and Odenweller, 1989). J. Potter (1984) uses channel correlation while J. Switzer *et al.* (1981) use the covariance between the channels; these methods are, however, applied to Landsat MSS data.

We will favor the "haze" parameter designed by Crist (1984) and refer to it as TC4 (the fourth Tasseled Cap parameter). This TC4 parameter has been derived (Crist et al., 1986) from simulated Landsat 4 TM values (based on field-measured spectra) and the Dave atmospheric model (the difference between "clear" and "hazy' being a five-fold increase in aerosol density). Because most of the information related to the ground pixels of the scene is explained by the first three Tasseled Cap parameters (TC1 = brightness, TC2 = greenness, TC3 = wetness), the direction of the maximum TC4 parameter should provide a means of measuring atmospheric effects without the confounding influence of the ground pixels. TC4 is computed from a combination of the six reflective spectral bands of Landsat 4 or 5 Thematic Mapper sensors. For Landsat 5 it reads (Crist et al., 1986)

$$\begin{aligned} TC4 = {} & 0.8461{*}TM1 - 0.7031{*}TM2 - 0.4640{*}TM3 \\ & - 0.0032{*}TM4 - 0.0492{*}TM5 - 0.0119{*}TM7 + 0.7879. \end{aligned}$$

This fourth parameter is, like the first three, not scene dependent (they were intended as scene normalizing parameters (Crist, 1984)) but depends only on sensor calibration. It is, however, not quantitatively related to some intrinsic quality of the haze like composition or density.

The histogram of the TC4 image (the real values computed for TC4 vary between about 30.0 and about 200.0, so that no extra stretching/compression of the dynamic needs to be performed for displaying the image at 8 bits) is typically bi-modal. The haze-free population yields a sharp-pointed distribution curve with a mode varying according to the general state of the atmosphere, situated at about TC4 = 40, whereas the hazy pixels give rise to a flat distribution curve limited towards the low values by the haze-free population and towards the high values by pixels with saturated values (DN = 255) usually represented by clouds (Figure 1; see also Figure 11 in Crist *et al.* (1986)).

It should be noted that, if the TC4 parameter of Crist quantifies the amount of haze present in a pixel, it doesn't correct an image for aerosols present.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 57, No. 10, October 1991, pp. 1297–1302.

0099-1112/91/5710–1297$03.00/0

RELATIONSHIP BETWEEN TC4 AND TM BANDS

Because the system determined by the haze and the ground relative to the incoming and outcoming radiations implies multiple scattering, the signal recorded from each ground pixel cannot be considered as the arithmetic sum of the radiance contributed by the haze and the radiance contributed by the ground pixel. We will, however, show that this complex relationship can be modeled globally and can be expressed mathematically through a rather simple law of correlation between the TC4 and the TM values.

EXPERIMENTAL DETERMINATION OF THE RELATIONSHIP

Because the variance due to the nature of the ground pixel acts independently from the variance due to the amount of haze, a straightforward computation of the relationship of the TC4 values to the TM values of a whole scene becomes subject to probability considerations (see Potter, 1984). In order to avoid this, we use a simplified data set where only one factor varies (the haze) while the others (TM value "below the haze") remain unchanged. The simplified set has been created by defining a geometric mask (training polygon) covering pixels (about 1000) belonging to a single taxonomic entity and thus displaying (except for the effect of the haze) a limited range of DN values (ideally, it should be only one). Ground pixels in this polygon are thus either haze-free or affected by variable amounts of haze readily visible on TM images in the visible range or on the image of TC4.

We collected data from two regions, both characterized by the presence of some kind of haze covering ground pixels displaying, locally, a narrow range of values: scene Path/Row 172/65-2 (22 July 1984, Karema, lake Tanganyika, Tanzania) and scene 201/39-1 (5 April 1987, Anti-Atlas, Morocco).

The image of Karema (Plate 1) displays a varied savannah landscape characteristic of western Tanzania. Bush fires give rise to smoke plumes extending over the clear waters of lake Tanganyika (Bardinet *et al.*, 1988). The DN values of these lake waters show the desired narrow range (4 to 6 DN according to TM band) of (low to very low) values. A polygon has been created containing only (hazy and non-hazy) lake water, thus excluding the variant DN values on the "continent."

The image of Anti-Atlas displays a semi-desert landscape and a comparatively high variety of rock and soil types. Some plateaux of tabular rocks, however, show the desired narrow (8 to 12 DN according to TM band) range of (rather high) DN values. Haze originates here from the presence of clouds. The polygon covers the thinnest (still transparent) parts of the clouds and different parts of the image display identical/similar rock formations (tabular and, therefore, minimally affected by shades only).

The scatter-plot of the DN values (contained in each polygon) in each spectral band with the corresponding TC4 values will give us an idea of the nature of the global relationship between TM data and haze.

Figure 2 shows that this relationship can be approximated by a linear function of slope A and intercept B. The line of best-fit calculated from the Moroccan image (A in Figure 2), where the range of TC4 values is the largest and the precision of the regression thus the highest, is of the type

$$TMx = A^{*}TC4 + B\text{: i.e.,}$$
$$TM1 = 1.88^{*}TC4 + 18.5,$$
$$TM2 = 0.89^{*}TC4 + 12.9,$$
$$TM3 = 1.02^{*}TC4 + 26.6.$$

Values for the other TM channels, calculated in the same way are

$$TM4 = 0.85^{*}TC4 + 30.1,$$
$$TM5 = 1.40^{*}TC4 + 52.1,$$
$$TM7 = 0.71^{*}TC4 + 41.8.$$

The range of TC4 values on the image of Karema is too small to be used in a regression line computation (B in Figure 2).

The independent factor B is of little significance here as it is

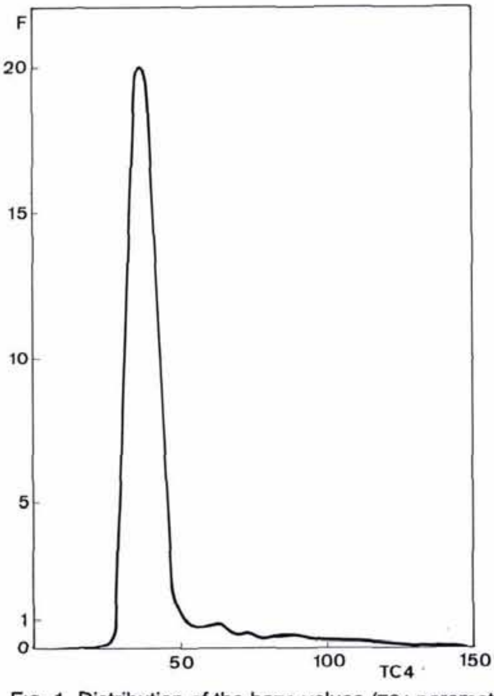

FIG. 1. Distribution of the haze values (TC4 parameter) over the scene of Morocco (F: frequency). TC4 values above 50.0 result from hazy pixels. The mode of non-hazy pixels is about 33.0. No values are observed below 21.0 or above 126.0.

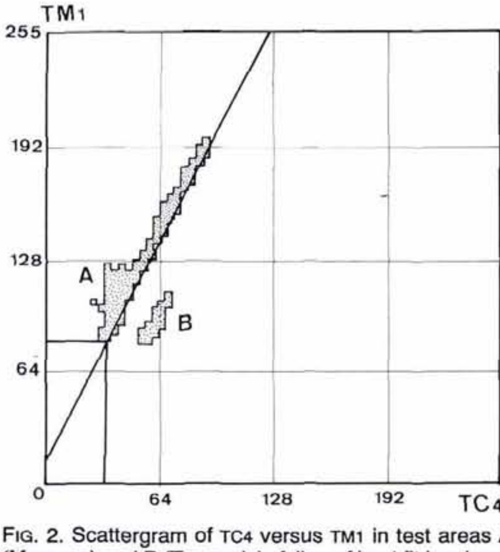

FIG. 2. Scattergram of TC4 versus TM1 in test areas A (Morocco) and B (Tanzania). A line of best fit has been drawn through the lowest DN values, thus allowing the determination of the relationship between haze (TC4) and TM values, subsequently, the correction of the latter. The "regional" threshold value for haze (TC4) appears to be 32.7 in Morocco (see text). Similar graphs can be drawn for TC4 versus the other TM bands.

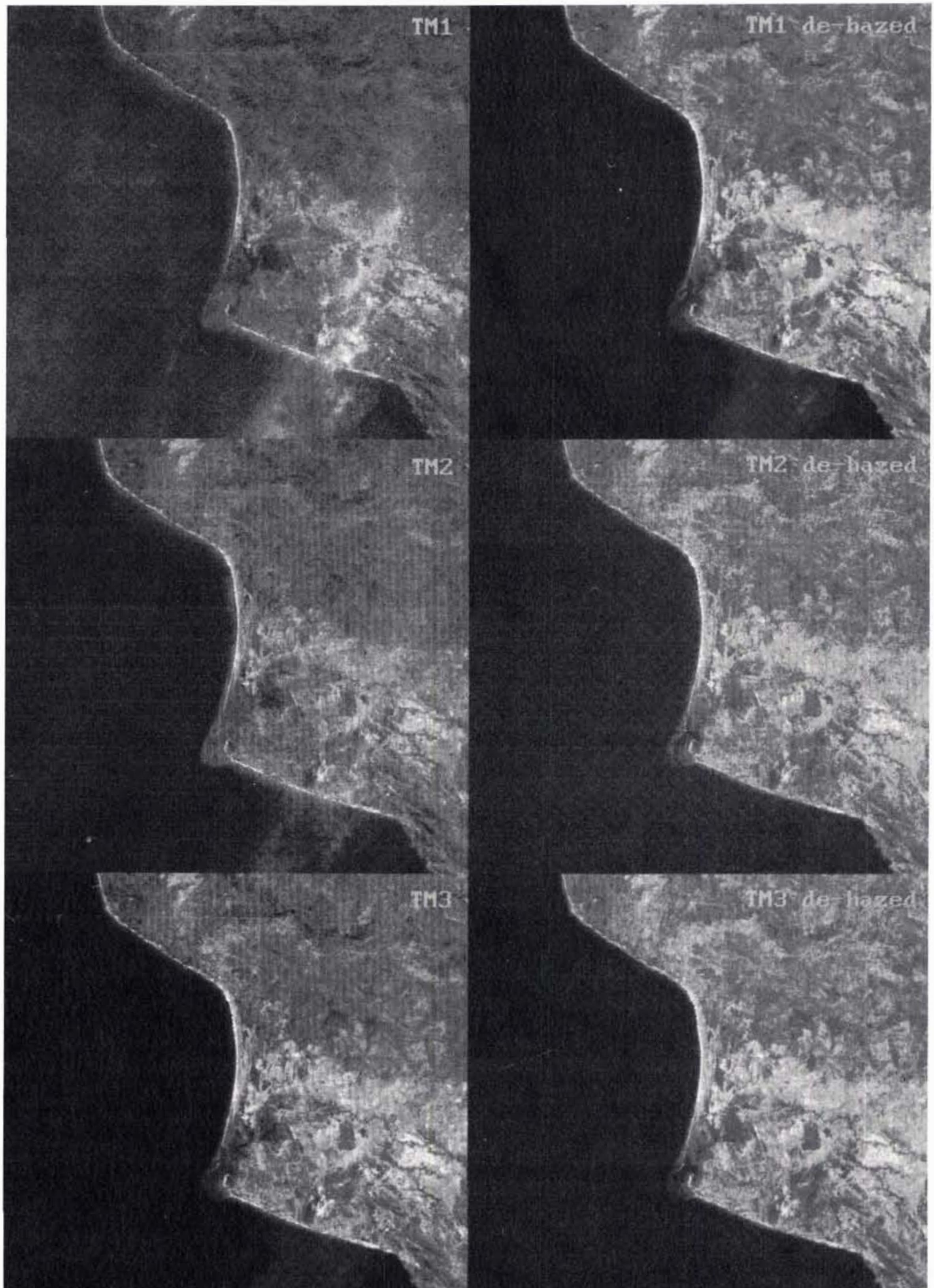

PLATE 1. From top to bottom: TM1, TM2, and TM3 images of the Karema (W. Tanzania) test site. To the left and right, respectively, the original and the haze-corrected image: A haze threshold (see text) of 32.7 has been used here.

only related to the DN values of the arbitrary chosen taxonomic class (training polygon).

CORRECTING A TM IMAGE FOR HAZE

DETERMINATION OF THE CORRECTION FACTOR

Because the relationship between TC4 and TMx is linear and because the TC4 value of a haze-free pixel is not zero, the weight of the correction must be the difference between the TC4 value of the actual pixel and the $TC4_o$ value of a haze-free pixel. This threshold can be determined experimentally. For instance, in the Moroccan scene, the non-hazy ground pixels in the polygon have a DN of respectively 80, 42, 60, 58, 98, and 65 in different TM bands, thus allowing to compute (from the equations above) an average threshold of 32.7 for TC4.

It should, however, be kept in mind that this threshold is scene-dependent because the TC4 parameter also varies according to the general state of the atmosphere; for instance, in another tropical area (Burundi), a value of 58 has been estimated at +1 sigma of the mean on the basis of the TC4 histogram of the whole scene (Lavreau, 1990). Nevertheless, lacking more appropriate data, and because a particular choice of threshold anyway only results in a shift of the resulting corrected values (see however below), the threshold value determined in Morocco has been used in the illustrated application on Karema (Plate 1).

Once a satisfactory haze-threshold $TC4_o$ has been determined, the relationship between any corrected $(TM_{x,i})c$ and its original $TM_{x,i}$ value (x for each spectral band, i for each pixel) can readily be computed from the slope A_x of each band/TC4 relation (Figure 3) : i.e.,

$$(TM_{x,i})c = TM_{x,i} - (TC4_i - TC4_o)\, A_x$$

Non-hazy ground pixels ($TC4_i - TC4_o = 0$) will thus theoretically not be affected. It is also obvious from the presented algorithm that only pixels with non-saturated DN value (i.e., less than 255) can be corrected significantly.

Because TC4 values lower than the actual $TC4_o$ can exist in the image (some variance exists in the TC4 image of a haze-free scene; see Figure 1), as in shaded areas, as well as non-hazy pixels with DN higher than $TC4_o$, as on particularly bright soils, a conditional application of the equation could be applied.

We have found out that using the mode of TC4 as threshold value gives equally good results; DN values lower than $TC4_o$, often due to shaded areas, indeed become brighter whereas sunny slopes become a little darker. Some alteration of the original DN values thus occurs anyway, but the importance of this alteration is small. This should nevertheless be taken into account when using de-hazed data while computing radiances from DN values.

The reasoning applies the other way round : a fixed $TC4_o$ value could be used as a scene normalizing factor both for the general and for the local atmospheric heterogeneity. One could, therefore, attempt to choose a very low threshold, perhaps unrelated to the actual mode of TC4. This would result in a correction of all ground pixels, including the non-hazy ones. This procedure should be avoided if the absolute DN value is of importance while subsequent processing or comparison with unprocessed images.

RESULTS

Plate 1 shows the effect of the de-hazing algorithm on the image of the Karema test site for bands TM1-3 (because the smoke plume is less conspicuous on the TM4-7 images, it has not been displayed).

Some unexpected effects are also visible: most of the striping and image noise present in TM1 and TM2 images have been

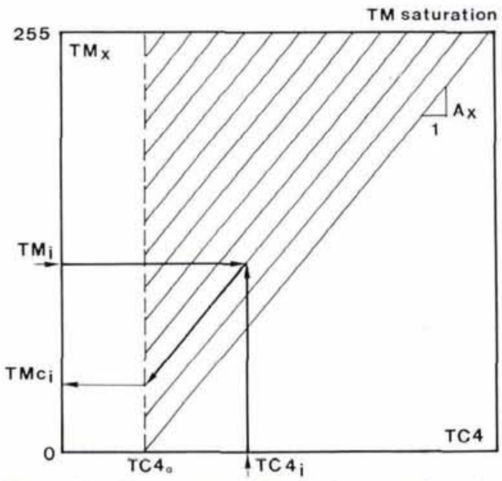

FIG. 3. Graphic representation of the correction algorithm for each pixel. A $TM_{x,i}$ (derived from each spectral band x of Landsat TM) and a $TC4_{x,i}$ (computed from the six spectral bands) pair of values defines a correction line of slope A_x. The corrected value $(TM_{x,i})c$ is given by the intercept of this line with the $TC4_o$ threshold.

removed (Plate 1). The rim of shallow water where the lake bottom is visible as well as the area of high chlorophyll activity (as on the spit of the triangular low-land area) has, however, also been erased. Some excess of correction has thus been made, partly because of a too low value for the haze threshold ($TC4_o$ = 32.7), but partly also because the TC4 parameter probably inadequately quantifies these taxonomic classes as haze.

The same algorithm has been applied to the already quoted area in Burundi, i.e., in an area different from the one where the TC4/TM calibration has been performed. The effect is still more dramatic here, where the haze is represented by small clouds on the one hand, and by a general veil covering the whole image on the other hand. Plate 2 shows that the clouds are almost eliminated whereas the haze veil has disappeared; the shade of the clouds however remains.

The effect of the correction algorithm on haze due to larger particles (dust) has not been tested, appropriate image data lacking.

We have implemented the de-hazing algorithm in our image preprocessing routines. The mode of the TC4 distribution, calculated from a window on a mainly haze-free area, is taken as threshold. Six input images are read and one to six output (corrected) images are created in each batch. Except for the disappearance of haze, no important alteration of the DN values is noticeable. When the corrected images are combined in color composites and compared with the combined uncorrected ones, more important changes appear because the corrected images show a higher degree of correlation. This can be expected because atmospheric effects typically act differently according to wavelength.

CONCLUSION

The haze correction algorithm presented results in a dramatic reduction of the aerosols covering (parts of) a Landsat TM image. Its effectiveness has been tested on clouds, haze, and smoke plumes. Its application needs no external parameters, but some care should be taken while using it in connection with quantitative applications.

REFERENCES

Bardinet, C., J. Lavreau, A. Le Page, I. Roggeri, and Ph. Trefois, 1988. Image analysis and lithological classification of the Karema area,

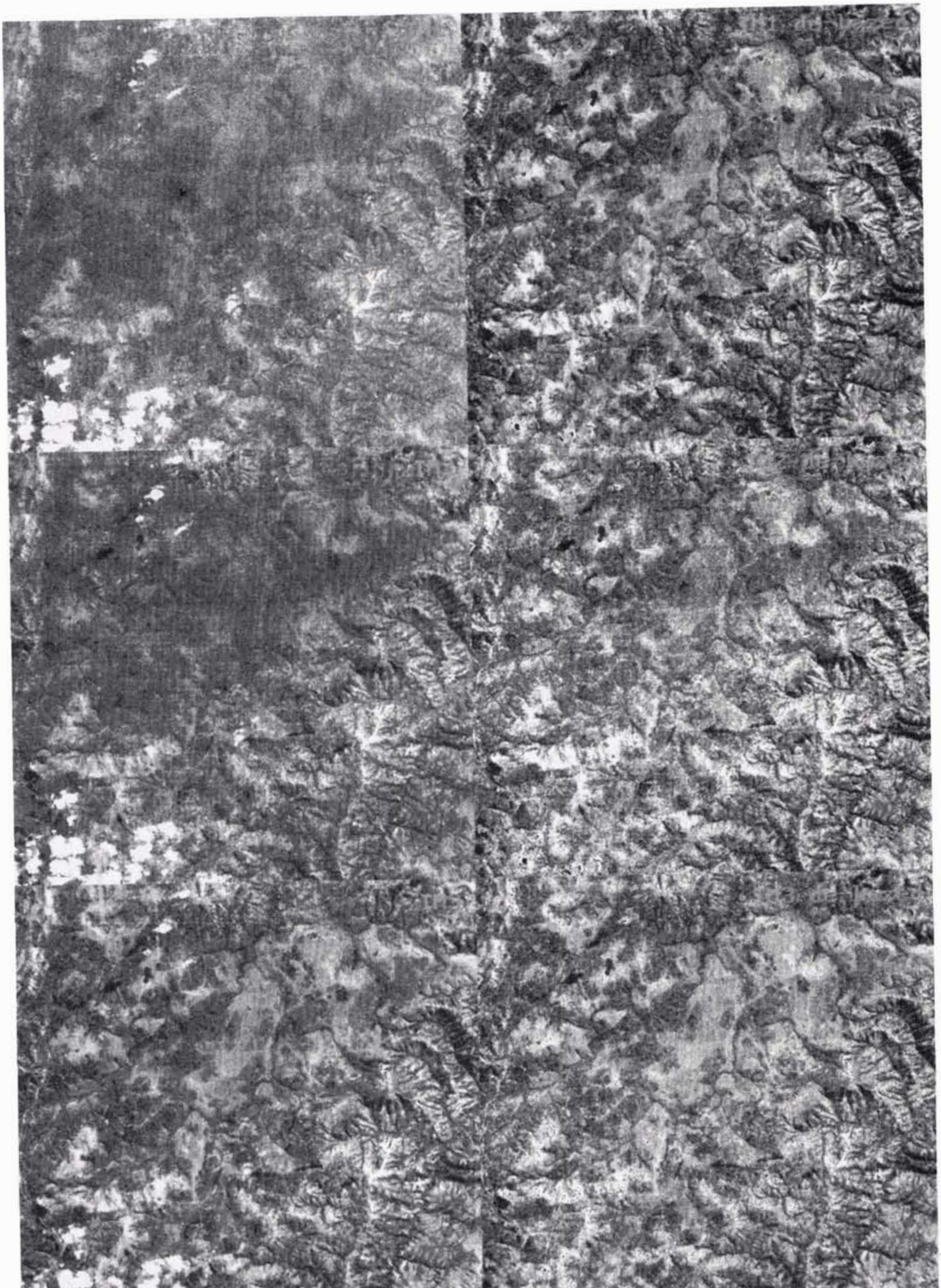

PLATE 2. From top to bottom : TM1, TM2, and TM3 images of Musongati region in S.E. Burundi. To the left and right, respectively, the original and the haze-corrected image. A haze threshold (see text) of 40 has been used here.

Tanzania, using Landsat TM data. *Image Analysis, Geological Control and Radiometric Survey of Landsat TM Data in Tanzania (The Unesco/IUGS GARS Program in Africa)* (J. Lavreau and C. Bardinet, ed.), Musée royal de l'Afrique centrale, Ann. série in 8°, Sc. géol., Tervuren (Belgium), 96, 79–98.

Chahine, M. T., 1983. Interaction within the atmosphere. (R.N. Colwell, ed.), *Manual of Remote Sensing*, vol. I, Am. Soc. Photogrammetry, Falls Church., 1232 p.

Crist, E. P., 1984. A spectral haze diagnostic feature for normalizing Landsat Thematic Mapper data. *Proceedings of the 18th International Symposium on Remote Sensing of Environment* (Paris), ERIM, Ann Harbor, Michigan, pp. 735–744.

Crist, E. P., R. Laurin, and R. C. Cicone, 1986. Vegetation and soils information contained in transformed Thematic Mapper data. *Proceedings of IGARSS '86 Symposium* (Zurich), European Space Agency SP-254, Paris, pp. 1465–1470.

Lavreau, J., 1990. Exploration de l'espace spectral d'une sous-scène de Thematic Mapper au Burundi (région de Musongati). Recherche du plan des sols. Géologie (*Bull. Soc. belge de Géologie*), Bruxelles, 99, 21–40.

Pettyjohn, W. A., 1980. Environmental geoscience. *Remote Sensing in Geology*, (B. S. and A. R. Gillespie, ed.), J. Wiley & Sons, New York, 702 p.

Potter, J. F., 1984. The channel correlation method for estimating aerosol levels from Multispectral Scanner data. *Photogrammetric Engineering and Remote Sensing*, 50, 43–52.

Rice, D. P., and J. B. Odenweller, 1990. External effects correction of Landsat Thematic Mapper data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 4, 355–368.

Switzer, P., W. S. Kowalik, and R. J. Lyon, 1981. Estimation of atmospheric path radiance by the covariance matrix method. *Photogrammetric Engineering and Remote Sensing*, 47, 1469–1476.

(Received 15 August 1990; revised and accepted 18 December 1990)