Reprinted from

Eighth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring

July 7-9, 1982

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

SPATIAL AND SPECTRAL SIMULATION OF LANDSAT IMAGES OF AGRICULTURAL AREAS

W.F. PONT, JR.

Environmental Research Institute
of Michigan
Ann Arbor, Michigan

I. ABSTRACT

A Landsat scene simulation capability was developed. The simulation employs a pattern of ground polygons each with a crop ID, planting date, and scale factor. Historical Greenness/Brightness crop development profiles generate the mean signal values for each polygon. Historical within-field covariances add texture to pixels in each polygon. The planting dates and scale factors create between-field/within-crop variation. Between field and crop variation is achieved by the above and crop profile differences. The Landsat point spread function is used to add correlation between nearby pixels. The next effect of the point spread function is to blur the image. Mixed pixels and misregistration are also simulated.

The simulation has been used to study the effects of small fields and misregistration on current Landsat-based crop proportion estimation procedures.

II. INTRODUCTION

The signal which the Landsat multispectral scanner generates is a function of many variables, few of which we have any control over. The ideal method of understanding a process is to hold all of the variables constant, except those under consideration. This method fails for the most part in the study of the Landsat signal-generation process with its seeming contradiction of vast amounts of data at the pixel level but a scarcity of data with unique combinations of factors such as scan angle, day of year, crop, field pattern, etc. Simulation is a tool which allows one to use combinations of assumed or known effects to infer the composite effect. The uses of a simulation include:

(1) The study of the interaction of known first order effects;

(2) Tests of procedures on data generated under known conditions, and

(3) Empirical estimation of model parameters when fitted to "real data".

The major motivation for the simulation model described here was the need for a capability to investigate, in detail, the effects of various factors on pixel values from small fields, boundaries between fields, and misregistered pixels. Both spectral and spatial properties were of interest. With this model any desired polygonal field pattern can be simulated and spectral-temporal characteristics can differ from field to field, even within a single crop type, and with within-field variances being included.

III. THE MODEL

Consider the point (x,y) on the ground at time t. Except for a set of area zero, (x,y) will be contained in the interior of a field.* Denote this field as k. The main effect which a sensor could detect is that of the crop at point (x,y). We denote the crop in field k as C_k. We use crop development profiles in Greenness and Brightness to simulate the mean crop response as a function of time since planting. Reference 1 gives the empirically estimated profiles used, while Figure 1 illustrates those for corn, soybeans, small grains, pasture, etc.

---

*e.g., a hedgerow between agricultural fields is, itself, considered to be a field of finite area.

Denote the profile for crop c as $P_c(.)$. Note that two fields with the same crop would not in general have the same profile value at time t due to different planting days. Denote the planting date for field k as T_k. The model further assumes that there are field effects beyond crop type and planting date due to soil characteristics, crop variety, fertilizer, etc. These additional between-field, within-crop sources of variability are viewed as geometric noise factors which scale each profile. Denote the scale factor for field k as U_k, where U_k is a random variable with a mean of 1. The profile at (x,y) is:

$$g(x,y,t) := U_k P_{ck}(t-T_k)+\varepsilon_{txy}$$

where ε_{txy} is assumed to be a bivariate normal with mean of zero. The model assumes that the covariance of ε_{txy} is a function of crop and time. This is reasonable if the dominant effect in within-field variation is due to crop-field effects. If sensor noise were the real dominant effect, then variances of the Landsat Bands 4, 5, and 6 would be proportional to the signal and the variance would be constant in Band 7.

One of the major problems encountered in multitemporal Landsat data is spatial misregistration between dates. The coordinate system changes between passes of the satellite. The point (x,y) in the satellite's coordinate system does not correspond to the same ground point for different passes. The relationship between ground coordinate system and that of the sensor's is non-linear. There are registration procedures which reduce the differences in coordinate systems; however, there is always a residual error in registration procedures. The model assumes the sensor coordinate system changes only by a translation between passes. If the ground coordinates are (x,y) then the sensor's coordinates at time t are $(x+x_t, y+y_t)$.

This form of misregistration is suitable for most applications using simulation. A more general form of misregistration could be simulated by warping the coordinates which define the fields.

The signal which the sensor receives is not g(x,y,t) but rather

$$f(x,y,t) = \iint g(x+x_t-r, y+y_t-s, t)p(r,s)drds$$

where p is the Landsat point spread function. p was derived in Reference 2 using the sensor's size, blur circle and properties of its three-pole Butterworth filter. Figure 2 gives a three-dimensional drawing of p and Figure 3 gives plots of p along the scan line and along track, at pixel center. The signals which the sensor allows us to observe are:

$$\{f(x+idx, y+jdy, t)\}_{\substack{i=1,N_x \\ j=1,N_y}}$$

Values for a 5x6-mile AgRISTARS segment are dx = 79M, dy = 57M, N_x = 196, and N_y = 117.

IV. IMPLEMENTATION

A. THE FIELD GEOMETRY

Each field is stored in the computer as a polygon. The vertices of all of the fields are contained in arrays, say $\{U_{kj}, V_{kj}\}$. Polygon (field) k is defined by the vertices $k_1, k_2, \ldots, k_{N_k}$, such that the points $\{U_{kj}, V_{kj}\}_{j=1,N_k}$ circumscribe field k in a counterclockwise direction. It is important that there be no gaps in adjacent fields and non-nil intersections can cause unexpected results. We assume that all fields are simply connected, but more general sets could be incorporated into the model easily.

A two-dimensional grid of points is assigned polygon identification. The point (x,y) is assigned to the first polygon whose winding number is positive. We view these points as subpixels. The polygon search begins with the polygon which contained the previous pixel. If only translation misregistration is to be simulated then this subpixel-to-field assignment only has to be performed once. If more general misregistration is to be simulated then the points $\{U_i, V_i\}$ can be replaced by $\{H_t(U_i, V_i)\}$ where H_t is the warping transform for time t. Examples of H_t are:

$$H_t(U,V) = \left(\sum_{q=0}^{5}\sum_{j=0}^{q} a_{qj}U^jV^{q-j},\ \sum_{q=0}^{5}\sum_{j=0}^{q} b_{qj}U^jV^{q-j}\right) \quad (1)$$

and

$$H_t(Z) = A_t(Z-Z_t)+Z_t \quad (2)$$

where

$$Z = u+vi,\ Z_t = u_t+v_ti$$

and

$$A_t = R_t e^{i\theta_t}$$

Functions of the form (1) are often used to correct geometric distortions in Landsat data (see Reference 3). Regression methods are often used to estimate the coefficients a_{qj}'s and b_{qj}'s. Since there are 21 terms in each coordinate of (1) there should be somewhat more than 21 control points used in the estimation, if estimates of all coefficients are desired. Stepwise regression methods tend to have good results with 5-9 control points. Functions of the form (2) represent a rotation of θ_t and a scaling by R_t about (u_t, v_t).

B. CROP RESPONSE AS A FUNCTION OF TIME AND FIELD

The crop for point (x,y) on the ground at time t is:

$$g(x,y,t) = U_k P_{ck}(t-T_k)+\varepsilon_{txy}$$

where

k is the field containing (x,y),

U_k is the scale factor for field k,

C_k is the crop growing in field k,

T_k is the time of planting,

$P_c(.)$ is the Greenness/Brightness response of crop c as function of time since planting, and

ε_{txy} is the within-field noise.

The polygon specific parameters U_k, C_k and T_k are saved in a file until all acquisitions are generated. U_k and T_k are viewed as random variables such that $E\{U_k\} = 1$ and the distribution of T_k is obtained from a crop calendar specific to the region being simulated. Empirical profiles were incorporated for grain, sunflower, corn, soybeans, and three types of grass/pasture/hay. New profiles can be added or old ones modified easily.

Presently the within-field error term is used only to add texture to the pixels contained in a given field. Data which would support an accurate estimation of the covariance matrix of ε_{txy} do not exist. The reason is that ground truth polygons often contain more than one field with the same ground truth code, while the field-finding algorithms are constrained to construct field-like regions with small within-field variances.

C. THE CONVOLUTION

The convolution of the sensor's point spread function blurs the image by adding correlations between nearby pixels. The sensor's response at point (x,y) and at time t is:

$$f(x,y,t) = \iint g(x-r,y-s,t)p(r,s)drds.$$

We use two different levels of approximations of f(x,y,t):

$$f_1(x,y,t) = \sum_{i=-16}^{48} \sum_{j=-16}^{16} g(x-\tfrac{i}{16}, y-\tfrac{j}{16},t)p_1(\tfrac{i}{16},\tfrac{j}{16})$$

where

$$p_1(\tfrac{i}{16},\tfrac{j}{16}) = \frac{p(\tfrac{i}{16},\tfrac{j}{16})}{\sum_{r=-16}^{48} \sum_{s=-16}^{16} p(\tfrac{i}{16},\tfrac{j}{16})};$$

and

$$f_2(x,y,t) = \sum_{i=-4}^{16} \sum_{j=-4}^{4} g)x-\tfrac{i}{4},y-\tfrac{j}{4},t)p_2(\tfrac{i}{4},\tfrac{j}{4})$$

where

$$p_2(\tfrac{i}{4},\tfrac{j}{4}) = \frac{p(\tfrac{i}{4},\tfrac{j}{4})}{\sum_{r=-4}^{16} \sum_{s=-4}^{4} p(\tfrac{r}{4},\tfrac{s}{4})}.$$

V. SUMMARY

The present understanding of several components in the Landsat signal-generation process allows the simulation of Landsat data. The simulation described in this paper allows for:

(1) Mixed pixels,

(2) Field geometry,

(3) Landsat point spread function,

(4) Crop development spectral profiles,

(5) Variation in planting dates,

(6) Within-field variation, and

(7) Misregistration.

The simulation has been used in small field research. Other applications include the simulation of other sensors, the test of new procedures, and the study of new crop mixes and field patterns.

REFERENCES

1. Crist, E.P. and W.A. Malila, Development and evaluation of an automatic labeling technique for spring small grains, Environmental Research Institute of Michigan, Ann Arbor, Michigan, Report No. SR-E1-04065, June 1981.

2. Dye, R.H., Restoration of Landsat images by discrete two-dimensional deconvolution. Presented at the Proceedings of the Tenth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, October 1975.

3. Wilson, C.L., Landsat derived maps of poorly mapped areas. Presented at the Third Annual International Users' Conference on Computer Mapping Hardware, Software and Data Bases, Cambridge, Massachusetts, July 1980.

Note: This work was sponsored under contract NAS9-15476 by the U.S. National Aeronautics and Space Administration, NASA Johnson Space Center, Houston, Texas 77058.

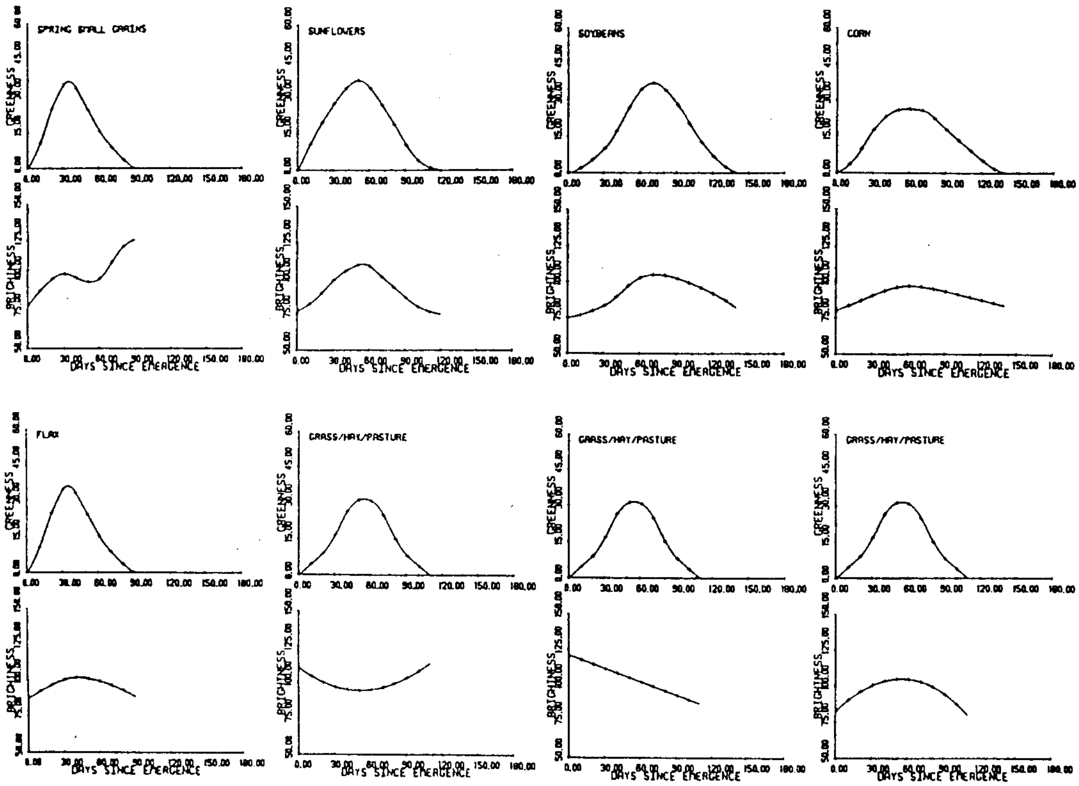


FIGURE 1. GREENNESS/BRIGHTNESS CROP PROFILES

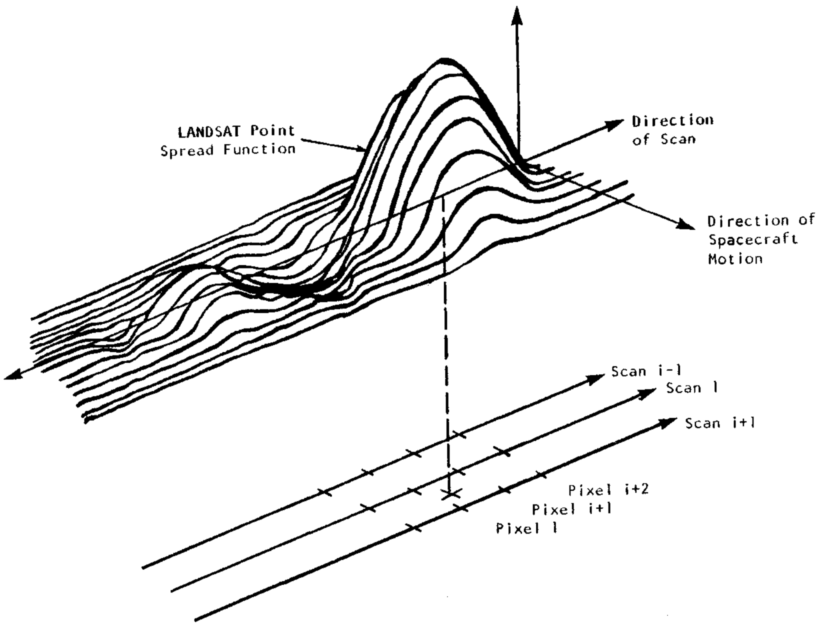
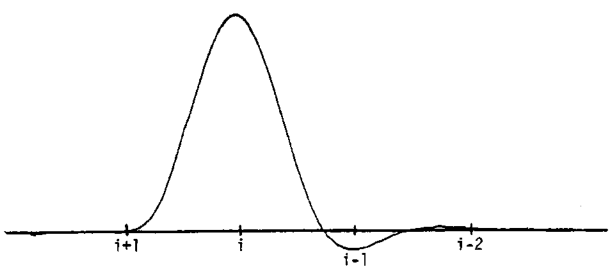
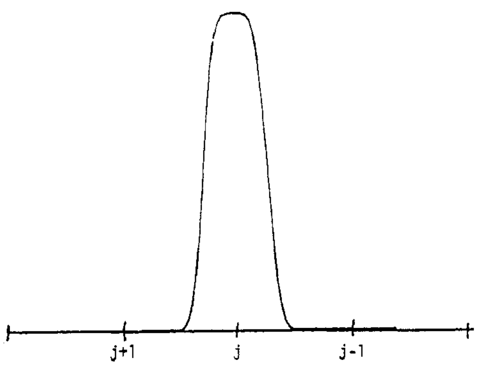


FIGURE 2. LANDSAT'S POINT SPREAD FUNCTION



(a) Landsat Along Scan Line Point Spread Function



(b) Landsat Along Track Point Spread Function

FIGURE 3. LANDSAT MARGINAL POINT SPREAD FUNCTIONS

W. Frank Pont, Research Mathematician, received the B.S. degree in Forestry from Idaho State University in 1973, M.S. degree in Statistics from Michigan State University in 1977, and subsequently carried out graduate research towards the Ph.D. degree in Statistics. Since joining ERIM in 1979, he has participated in research in remote sensing.