Spatial Thresholds, Image-Objects, and Upscaling: A Multiscale Evaluation

G. J. Hay,* K. O. Niemann,* and D. G. Goodenough†

When examining a remotely sensed signal through various scale changes, what is the most appropriate upscaling technique to represent this signal at different scales? And how can this be validated? Solutions to these questions were approached by examining how the 660 nm signal of six forest stands vary through four different scales of same-sensor imagery, four traditional resampling techniques, and a new object-specific resampling technique. Analysis of the original and modeled datasets suggests that appropriately upscaled imagery represents a more accurate scene-model than an image obtained at the upscaled resolution. Results further indicate the need for a multiscale approach to feature extraction and upscaling, as no single spatial resolution of imagery appears optimal for detecting or upscaling the varying sized, shaped, and spatially distributed objects within a scene. By employing the human eye as a model, we describe a novel object-specific approach for addressing this challenge. Upscaling evaluation is based on visual interpretation, an understanding of the applied resampling theories, and the root mean square error results of 6000 samples collected from a 10 m CASI scene, and from 1.5 m, 3 m, and 5 m same site CASI images upscaled to 10 m. Potential application of this object-specific approach in hierarchical ecosystem modeling is also briefly described. ©Elsevier Science Inc., 1997.

INTRODUCTION

In an effort to better understand and predict how our planet functions as an integrated system, biophysical and ecological information derived at the scale of the individual organism must be extrapolated to the regional and global scales of climate models (Wessman, 1992). As solutions for integrating multiscale data and models are sought (Turner et al., 1989), scale-related concepts are rising to the forefront in earth system sciences where they have been discussed in biology, geography, geomorphology, hydrology, landscape ecology, and meteorology (Clark, 1990; Turner et al., 1991; Lam and Quattrochi, 1992). While a single ubiquitous definition of scale does not exist, some 15 distinct meanings do appear in the Oxford English Dictionary, where several more readily lend themselves to computation and hypothesis testing than others. One definition which can be applied to the time, space, and mass component of any quantity is that scale denotes the resolution within the range of a measured quantity (Schneider, 1994). This is an effective definition as it encompasses two very important, interacting facets of scale: resolution and range. Resolution or "grain" refers to the finest distinction that can be made in an observation set, while range or "extent" refers to the span of all entities that can be detected in the data (Allen and Hoekstra, 1991).

In remote sensing, a fundamental characteristic of an image is the spatial resolution, or the size of the area on the ground from which the measurements that compose the image are derived. In this sense, spatial resolution is analogous to the scale of the observations (Woodcock and Strahler, 1987) and the two terms will be considered synonymous throughout this paper. Similarly, the adjectives fine and small will be used in conjunction with scale to represent microscopic areas; coarse and large will be used to refer to macroscopic spatial extents, and high (H)- and low (L)-resolution will be used as described by Strahler et al. (1986): H-resolution refers to the cells of an image being smaller than the real-world elements composing the scene; thus the elements (i.e., tree crowns) may be individually resolved within the image, while L-resolution refers to the cells of the image...
being larger than the elements, and therefore unresolvable. We also define the term image-objects (Hay et al., 1996) as individually resolvable entities within an image that are perceptually generated from H-resolution pixel groups, where each group possesses an intrinsic size, shape, and geographic relationship with real-world scene components.

Within remote sensing literature, scale-related topics include spatial structure and sampling design (McGwire et al., 1993), pattern and statistical scale approaches (King, 1991; Turner et al., 1991), spatial resolution strategies (Forshaw et al., 1983; Townshend and Justice, 1988), scale and information measures (Woodcock and Strahler, 1987; Jupp et al., 1988; 1989; Raffy, 1994; Wieczorek, 1992), classification accuracy (Irons et al., 1985; Marceau et al., 1994a,b), and upsampling (Aman et al., 1992). Here, the term upsampling refers to resampling techniques designed to transform an image collected at a high spatial resolution (Strahler et al., 1996), to a lower spatial resolution representation of the same image. Ideally, such transformations are intended to reduce the data size of an H-resolution image while maintaining its inherent information content at a lower spatial resolution. Unfortunately, this goal is somewhat contradictory as upsampling involves generalization techniques which tend to produce more homogeneous, variance reduced, thus inherently lower information content datasets (Wieczorek, 1992).

While a host of different resampling or scaling techniques exist in the literature (Menemenley and Box, 1987; Turner et al., 1991), few are readily available to users, and very little instruction regarding the appropriateness of these techniques to different remote sensing data types exists. In many cases, the only available resampling algorithms for remotely sensed data have been directly copied from classic image processing interpolation techniques and included in commercial remote sensing image analysis packages, even though they are not theoretically developed for remotely sensed imagery (Moreno and MeBra, 1994). This often leads to an inappropriate use of these routines by inexperienced users, similar to that seen with computer statistical software and novice users. In fact, as no proven theory of spatial scaling exists (Schneider, 1994) the opportunity to produce large volumes of impressive nonrepresentative data is enormous. A classic example includes the upsampling routinely performed on Advanced Very High Resolution Radiometer (AVHRR) data, where pixels representing an integrated spectral response from a nominal 1 km² areal extent are aggregated to 20 km² pixels, for weekly use as a global vegetation index (Justice et al., 1989).

In a broad context, scaling requires the identification of process nonlinearities with change in scale, the range in scales where linearity may hold, and the properties that may be coherent between scales (Wessman, 1992). While it is inaccurate to state that a process is restricted to any particular scale, it is possible to point to specific time and space scales at which one process prevails over another (Schneider, 1994). Allen and Starr (1982), O'Neill et al. (1986), Holling (1992), and others have suggested that a landscape forms a hierarchy which contains breaks in object sizes, object proximities, and textures at particular scales, resulting from (different) structuring processes exerting their influence over defined ranges of scale. Several remote sensing studies support these claims (Borrough, 1981; Clark, 1990; Marceau et al., 1994a,b), and have suggested that the landscape possesses critical thresholds at which ecological processes change qualitatively (Puech, 1994; Sonnian, 1994).

Unfortunately, as spatial scales vary, spectral variations captured by sensors differ nonlinearly with a trajectory that is incompletely understood, often resulting in substantial scaling error when integrating data and models from different disciplines, time, and spatial scales (Gardner et al., 1982; King, 1991; Rastetter et al., 1992). Kimes et al. (1993) have explored the reflectance errors associated with spatial averaging by using a simulated scene derived from high-quality bidirectional reflectance measurements that have been processed into grid cell measures of hemispherical reflectance. Results indicate that errors could be restricted to the 5–10% range required for use in global climate models if homogeneous regions within each cell could be identified, and the resulting estimates weighted by the proportional area of each cover type. One of their greatest challenges lies in being able to divide the scene into “homogeneous” patches of ground cover for which detailed angular measurements are available.

Though originating from a completely different theoretical approach than that described by Kimes et al. (1993), this article describes a novel variance-weighted upsampling technique—based on a model of the human eye (Hay, 1993)—that is capable of identifying homogeneous regions within an image by defining multiscale spatial thresholds at which the spectral variance of image-objects are scale-dependent, and using these scale dependent spatial measures as weighting functions so that fine scale data can be representatively modeled within a user-defined upscaled image. This technique is also able to automatically generate an image that models the maximum multiscale information within the scene, and an image that models multiscale scene components at their individually unique next coarser scales, with application for any scale of imagery. The theory behind this technique is based on an understanding of the relationship between multiscale image variance and high-resolution image-objects, and will be further discussed in the Methodology section.

The primary objectives of this study result from asking the following questions: When examining a signal through various scale changes, what is the most appropriate upsampling technique to represent this signal at different scales? And how can this be validated? Solutions
to these questions were approached by examining how the 660 nm signal (wavelength center ±5 nm) of six forest stands vary through four different scales of same-sensor imagery and five resampling techniques.

**IMAGE, GENERATION, SELECTION, AND PROCESSING**

Though dramatically smaller than the terabytes of imagery expected daily from the Earth Observing System platforms, the 1.03 GB dataset collected for this study offers substantial storage, retrieval, manipulation, and analytical challenges to the user. This section summarizes the procedures involved in image acquisition, study site and wavelength selection, and the application of atmospheric and geometric corrections.

**Image Acquisition**

By varying flying height, ground speed, and sensor integration times (Table 1), 11 multiscale Compact Airborne Spectrographic Imager (CASI) band sets, each composed of eight channels totaling some 1.03 GB were collected during 20:10-21:40 h (GMT) over the Sooke Watershed, British Columbia, Canada on 1 August 1993 (Fig. 1). Data were collected over both north-south and south-north flight paths.

The CASI is a pushbroom sensor designed to operate from light aircraft and helicopters, with data capture capabilities based on a two-dimensional (2D) frame transfer CCD array. This 2D array allows the sensor to function as both a multispectral imager (spatial mode), and an imaging spectrometer (spectral mode) sensitive in the visible and near-infrared portions of the electromagnetic spectrum (418-926 nm). In both modes, the system offers user programmable spectral band sets (in terms of wavelength and width) with a sampling interval of 1.8 nm and a 12-bit dynamic range (padded to 16 bit). In spatial mode, the sensor provides a maximum capability of 15 programmable bands, while spectral mode offers a maximum of 288 bands in up to 39 different viewing directions across the swath (Gower et al., 1992). Since these data were acquired, an enhanced CASI instrument has become available with up to 101 look directions, and irradiance probes on the roof and belly of the aircraft capable of collecting up- and downwelling hemispherical radiance measurements in addition to normal image acquisition. Data from the downwelling probe enable conversion to image reflectance, while the two probes in conjunction provide the potential for the direct measurement of at-sensor hemispherical reflectance (McDermid, 1995).

Unfortunately, data acquisition over this site was plagued by equipment difficulties and conditions of severe wind and turbulence resulting in image smear and distortions during aircraft rolls greater than ±10° from horizontal (0.1° precision). Data were further compromised by a malfunction in the aircraft pitch gyro which removed the option for pitch correction during processing. Imagery were able to be corrected for yaw and roll errors (less than ±10°), and sensor-inherent radiometric errors. With the addition of differential Global Positioning System data, a first-order geometric correction (Zacharias et al., 1994) was applied to the multiscale imagery, providing for resampling into north-up square pixels at four spatial resolutions: 1.5 m², 3 m², 5 m², and 10 m² (Table 1).

**Wavelength and Site Selection**

As the dimensionality, spatial resolution, and volume of datasets increase, routine image analysis becomes increasingly more complex and time-consuming. Even when working on a dedicated workstation with 64 MB of RAM, a 21-in. monitor, and a 4 GB storage medium, the task of visually locating a plausible study site within the swath proved to be nontrivial. In total, 11 different datasets (each 8 channels deep) ranging from 500 pixels×800 lines (9.9 meg) to 3334 pixels×6000 lines (327 meg) required visual evaluation before a location with minimal roll error through all scales of imagery could be selected.

Although the finest resolution imagery covered a swath path no larger than 9.0 km×0.6 km, its sheer digital size (327 meg at 1.5 m spatial resolution) necessitated the selection of a smaller subimage for analysis. After extensive scrutiny, a single channel, 1200 pixel×900 line north–south subimage was selected and extracted from the 1.5 m band set. Its corresponding location and wavelength were also extracted within the north–south 3-m.
5-m, and 10-m band sets, resulting in four images covering the same location at different spatial resolutions, but with equivalent spectral resolutions (wavelengths). A CASI band centered at 660 nm (±5 nm) was chosen for analysis as it represented both the minimum chlorophyll a reflectance signal and the absorption maximum of solvated chlorophyll a (Kirk et al., 1978). It should be noted that while the low trough in actual spectra associated with plants appears closer to 675 nm rather than the 655-665 nm as used, the selection of band widths and locations was limited to those predetermined for a prior mission.

**Geometric Correction**

Once corresponding subimages were isolated and extracted from all four scales of imagery, the 10 m subimage was geocorrected to imported road and forest polygons vectors (with an inherent locational error of ±10 m). The remaining 5 m, 3 m and 1.5 m subimages were then geometrically corrected to this new 10 m image, as visual cues for ground control point (GCP) selection were more abundant here than in the vector data. In all channels, the locational error after resampling with the nearest neighbor algorithm—as the original DNs remain unaltered (Jensen,
Class 1  Class 2  Class 3  Class 4  Class 5  Class 6
Description Mature Mature Young Young Immature Immature
Spatial Pattern Clumpy Clumpy Uniform Mottled Smooth Smooth
Dominant Species DF DF DF DF DF DF
Age (yrs) 141 - 250 141 - 250 33 23 17 17
Height (m) 55.5 - 64.4 37.5 - 46.4 19.5 - 28.4 10.5 - 19.4 10.5 - 19.4 0 - 10.4
Canopy Diameter (m) 6 - 10 6 - 10 3 - 5 2 - 4 1 - 3 1 - 3
Crown Closure (%) 46 - 55 56 - 65 66 - 75 16 - 25 46 - 55 26 - 35

Figure 2. Description of Six Forest Class Characteristics.

Atmospheric Correction
After examining the entire multiscale band set and being satisfied that all wavelengths behaved according to known scattering theory (Slater, 1980), a first-order atmospheric correction was individually applied to each of the four channels in the form of a dark-body extraction technique (Avery and Berlin, 1992). Once applied, each of the four channels were resampled to a common 1.5 m spatial resolution (using the nearest-neighbor algorithm) and then combined within a single dataset to facilitate the extraction of spatially explicit samples. This resulted in a dataset composed of four geometrically and atmospherically corrected, same-wavelength channels, each of which visually models the scene from different flying heights, while being digitally modeled at the same spatial resolution (i.e., 1.5 m). Since the L-resolution data were resampled into a finer scale with nearest-neighbor, the DNs were not compromised; that is, a 3 m×3 m pixel would now be represented by four 1.5 m×1.5 m pixels, each containing the same DN. These geometrically and atmospherically corrected channels were then considered accurate representations of the scene, and all subsequent analysis was performed on them.

Due to the pushbroom nature of the CASI sensor, path radiance effects, that is, the light scattered by the atmosphere into the sensor’s field of view without being reflected from the surface, tend to produce a brightening of the across-track DNs located at the image edges due to the longer path (Kaufman, 1989). While newly devel-
opened software were available to correct for this effect (Niemann and Sykes, 1995), none were applied to the data, as two (across-track) transects sampled over all four corrected scales appeared with minimal curvature when modeled by a second-order polynomial, indicating limited path radiance effects.

**METHODOLOGY**

The task of assessing the most appropriate resampling technique to represent a 660 nm signal at different scales was approached in two steps. The first involved applying five different resampling strategies to three scales of imagery (1.5 m, 3 m, and 5 m), which produced 15 unique images, each of which were upsampled to a 10 m spatial resolution. The second step required determining the root mean square error (RMSE) of samples extracted from six different forest classes within the modeled data (upscaled) and those obtained from equivalent locations within the original 10 m imagery. The upsampled image with the lowest RMSE was then considered the most accurate upscaling routine of those tested, as it generated
DNs with values most like the original coarse resolution image. The first of the following two subsections describes the traditional resampling algorithms used, while the second describes the theory and application of the variance-weighted upscaling routine.

Traditional Resampling Algorithms

Nearest-neighbor, bilinear interpolation, and cubic convolution were available in a commercial image processing package, while nonoverlapping averaging was written for this study.

- In nearest-neighbor (NN), or zero-order interpolation, the DN of the pixel closest to the location of the original input pixel is assigned to the DN value at the output pixels location.
- In bilinear (BIL), or first-order interpolation, a DN is assigned to an output pixel by interpolating DNs in two orthogonal directions within the input image. Essentially, a plane is fit to the four pixel values nearest the location of the pixel in the input image; then a new output DN is computed based on the weighted distances to these points.
- In cubic convolution (CC), resampling occurs in much the same manner as bilinear interpolation, except that the weighted values of 16 pixels surrounding the location of the pixel in the input image are used to determine the value of the output pixel (Jensen, 1986).
- In nonoverlapping averaging (AVG), the mean of the DNs within a nonoverlapping square window (beginning at the origin) are calculated and assigned to the first location in the output image. The window in the input image is then moved a distance equivalent to the specified window size, a new mean is calculated, assigned to the next location in the output image, and the kernel is iterated until the original image has been completely sampled.

Variance-Weighted Upscaling: Theory and Application

Allen and Hoekstra (1991) suggest that scale is not a property of nature alone but, rather, is something associated with observation and analysis and that the scale of a process is fixed only once the actors in the system are specified by the observer. But how do we define these actors in a pixelated image void of topology? And what happens when the actors are themselves multiscale, and interact in a nonlinear fashion? These are not trivial queries and have important implications for the development and linking of hierarchical ecosystem models driven by multiscale remotely sensed data. It is also critical to recognize that spatial phenomena cannot be studied independently of the sampling system used to detect and measure them, as modifications to the sampling frame induce changes, both in the phenomena themselves and their subsequent interpretations (Marceau et al., 1994a).

To satisfy both the sampling dependent nature and multiscale variability of spatial phenomena, we have chosen image-objects to be our vehicle for defining scale, and employ heuristics based on the mechanics of the human eye and multiscale image variability to define and upscale them. The following section briefly describes the theory and methods developed to achieve this.

By employing the human eye as a model (Hay, 1993) we propose an upscaling routine that uses variable sized object-specific windows to analyze and incorporate the influence of different sized, shaped, and spatially distributed objects within the upscale image, rather than upscaling based exclusively within an arbitrary fixed sized window. We recognize that when a sensor views a scene, the recorded radiance represents an integration of the spectral reflectance characteristics of the corresponding objects in the scene convolved within the spatial resolution of the sensor (Jupp et al., 1988: 1989). To model these effects within an upscaling window, we use image-objects to model real-world objects; the spatial extent of the upscaling kernel to model the sensor’s static instantaneous field of view (IFOV); and an area weighting scheme based on the unique spatial characteristics of image-objects to model the spectral influence of neighboring cells [i.e., the adjacency effect (Kaufman, 1989)].

The variable object-specific basis of this approach is far from new. In the early 1980s, David Marr’s (1982) pioneering theories on image processing and the human visual system indicated that intensity changes occur at different scales in an image, such that their optimal detection requires the use of operators of different sizes. He also theorized that sudden intensity changes produce a peak or trough in the first derivative of the image. Consequently, a vision filter [according to Marr (1982)] requires two characteristics: It should be a differential operator, and it should be capable of being tuned to act at any desired scale (Graps, 1995).

To meet these criteria, an image-object is considered an entity composed of primitives (in this case, individual pixels) more similar to it, than dissimilar. Where the heuristics determining the threshold of “similarity” are based on the novel concept that all pixels within an image are H-resolution samples of the scene-objects they model, even though both H- and L-resolution information potentially exist for each object within the image. The importance of this rule is that spatially near pixels elicit a strong degree of spectral autocorrelation. Therefore, when we plot the digital variance of samples obtained within a varying sized window while centered on an image-object of known size, a distinct break, or threshold, in variance is observed over increasing window sizes, which (threshold window size) strongly relates to the objects known size. This is similar to analysis using linear variograms but is not limited by the lag/accuracy problem (Hay et al., 1996), or the difficulty of interpre-
Conceptually, this variance threshold represents the end of one scale—that is, the maximum zone of influence (i.e., area) at which the pixel under analysis is spectrally and spatially related to its neighbors—and the beginning of the next scale—which defines a new image-object, of which the mean value and pixel location of the current threshold window will be an H-resolution member (see mean dataset in the Discussion section). For example, Class 4 tree crowns (2-4 m) are poorly suited to be modeled by the 5 m data, as these data are L-resolution with respect to this class [i.e., the objects (tree crowns) composing the class are much smaller than the spatial sampling scheme (pixel size)]. But if the pixels over this forest class are considered as H-resolution and their DN variance is examined at increasing window sizes until a threshold in variance is reached, a unique (potentially optimal) spatial extent can be determined for each H-resolution pixel which indicates how it (a single pixel representing integrated tree and background signals) is associated with a portion of an object existing at the next coarser object scale, that is, gap or stand parameters. A measure of the window area obtained at this variance threshold is then used as a weighting value to model the influence of the image-object within the upscale kernel. This is based on the supposition that when upscaling an image, the H-resolution image-objects within a scene should have more influence on the integrated signal than
Variance Threshold / Window Size Measures

![Graph showing variance threshold measures vs. window size](image)

Figure 5. This graph depicts plots of variance threshold measures vs. window size for individual trees numbered 15-22 as illustrated in Figure 3, and also illustrates the variability within a natural scene. Values are derived from a varying sized window centered on their respective tree centers. Measures similar to these were used to develop object-specific threshold heuristics.

In the second process, weighting occurs in the following manner: A user determined, fixed-sized, square moving kernel is evaluated over the newly generated area dataset. Within this kernel, each unique area value per pixel is identified and divided by the sum of the total area values within the specified upscaling window (i.e., upscaling from 1 m² to 10 m² data requires a 10x10 pixel kernel). This results in an individually weighted area value per pixel. Each weighted value is then multiplied by the DN at the same coordinates in the original CASI imagery, producing an object-specific area-weighted spectral value. This process is applied to all pixels within the upscaling kernel, which are then summed and assigned to the appropriate location in a new up-scaled image file. The kernel is then moved over the area dataset a distance equivalent to its static window size, and the process is iterated until the entire dataset has been assessed and a new up-scaled image produced.

Initial analysis and heuristic development were conducted using 1.2 m CASI imagery from a previous study (Hay et al., 1996) which represented over 150 trees with known spatial locations (Fig. 4), and corresponding measurement data (Hay and Niemann, 1994). Due to the high variability exhibited in this natural scene (Fig. 5), a digital forest model was generated—composed of 373 objects (tree crowns) with known sizes—and the described
theory was applied to it with excellent results. To account for the inherent variability within the natural scene, we returned to the \( H \)-resolution imagery and further developed and refined the variance/kernel-size heuristics to agree with known forest measurement data. We then applied these refined methods to this study for further analysis.

RESULTS AND DISCUSSION

Area, Mean, and Variance Datasets
The area dataset (Fig. 6) models (both numerically and thematically) the relationship between each pixel (grain), and the spatial area of influence, or structural extent of the image-object it is presently an \( H \)-resolution member of. While it does not indicate that the processes which produce these unique spatial extents (and corresponding thematic patterns) are related, inference may be possible with more detailed ancillary data. In Figure 6, the brighter the tone, the larger the area of influence. In the 1.5 m area image (top left inset), several tones corresponding to different sized tree crowns—which include both illuminated and shadowed portions—are represented within homogeneous spatial units, whereas typical edge detecting algorithms would segment the two. Other thematically defined areas directly correspond to canopy gaps, forest classes, forest stands, road systems, and logged areas. In the 3 m, 5 m, and 10 m area datasets, each increase in scale reveals less overall tonal variation within the same spatial extent, and may be indicative of how specific landscape structuring processes with unique spatial extents dominate others as resolution changes. We are currently investigating techniques to map these varying spatial fields over increasing scales in an effort to link them with appropriate scale-specific biophysical models, in part, by using output generated from the mean dataset.

In the mean dataset (Fig. 7) each pixel represents an average spectral value of all the pixels within the unique maximum spatial extent (or area) of the image-object it resides in. Though only briefly mentioned in this paper, its generation provides an important juncture in this upscaling paradigm. By adopting the human eye model, discarding the concept of a user-defined static upscaling window, and iteratively generating upscale images derived from the mean of the pixels within the threshold window, we can produce an object-specific upscaled image which hierarchically models the next scale(s) of image-objects within a digital scene (at multiple resolutions) as the human eye would see it, rather than as a sensor would arbitrarily define it. Iterations on mean datasets will produce corresponding area and variance images, each with their unique spatial patterns, which, when compared over different scales, can be used to describe how the extent of different structures, or objects, dominate over others with changes in scale. It is important to note that each scale of mean imagery will be optimally upcaled for the next level of each of the many unique sized image-object composing it, as the algorithm allows the statistics of each image-object within the scene to determine its next most optimal spatial extent, rather than being arbitrarily upcaled by a convenient static kernel. In essence, we are using grain (pixel size) to define extent (object influence or footprint), which then defines the scale of the next (coarser resolution) image-object it is presently an \( H \)-resolution member of. Obviously, it will be critical to examine and define the exact spatial resolution each pixel represents as it moves through this hierarchy of object-specific scale changes.

While aware that the entities that emerge in a dataset are scaled by virtue of the observation protocol and the filters applied to the dataset during analysis (Allen and Star, 1982), we believe that if this human vision approach proves appropriate, then the changing patterns generated at each new scale of image-objects, in combination with ancillary data, that is, soils, slope, aspect, vegetation cover, climatic variables, etc., may provide a new tool for gaining a greater understanding of landscape processes from pattern. Further explanation falls beyond the scope of this article, though these ideas are an issue of current research (Niemann and Hay, 1996).

Each pixel in the variance dataset (Fig. 8) represents a unique variance value obtained when a series of threshold-old heuristics are met. As indicated, these heuristics are based on the shape of variance measures plotted over changing kernel size; consequently, each pixel is also related to a uniquely calculated spatial extent which defines the maximum scale of an image-object it is an \( H \)-resolution member of. In this dramatically textured image (Fig. 8), dark DNs represent spatial extents of low spectral variance (i.e., relative homogeneity), while bright DNs represent areas of high spectral variance (i.e., relative heterogeneity). This dataset provides an excellent visual measure of how well the object-specific algorithm operates to delineate objects. Essentially it performs as an edge detector, where image-objects (dark pixel groups) reside within edges (bright pixels). Visual analysis of this scene reveals a large amount of new information, particularly of physical boundaries, or ecotones that were not immediately obvious within the original imagery, but are confirmed by ground surveys and aerial photography. Based on the well-defined edge effects that appear in this image, we feel highly confident in the ability of this technique to define image-objects that model realworld scene components.

RMSE and Visual Interpretation
The root mean square error (Algorithm 1) represents the average difference in DNs per class between the corrected 10 m dataset \( (R_c) \) and an upscaled dataset \( (R_u) \) generated by one of the five resampling algorithms. Once the 1.5 m, 3 m, and 5 m data were upscande to a
10 m spatial resolution, 50 samples \((N_{obs})\) per six forest classes were evaluated for each algorithm (five total), for each scale of imagery (for a total of 6000 samples). Samples were selected away from class boundaries, and at a minimum distance of 20 m (2 pixels) apart from each other within the original and upscaled 10 m datasets:

\[
\text{RMSE} = \left( \frac{\sum (R_y - R_i)^2}{N_{obs}} \right)^{1/2},
\]

(1)

Based on the initial premise that the upscaled image with the lowest RMSE is the most accurate upscaling routine of those tested, RMSE results (Table 2) indicate that the variance-weighted technique (SUP) is the most accurate upscaling algorithm of those tested, as it produces an upscaled image with the lowest RMSE in 10 out of 18 classes over all forest types and ranges of scale tested. In the eight times it did not obtain the lowest RMSE, it produced six values with the second lowest. Simple nonoverlapping averaging provided the second most accurate class results, with the lowest error values.
This collage illustrates each of the four scales of mean data, where each pixel represents an average spectral value of all pixels within the unique maximum spatial extent (or zone of influence) of the image-object it resides in. For comparative purposes, each subimage represents the same location as the area data set (Fig. 5). Illustrated in the right image and at the top left are the 1.5 m data. Below it are the 3 m, 5 m, and 10 m datasets respectively.

7 out of 18 times. Although the RMSE values appear exceptionally high, they are a valid measure when considered that the data are 16-bit spectral values. There are also the possibilities of class samples at a particular scale being misregistered at another scale due to roll error, image smear (see Image Acquisition subsection), and geometric error.

Regarding image smear, it is interesting to note that visual analysis of the variance and area datasets indicate that the variance-weighted algorithm considers areas of image smear—usually a composite of several tree crowns, which appear through different scales and forest classes over the study area—as a single image-object, just as a human interpreter would. From an evaluation perspective, this results in different window sizes, variance and area measures, weighting functions, and RMSE values being generated over these smeared areas, from what would be the case if crowns were individually assessed. This characteristic was only noted as we attempted to understand why several RMSE results for SUP indicated
Figure 8. This image illustrates the spectral variance dataset which is produced from a contextual analysis over increasing spatial extents until a threshold heuristic is met. Bright areas represent locations of high variability (i.e., edges), while dark areas represent more homogeneous locations. Multiscale images are included to visually illustrate changing information content as data are generalized due to a coarsening of the sensor's spatial resolution. Illustrated in the right image and at the top left are the 1.5 m data. Below it are the 3 m, 5 m, and 10 m datasets.

substantially different values from those of AVG. By recognizing how the algorithm interpreted the “smeared” objects over sampled classes, we consider these uncharacteristic RMSE values (i.e., Table 2: intersection of Class 4, and 5–10 SUP) as further support for how well the object-specific approach works.

When the upscaled images (Fig. 9) are visually compared to the original data collected at 10 m, greater landscape structure are apparent in the resampled images though this may be difficult to assess on the printed medium. In several cases this structure is noise generated as a function of the algorithm used (i.e., NN). In other cases, these spatial structures represent actual scene components, and may be explained by the relationship between the sensor's ground resolution and the size of the scene objects. Slater (1980) indicates that if scene objects are less than one quarter the ground resolution of the remote-sensing system, they are imaged at the size
of the point-spread function of the system. Since this function describes how a point of light is spread by lens aberrations—which in modern lenses tends to be very small—their influence is essentially negligible. Similarly, Marr (1982) states that the features of an object that are much smaller than the primitives used to describe it (i.e., the resolution cell of the sensor) are not just inaccessible, they are completely omitted from the description. Therefore, determining the most accurate upsampling routine based on the lowest RMSE value (as we have done) may not be suitable for evaluating this type of data, as upscaled datasets incorporate fine object detail from their H-resolution parent images which does not exist within the original L-resolution image. Thus they can never be the same.

Understanding this perspective is important as it implies that traditional evaluation techniques that compare the original (i.e., 10 m data) with the model (i.e., 1.5–10 m upscaled data) may not be suitable for evaluating this type of data, as upscaled datasets incorporate fine object detail from their H-resolution parent images which does not exist (or is sufficiently negligible) within the original L-resolution image. Thus they can never be the same.

Visual analysis of the upscaled and original images (Fig. 9) and an understanding of the relationship between IFOV and scene objects suggest that, with a theoretically sound resampling routine such as the variance-weighted algorithm, an upscaled image is more representative of the scene it models than an image captured at the same upscaled resolution, as an upscaled image incorporates information that the sensor filters out, or is “blind to” in the original L-resolution image. Consequently, it may be more appropriate to focus future sensor development towards collecting data at the finest spatial resolution possible (i.e., the EarthWatch Early Bird system with 3 m panchromatic planned for late December 1996, and the submeter QuickBird platform scheduled for late 1997) and algorithm development focused on upsampling routines, rather than collecting a host of different spatial resolution data each with their own sensor specific challenges. Obviously there are numerous detection, retrieval, storage, and analysis considerations that require attention before this occurs, but this is where we believe the future lies.

**Different Optimal Spatial Resolutions for Different Objects**

Just as the lenses of our eyes change shape to “optimally” resolve different sized, shaped, and spatially arranged objects within a scene, no single spatial resolution of digital imagery provides an optimal (i.e., perfect, or exact) resolution for examining the varying sized, shaped, and spatially arranged image-objects it models. Rather, different optimal spatial resolutions exist for different object characteristics, suggesting the need for a multiscale approach for detection and analysis (Marceau et al., 1994; Hay et al., 1994; 1996). Here the term *optimal* is used ideally, with reference to many suitable scales from which to choose. It is not used in the sense of selecting the most satisfactory from a limited set of potentially inappropriate scales.

From a spectral classification perspective—where low within-class variance is desired, as it tends to result in a higher classification accuracy when using traditional classifiers—the spatial resolution that provides the minimum variance may be considered as optimal (Marceau et al., 1994b). By using this theory as a guide, an analysis of each of the six forest classes through the four corrected wavelengths (Fig. 10) confirms that no single spatial resolution of imagery can satisfy this criteria for all.

### Table 2. Root Mean Square Error Results

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5–10 CC</td>
<td>613.67</td>
<td>801.02</td>
<td>223.78</td>
<td>1114.31</td>
<td>415.03</td>
<td>530.24</td>
</tr>
<tr>
<td>1.5–10 BIL</td>
<td>529.90</td>
<td>727.21</td>
<td>209.52</td>
<td>1046.74</td>
<td>396.09</td>
<td>512.43</td>
</tr>
<tr>
<td>1.5–10 NN</td>
<td>580.83</td>
<td>756.51</td>
<td>263.40</td>
<td>1182.93</td>
<td>448.27</td>
<td>534.16</td>
</tr>
<tr>
<td>1.5–10 AVG</td>
<td>325.56</td>
<td>556.64</td>
<td>158.57</td>
<td>539.61</td>
<td>318.01</td>
<td>356.22</td>
</tr>
<tr>
<td>1.5–10 SUP</td>
<td>394.19</td>
<td>520.49</td>
<td>152.94</td>
<td>939.02</td>
<td>360.44</td>
<td>271.92</td>
</tr>
<tr>
<td>1–10 CC</td>
<td>424.21</td>
<td>637.24</td>
<td>290.22</td>
<td>962.73</td>
<td>487.05</td>
<td>541.50</td>
</tr>
<tr>
<td>1–10 BIL</td>
<td>364.40</td>
<td>601.39</td>
<td>253.74</td>
<td>936.40</td>
<td>442.90</td>
<td>526.33</td>
</tr>
<tr>
<td>1–10 NN</td>
<td>373.51</td>
<td>649.35</td>
<td>305.97</td>
<td>983.48</td>
<td>440.05</td>
<td>538.81</td>
</tr>
<tr>
<td>1–10 AVG</td>
<td>323.96</td>
<td>493.62</td>
<td>681.44</td>
<td>756.19</td>
<td>495.67</td>
<td>446.12</td>
</tr>
<tr>
<td>1–10 SUP</td>
<td>389.10</td>
<td>446.00</td>
<td>261.37</td>
<td>909.46</td>
<td>400.41</td>
<td>537.57</td>
</tr>
<tr>
<td>5–10 CC</td>
<td>474.11</td>
<td>598.17</td>
<td>234.81</td>
<td>965.05</td>
<td>418.33</td>
<td>370.55</td>
</tr>
<tr>
<td>5–10 BIL</td>
<td>435.51</td>
<td>445.49</td>
<td>197.05</td>
<td>892.17</td>
<td>384.43</td>
<td>350.48</td>
</tr>
<tr>
<td>5–10 NN</td>
<td>448.43</td>
<td>527.79</td>
<td>229.66</td>
<td>939.54</td>
<td>410.55</td>
<td>399.16</td>
</tr>
<tr>
<td>5–10 AVG</td>
<td>400.77</td>
<td>502.83</td>
<td>209.25</td>
<td>784.19</td>
<td>437.83</td>
<td>423.32</td>
</tr>
<tr>
<td>5–10 SUP</td>
<td>394.23</td>
<td>466.32</td>
<td>213.11</td>
<td>1012.09</td>
<td>356.00</td>
<td>312.58</td>
</tr>
</tbody>
</table>

Bold figures represent the lowest RMSE value, thus the most appropriate upsampling algorithm for each class over the 3 upscaled image sets.
Figure 9. This collage illustrates the visible differences between the original 10 m CASI imagery, and the datasets generated by resampling from 1.5 m to 10 m. Viewing from top left to bottom right, the illustrations are: 10 m original CASI imagery, 1.5-10 m cubic convolution (CC), 1.5-10 m bilinear interpolation (BIL), 1.5-10 m nearest neighbor (NN), 1.5-10 m nonoverlapping averaging (AVG), and 1.5-10 m variance-weighted upscaling (SUP).

classes. In class 1, the minimum variance is reached with a spatial resolution of 10 m. In classes 2, 3, and 5, minimum variance is obtained with the 1.5 m dataset, and, in classes 4 and 6, the 3 m data provides the lowest measure of variance.

From an information perspective—where high within-class variance is sought—there is also no single scale of imagery which provides a ubiquitous optimal spatial resolution for all forest classes. Classes 1 and 6 illustrate a measure of maximum variance in the 1.5 m dataset, class 2 at a spatial resolution of 10 m, classes 3 and 5 at 3 m, and class 4 at 5 m.

It is also interesting to note (Fig. 10) that the finest spatial resolution imagery does not always produce the largest measure of class variance, for example, low resolution 5 m data provides the greatest within-class variance in class 4, rather than high resolution 1.5 m data, and may be explained in the following manner. Minimum variance primarily results from two conditions: When pixels are L-resolution, thus the signal from many objects are integrated producing a smoothed response, or, when the signal is H-resolution, thus many pixels represent the same object with a greater likelihood of closer pixels having a similar spectral value. But when the pixel size is nearly the same size as the scene objects, that is, 3 m data in classes 3 and 5, the 5 m data in class 4, the 1.5 m data in class 6, and the 10 m data in class 2 (Figure 2), variance tends to be higher as the potential for a pixel to directly sample a single object (i.e., tree crown) is reduced. Instead, the pixel possesses the probability of representing the object, background, or combination of the two, resulting in increased spectral variance.
Variance of Six Forest Classes, Through Four Scales of Imagery.

Figure 10. This graph plots variance change in six forest classes through four scales of corrected imagery. A sample size of 50 pixels per class, per wavelength were selected (1200 total). Note that no single scale is optimal for all classes over all scales.

Class variance is a function of the size, shape, spatial arrangement, and spectral characteristics of the objects within the class, in relation to the spatial resolution of the sensor. While this theoretically sounds concise, the heterogeneity of varying sized objects within a single class can result in a host of difficult-to-predict outcomes. For example, in class 1 the 1.5 m data should result in a low variance measure due to a high correlation of high-resolution samples; instead, it produces the highest variance without being the same size as the objects. This may occur because the crown density is relatively low; thus the opportunity to sample a greater number of pixels with both similar tree, and background (shadow) values increases, resulting in an increased variance between the two.

Algorithm Caveats

Of the algorithms tested, it is apparent that NN, BIL, and CC are unsuited for upscaling remotely sensed data for several fundamental reasons. None of these techniques take into consideration the influence of neighboring pixels outside the upsampling kernel, and, in cases where upsampling requires a window size greater than 5x5, they generate an upscaled value from a sample which may not accurately represent the area they model. For example, within a 5x5 kernel, NN only samples a single pixel, BIL samples 5 pixels, and CC samples 17 out of 25 pixels. When these pixels are considered as actual landscape samples ranging in nominal area from 1 m² to 1 km² (i.e., CASI to AVHRR), their nonrepresentative influence on the upscaled image and/or their subsequent effect in ecosystem models could be dramatic. Also, if the upscale kernel size were increased beyond 5x5, these algorithms would still sample the same number of clustered pixels, resulting in an even less representative upscaled measure. We discuss these algorithms to illustrate that, even though they are included in many commercial image analysis packages under the generic guise of "resampling," they are inappropriate for resampling to coarser scales (i.e., upscaling).

In nonoverlapping averaging, all pixels within the upsampling kernel are evaluated, resulting in a more representative dataset than the previous three algorithms. The greatest concern with this algorithm is that averaging assumes that the larger scale system behaves like the average smaller scale system, which, as illustrated, by the differences in area thresholds over varying scales (see Fig. 6), is not a valid assumption. Over small spatial extents this technique is certainly convenient, and in many cases may provide a suitable model (i.e., as used to generate the mean dataset), but over large extents this assumption is not valid. Further discussion regarding the caveats involved in averaging and scale may be found in O'Neill (1988) and Beven (1995). In the variance-weighted technique all pixels within a user-defined upscale kernel are evaluated as well as how these pixels—representing parts of objects outside the arbitrary upsampling perimeter—influence those inside.

CONCLUSION

In an attempt to define and validate the most appropriate upscaling technique to represent a remotely sensed signal at various scale changes, we have examined how a 660 nm signal of six forest stands vary through four different (same-sensor) scales of imagery, and five resampling algorithms. From this study the following conclusions can be drawn:

- Just as a ubiquitous literary definition of scale does not exist, neither does a single optimal scale that is capable of accurately describing the multisized, shaped, and spatially distributed objects within a remotely sensed scene. We suggest that by sampling image-objects under iteratively growing object-specific kernels, unique optimal scales and information can be defined for them.
- The spatial extent (or object influence) of multiscale image-objects can be defined by the similarity measures of the elements (pixels) that constitute them. These spatial measures can then be effectively used to representatively upscale data. This novel approach also provides a mechanism whereby grain and extent may be iteratively recalculated to determine the next scale(s) of image-objects within a dataset (i.e., analysis using the mean dataset).
- As the grain (spatial resolution) of an image becomes coarser, while the spatial extent viewed remains constant, the scene tends to move towards...
a level of reduced variance as scene components become finer portions of the next level of larger, more dominant landscape objects (see Fig. 6).

Providing that an appropriate resampling algorithm is applied, an upsampled image represents a more accurate model of a scene than an image obtained at the upsampled resolution. This occurs because upsampled images incorporate higher-resolution detail that does not exist in the coarser scale image due to the relationship between sensorIFOV and scene component size. This suggests that while the RMSE is a suitable technique for evaluating a data model against the original data, it may not be suitable for assessing the most appropriate upscaling routine as we have done, as the model (i.e., upsampled image) can never be the same as the original data against which it is evaluated. These ideas also suggest that new evaluation criteria need to be established to verify the appropriateness of upsampled data and that future platform and algorithm development should focus on higher-resolution spatial data and upscaling techniques.

Nearst neighbor, bilinear interpolation, and cubic convolution resampling algorithms are not suitable for resampling remotely sensed data to a coarser spatial resolution (i.e., upscaling), especially when the upscale factor is greater than 5. This is important to recognize, as these algorithms continue to be built into commercial remote sensing image analysis packages and inappropriately find their way into upscaling analysis.

Nonoverlapping averaging appears a simple method for approximating the spectral response in an upsampled image, though caution should be used as averaging assumes that the larger-scale system behaves like the average smaller scale system. When in reality, different structuring processes occur in the landscape at different (often unique) scales in a nonlinear fashion, as illustrated by observing the changing spatial extents defined in the area datasets (Fig. 6).

A new variance-weighted upscaling technique has been described that produces a generalized spectral response more similar to the original coarse resolution signal than the response produced by nearest neighbor, bilinear interpolation, cubic convolution, and nonoverlapping averaging. This technique also introduces an object-specific approach to upscaling, with the ability to generate additional datasets (from a single-channel input) which may be beneficial as logic channels in traditional classification strategies, and as guides and input for developing a hierarchy of linked scale-specific models.

Large volumes of high resolution multiscale data offer challenges to the user that are more time-consuming, complex, and computationally intensive than single scene analysis. Each scale of data essentially provides the analysts with a new single-scene project to understand and integrate with the other scales. Multiscale analysis is not trivial.

This article is part of a research agenda directed towards utilizing multiscale remotely sensed imagery to quantitatively identify the spatial extent of critical landscape thresholds as it is hypothesized that they define specific regions of scale dependence at which unique scale-limited models are needed (Holling, 1992). As illustrated in this article, these threshold qualities may also be used as upscaling inputs, potentially capable of providing the key for translating results obtained from a model at one scale, to the parameter set of a model operating at the next scale (Wickland, 1989). Further research will be directed towards refining the heuristics used; following the mean dataset through a hierarchy of scale changes: examining multispectral scale changes rather than a single band scale change: and determining if and how datasets produced from these object-specific thresholds relate to landscape processes.

The authors gratefully acknowledge the support of the Advanced Forest Technologies Group at Natural Resources Canada, and grants and scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC), Forestry Canada, and a Presidents Research Scholarship from the University of Victoria. Special thanks is also extended to Dr. Jim Czencer at the Institute of Ocean Sciences, and to three anonymous referees for their helpful comments.

REFERENCES


