

What's New in IDRISI Selva!

Selva. In several languages, the word means dense tropical forest, a biome of immense importance that is strongly coupled to the Earth's atmosphere and hydrosphere. With this 17th release of the IDRISI geospatial software for monitoring and modeling the Earth system, Clark Labs is introducing a series of special tools for the analysis, monitoring, and assessment of coupled systems.

Selva Overview

The Selva Edition includes a major revision to the Land Change Modeler (LCM) to support REDD (Reducing Emissions from Deforestation and Forest Degradation) projects, allowing the atmospheric impact of deforestation to be assessed. Additionally, a major enhancement to the Earth Trends Modeler (ETM) includes an array of pattern decomposition tools for

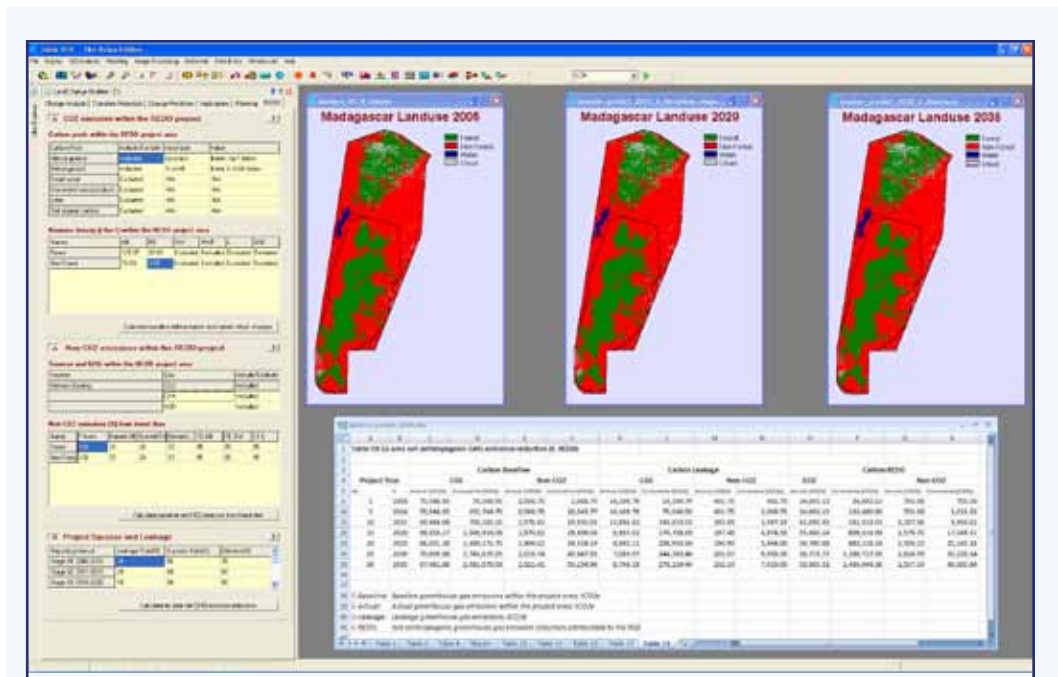


Figure 1: The new REDD tab in the Land Change Modeler (LCM). The REDD tab provides a full accounting of CO₂ and non-CO₂ baseline emissions for a REDD project area and the reductions that would be expected as a result of REDD project activities. A set of 19 tables is generated following the logic of the World Bank's BioCarbon Fund methodology. However, the procedures are inherently compatible with most methodologies in use within the voluntary carbon market sector.

the analysis of coupled systems (such as the oceans and atmosphere) over time. Selva also includes major enhancements to the display subsystem of the IDRISI suite, including image pyramids and support for mega images.

Land Change Modeler Enhancements

The Land Change Modeler has become a major tool for many of our users, and with this release, we have expanded and enhanced its capabilities. These include:

A new REDD tab to support projects aimed at Reducing Emissions from Deforestation and Forest Degradation (Figure 1). The new REDD tab is intended to support the various methodologies being reviewed and approved by the Verified Carbon Standard (VCS) for the voluntary market. The REDD tab facilitates the estimation of baseline emissions from various carbon pools and allows the calculation of deferred emissions and carbon credits. It also creates formatted tables in an

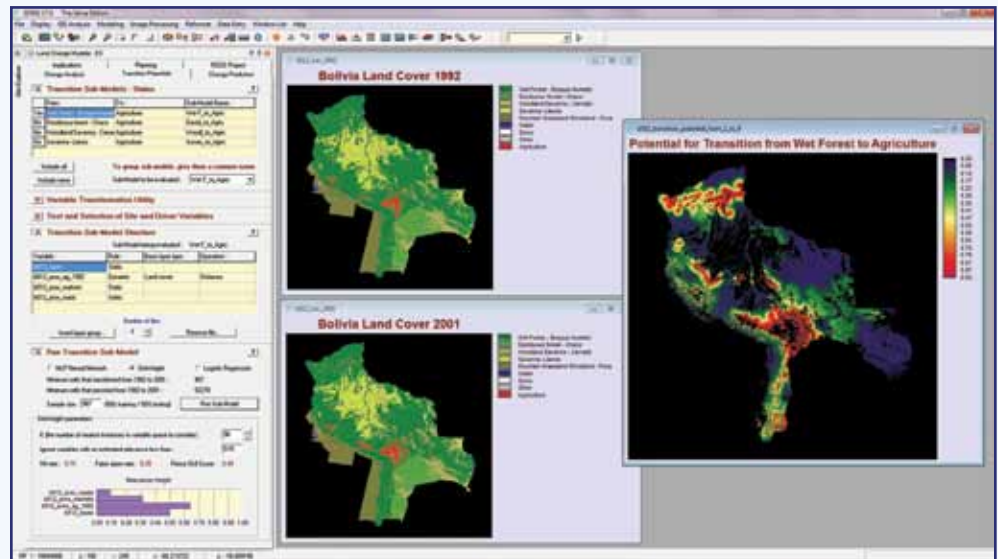


Figure 2: The new SimWeight empirical transition potential modeling procedure in LCM. SimWeight is based on a modified K-nearest neighbor machine learning algorithm. It offers the strength of modeling that is normally associated with the Multi-Layer Perceptron neural network without its complexity. There are only two very easy parameters – the sample size to use and the k parameter – the number of similar cases in the sample to consider in estimating the transition potential. Further, with SimWeight, the black box of machine learning has been opened, with feedback on the relative importance of each variable and a measure of the skill (based on an analysis of automatically designated validation data).

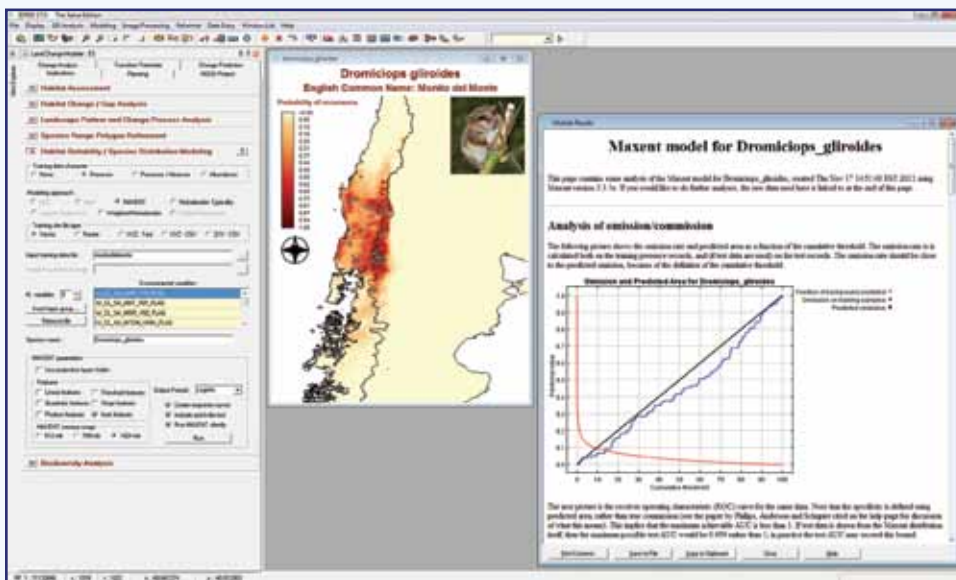


Figure 3: An example of the Maxent interface in LCM and its output. In this example, the range of *Dromiciops gliroides* (Monito del Monte) is modeled based on a collection of observation points (shown as points) and a set of nine environmental variables, such as annual mean temperature and precipitation. This new Maxent option extends the existing group of species distribution modeling tools such as the Multi-Layer Perceptron neural network, Mahalanobis Typicalities and Logistic Regression.

Excel spreadsheet format.

A new pioneering land cover change modeling procedure following the recently published SimWeight methodology.¹ SimWeight (Figure 2) is a machine learning procedure that has proven to yield results that rival that of the Multi-Layer Perceptron with minimal (and easily understood) parameters. The SimWeight procedure also gives feedback on the relative power of explanatory variables, as well as a skill measure based on a *leave x% out* validation procedure.

An integrated link to the popular Maxent procedure for species distribution modeling (Figure 3). Maxent uses a maximum entropy empirical modeling procedure and provides a wealth of feedback on the importance of the explanatory

¹ Sangermano, F., Eastman, J.R., Zhu, H. (2010) "Similarly weighted instance based learning for the generation of transition potentials in land change modeling." *Transactions in GIS*, 14, 5, 569-580.

variables and the user's parameter selection.

A new procedure called Harmonize has been added for those instances when the input land cover layers do not correspond as needed. The Harmonize tab (which only launches when a problem is detected) coordinates the land cover layers in terms of their spatial characteristics (projection, resolution and extent), background masks and categorizations.

Earth Trends Modeler Enhancements

With the Selva release, we have incorporated a major expansion of our spectral decomposition capabilities, continuing IDRISI's groundbreaking leadership in Earth System Information Science.

Both Principal Components Analysis (PCA) and Empirical Orthogonal Teleconnection (EOT) analysis now offer extended modes where multiple data series can be analyzed simultaneously. With Extended PCA (EPCA, also known as Extended EOF, or EEOF) and Extended Empirical Orthogonal Teleconnection (EEOT) analysis, one could, for example, analyze a time series of sea surface temperature imagery simultaneously with temperature in multiple levels of the atmosphere to search for coupled ocean-atmosphere patterns (Figure 4).

All EOT-based outputs now also indicate the percent variance explained.

Furthering the logic of extended analysis, Selva also offers Multichannel Singular Spectrum Analysis (MSSA – Figure 5) and Multichannel Empirical Teleconnection (MEOT) analysis. These modes are similar to EPCA and EEOT except that the multiple series involved are the same series at varying lags. This is a form of analysis that operates in phase space and is particularly effective in analyzing patterns that evolve in space

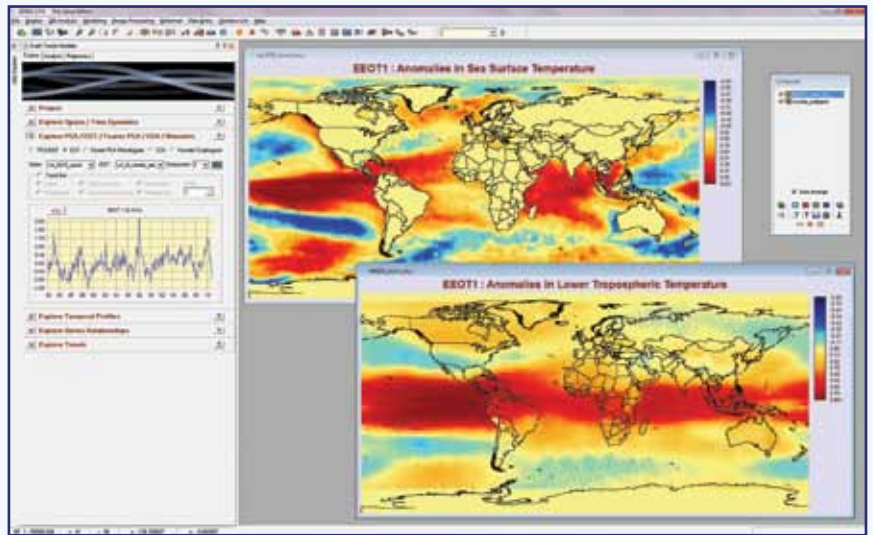


Figure 4: An Extended Empirical Orthogonal Teleconnection (EEOT) analysis of anomalies in sea surface temperature (SST) and anomalies in lower tropospheric temperature (TLT) over the 1982-2010 period. The first EEOT is clearly the El Niño phenomenon. The warming in the equatorial Pacific is very evident in the SST anomalies. However, what is interesting is that the atmospheric response is far more extensive, with positive anomalies in TLT being spread throughout the tropics – a coupled ocean-atmosphere phenomenon known as the atmospheric bridge.

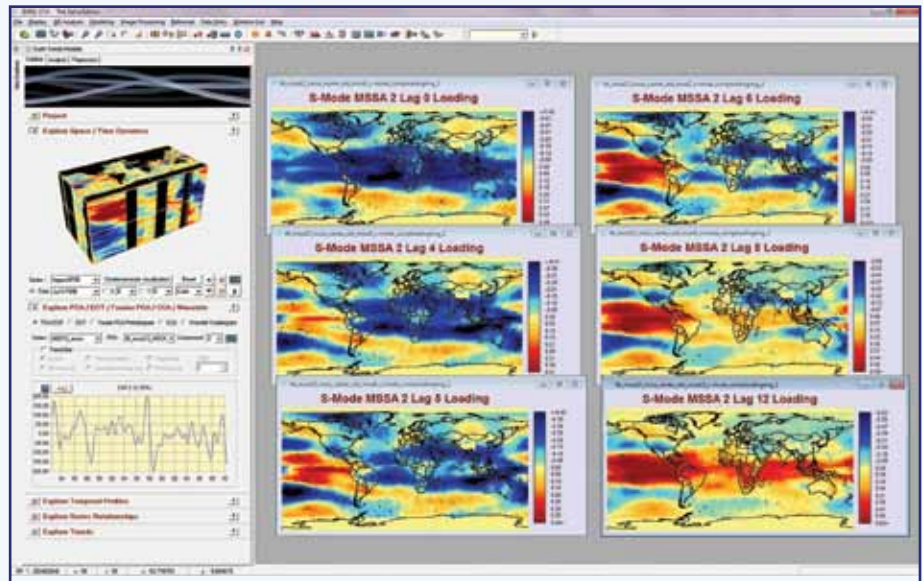


Figure 5: A Multichannel Singular Spectrum Analysis (MSSA) of monthly anomalies in lower tropospheric temperature (1982-2010) using an embedding dimension of 13 months. The embedding dimension establishes a time frame of focus in the analysis of evolving spatio-temporal events, and also indicates that the analysis will be run as an Extended PCA on 13 lagged versions of the same series. In this example, the temporal graph again shows an obvious relationship to El Niño, but instead of a single loading image, this analysis has one for each of the 13 lags (only six are shown). These images are like the frames of a movie loop that show the evolution of the lower tropospheric response to the El Niño/La Niña phenomenon – the genesis of the atmospheric bridge phenomenon illustrated in Figure 4.

and time such as many climate teleconnection patterns.

The PCA, EPCA and MSSA procedures in ETM also offer both T-mode and S-mode orientations for analysis – the first GIS/Image Processing software system to offer both. T-mode analysis is the mode most familiar to geographers, where the time series is organized as a sequence of images over time. S-mode analysis is more familiar to atmospheric and ocean scientists, where the data are organized as a series of one dimensional profiles over time, one at each pixel location. This distinction may seem to be a simple change in perspective, but it has major analytical implications,² as illustrated in Figure 6. Each orientation offers different insights into the structural patterns that underlie the series.

IDRISI was a pioneer in offering both standardized and unstandardized PCA and

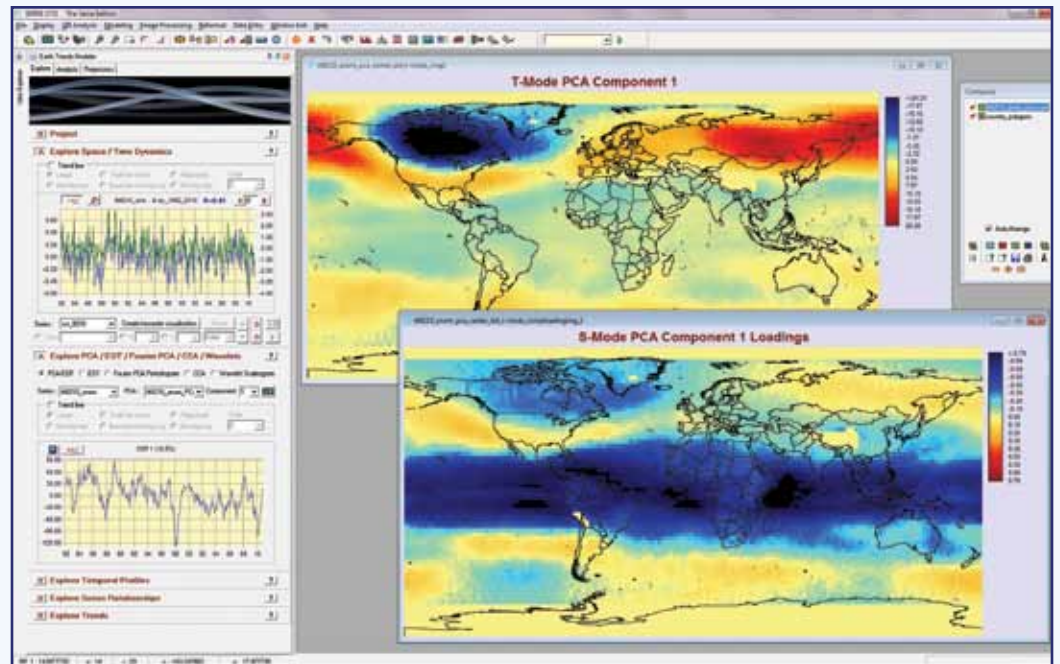


Figure 6: An example of the difference between T-mode and S-mode analysis. In both cases, a PCA has been conducted on an archive of monthly anomalies in lower tropospheric temperature (TLT) from 1982 to 2010. The top image (and the blue line in the top graph) represent the first component of a T-mode analysis where the variables analyzed are time slices (maps). The spatial pattern is that of the Arctic Oscillation (the green line in the top graph is the tabulated Arctic Oscillation Index and the calculated correlation is 0.41, as indicated by the new lag correlation monitor). The bottom graph and image represent the first component from an S-mode analysis, where the variables are pixels over time. Here we see the pattern of the atmospheric bridge that was found in Figures 4 and 5 (but inverted in this view). GIS and remote sensing scientists have traditionally used T-mode analysis while atmospheric and ocean scientists have more typically used S-mode. However, as is illustrated in this example, the differences can be profound and analysis with both modes is essential for Earth System Science applications. IDRISI is unique in offering these two modes to the GIS and remote sensing communities.

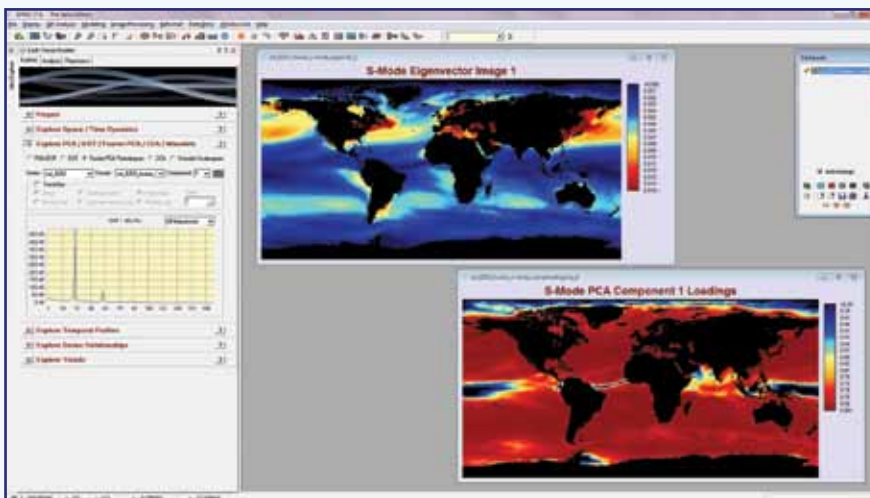


Figure 7: An analysis of monthly sea surface temperature from 1982-2010 using the revised Fourier-PCA procedure. The components from this procedure are periodograms with the x-axis representing frequency as wave numbers. This first component shows a strong peak at 29 waves over the 29-year series (with a couple of minor harmonics). It therefore represents the annual cycle. The loading image (lower right) shows that almost all ocean regions (with the exception of the poles and the thermal equator) exhibit prevalent seasonality. However, the eigenvector image shows that the seasonal effect is most extreme in the northern mid-latitudes. Leading modes of the Fourier-PCA analysis, such as this one, show major patterns of variability that can be described as combinations of sinusoidal waves.

EOT. With this release, we add the option to uncenter the analysis. Centering refers to the removal of a mean from a data set. Centering over space (removal of the image mean from each image) detrends the series over time and centering over time (removal of the pixel-specific mean over time from each pixel) detrends over space. The effect can profoundly alter the analytical outcome. With the Selva edition, IDRISI users are now in control of this feature.

With Selva, we also introduce Canonical Correlation Analysis (CCA). CCA is somewhat similar to PCA, but it seeks components for two image series such that the component images (known as canonical variates) are maximally correlated. However, we have extended classical CCA to include pre- and post-processing steps especially designed for image time series analysis. As such, our implementation provides both S- and T-mode orientations, and provides another valuable tool for the analysis of coupled earth systems (Figure 8).

With the prior Taiga release, we introduced Fourier-PCA – a form of image time series-based Fourier analysis that used Principal Components to organize and present a form of spatial/spectral periodogram. With the introduction of S-mode analysis in the Selva Edition, we have made the decision to switch the PCA step in Fourier-PCA from T-mode to S-mode. The results are far superior (Figure 7).

With the Trend Analysis and Seasonal Trend Analysis procedures, we have also introduced the recently-published Contextual Mann-Kendall (CMK) trend significance measure.³ CMK uses spatial autocorrelation to extend the power of trend significance testing.

Several elements of ETM include the ability to

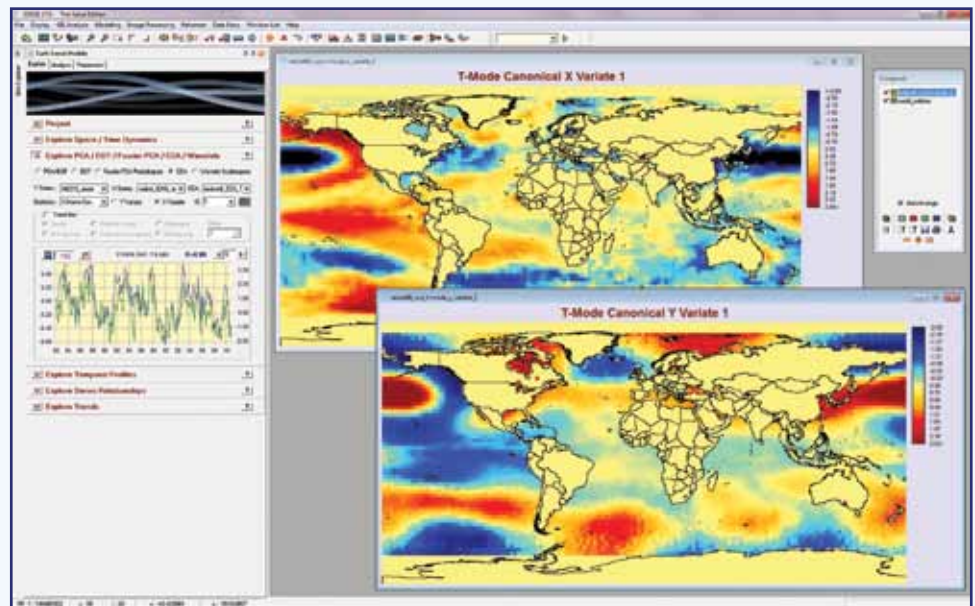


Figure 8: The result of a T-mode CCA analysis of monthly anomalies in sea surface temperature versus monthly anomalies in lower tropospheric temperature from 1982-2010. T-mode analysis is a new mode being introduced in the Selva Edition. Traditional S-mode CCA looks for recurrent temporal patterns across the two data series. In T-mode analysis, it is looking for recurrent spatial patterns. In this global analysis, it turns out that the predominant global spatial pattern is that of the Pacific Decadal Oscillation (PDO). In the graph, the index to PDO published by NOAA is plotted (in green) on top of the X-variate (sea-surface temperature) homogenous correlation (in blue). The correlation is 0.80. An equivalent S-mode CCA finds that the familiar ENSO (El Niño Southern Oscillation) pattern is the most prevalent pattern over time.

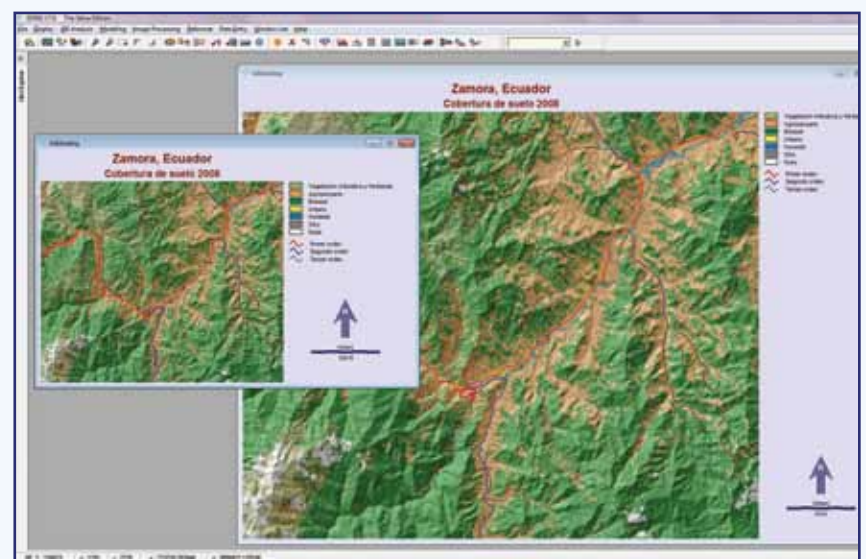


Figure 9: With the new auto-arrange feature, IDRISI automatically arranges ancillary map components such as titles, legends, scale bars and north arrows. Thus, when map windows are resized (such as in this illustration), ancillary components are automatically rearranged. For times when non-conventional layouts are desired, autoarrange can be turned off for any individual map window using Composer.

³ Neeti, N., and Eastman, J.R. (2011) "A Contextual Mann-Kendall approach for the assessment of trend significance in image time series," Transactions in GIS, 15, 5, 599-611.

view one-dimensional time series (such as PCA loadings, EOTs, etc.) and to optionally superimpose a second series. In these instances, we have added an interactive facility to slide the second series forwards or backwards in time and display the lagged correlation between the two series. This is an invaluable tool when searching for explanatory variables (Figure 6).

For missing data interpolation, we have extended the linear interpolation option to allow bridging over gaps of any specified duration (as opposed to just *one-time-slice gap*).

The Inverse PCA denoising option now also offers a choice between S-mode and T-mode. For small regional study areas, T-mode denoising tends to work best. However, for large regions and global studies, S-mode is often the better choice.

The Generate/Edit series options now also include a spatial subsetting tool.

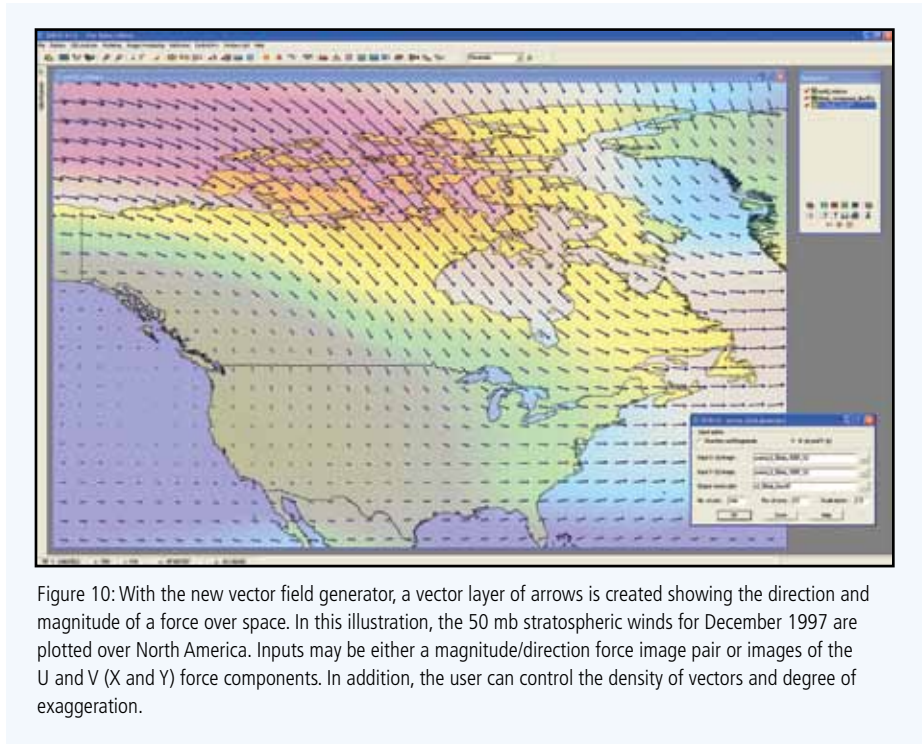


Figure 10: With the new vector field generator, a vector layer of arrows is created showing the direction and magnitude of a force over space. In this illustration, the 50 mb stratospheric winds for December 1997 are plotted over North America. Inputs may be either a magnitude/direction force image pair or images of the U and V (X and Y) force components. In addition, the user can control the density of vectors and degree of exaggeration.

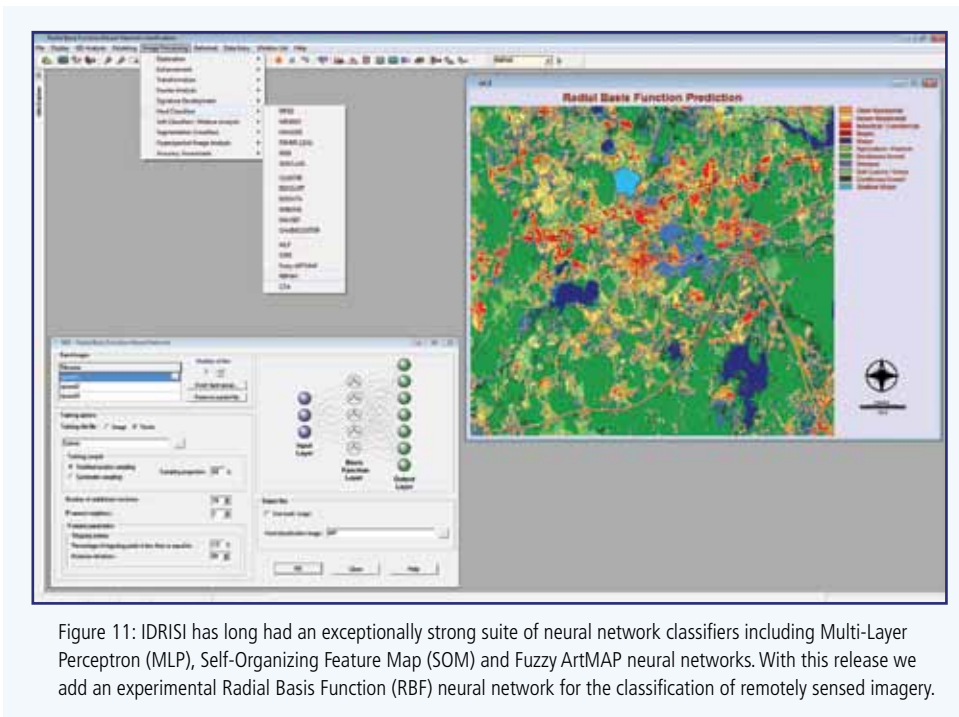


Figure 11: IDRISI has long had an exceptionally strong suite of neural network classifiers including Multi-Layer Perceptron (MLP), Self-Organizing Feature Map (SOM) and Fuzzy ArtMAP neural networks. With this release we add an experimental Radial Basis Function (RBF) neural network for the classification of remotely sensed imagery.

Display Enhancements

Probably the most immediately noticeable features of Selva are the new display elements.

With the Selva edition, IDRISI breaks through the Windows 32-bit display architecture, with the ability to now both process and display images up to 2 billion rows by 2 billion columns. This would be the equivalent of displaying an image the size of a Landsat scene (185 km on a side) with a pixel resolution less than that of a human hair! To support the rapid display of such images, IDRISI has also introduced support for an image pyramid – a multiple resolution image that allows the rapid display of large images regardless of the level of zoom. Unlike other systems, where the image pyramid is in a separate file, IDRISI incorporates the pyramid into the same

image file. If you copy an image to another location, you can be assured that the pyramid will move with it.

IDRISI now includes an auto-arrange feature whereby map elements such as titles, legends, scale bar, insets, etc. are automatically arranged. By default, the auto-arrange feature is on for any map window, but it can be turned off (in Composer), allowing the manual positioning of elements for a map composition (Figure 9).

You will also note from Figure 10 that the Composer window has changed. It not only has a new interface design, but it is sizeable in order to better handle long filenames and compositions with many layers.

Map windows can now be very simply resized by pulling out or pushing in the lower-right corner. This is in addition to the traditional options of using the home and end keys.

Another new display feature is the ability to display vector fields. Figure 10 illustrates a vector field of 50 mb stratospheric winds for December 1997. As you will note from the vector field generator dialog box, the inputs can be of two types – a magnitude and direction force pair (such as slope and aspect) or as the X and Y (U and V) components of the force.

New Analytical Modules

Alongside the additions made in ETM and LCM, changes have also been made to the extensive set of independent analytical modules IDRISI provides.

A Radial Basis Function (RBF) neural network classifier (Figure 11) has been added to complement the Multilayer Perceptron (MLP), Self-Organizing Map (SOM) and Fuzzy ArtMAP neural networks previously available in IDRISI. This experimental

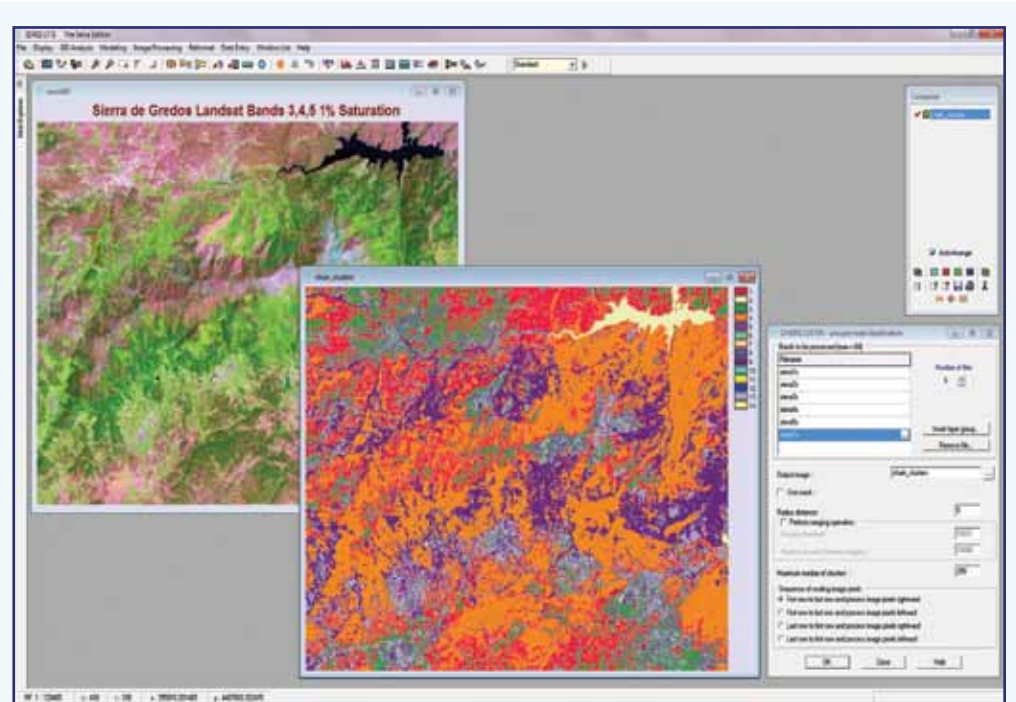


Figure 12: The new Chain Cluster procedure adds to IDRISI's already extensive set of unsupervised classification procedures. The technique, which has strong similarities to the logic of the Fuzzy ArtMAP neural network procedure, is particularly simple to parameterize when used with standardized image inputs. In this illustration, 14 clusters have been naturally uncovered based on a logic that clusters should consist of pixels within 5 standard deviations of the cluster centroid in band space.

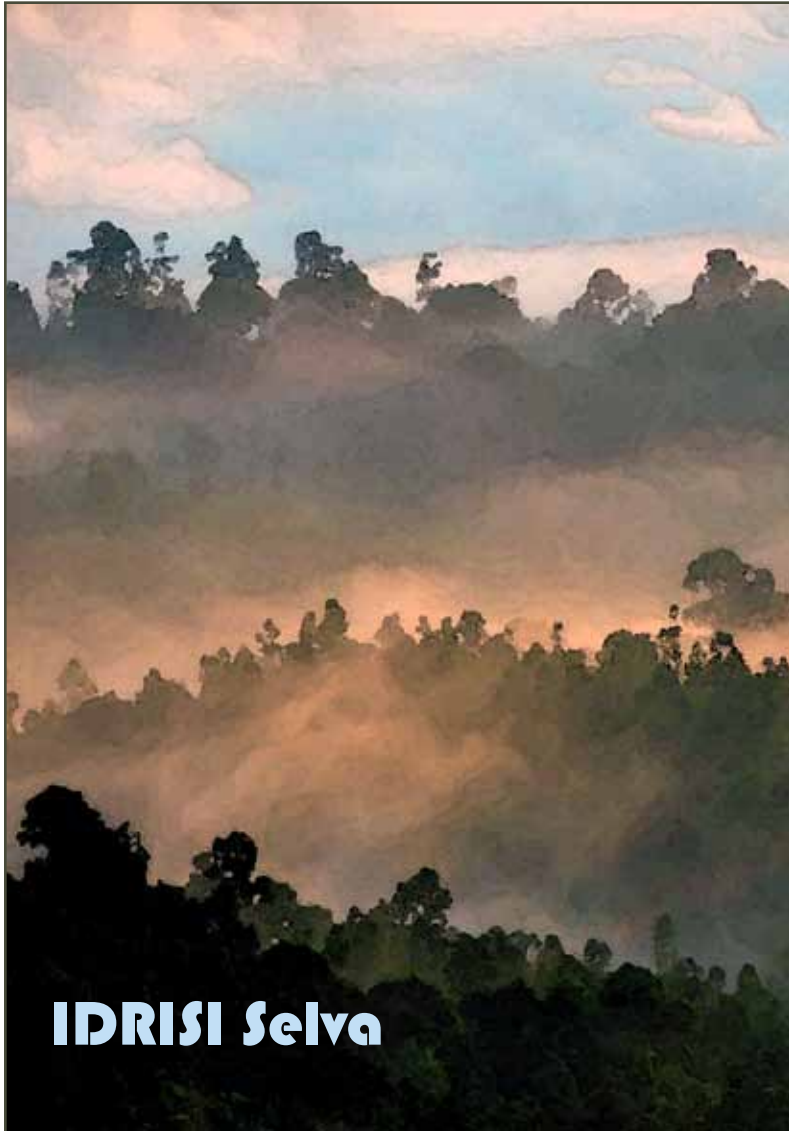
module continues the cutting-edge work of Clark Labs on machine learning procedures.

A Chain Clustering procedure has been added to the basic clustering tools available in IDRISI. This very simple procedure has some useful statistical properties that make it especially easy to use (Figure 12).

A Durbin-Watson module has been added independently of ETM to map locations where serial correlation is present.

A new high-precision rank-and-slice technique has been added to help support decision making procedures. It pulls out the top ranked pixels according to a specified threshold. The result is exact, and is now used internally by LCM. However, it is also available as an independent module called TOPRANK.

A new general purpose PCA module now offers distinctions between S-mode and T-mode, even for multivariate cases not involving time series. The options for standardized/unstandardized, centered/uncentered and forward/inverse transformation are also included.



Contact us at:
Clark Labs
Clark University
950 Main Street
Worcester, MA 01610 USA
Tel: +1.508.793.7526
Fax: +1.508.793.8842
Email: clarklabs@clarku.edu

Order online:
www.clarklabs.org

A general purpose Canonical Correlation Analysis procedure, called CANCOR has been added for the analysis of pairs of data sets. To avoid confusion, the module previously called CCA (Canonical Components Analysis) has been renamed CANCOMP.

New or Revised Import/Export Modules

Support for KML files (Keyhole Markup language, used by Google) has now been extended to include point, line and polygon files as well as raster images. Both import and export are supported.

MODIS users will love this module: MODISCONV. MODISCONV imports and converts MODIS tiled imagery to IDRISI raster format. The files are imported and then the tiles are mosaicked, with options to mosaic tiles of different geographic extent. The mosaicked output can then be projected to a different reference system. All these tasks can be automated and make effective use of multi-core CPUs to speed the processing.

The MODISQC module has been enhanced. Quality control maps can now be generated for vegetation indices using MODIS collection 5, land surface temperature collections 4 and 5, and Albedo collection 5.

Improved Modules

Every so often, we take the opportunity to substantially revise modules that we think could benefit from a different approach (a process known in the software industry as *refactoring*). The biggest refactoring we have undertaken is to revise every analytical module so that it can process images larger than 32,000 columns and rows (in fact, up to 2 billion columns and rows). With the Selva release, we have also focused on the core of the distance-based routines: DISTANCE, COST, VARCOST, DISPERSE and BUFFER. The optimization is substantial. For example, the refactored DISTANCE module now runs 26 times faster than before. Similarly, we have improved the precision of the change allocation procedure used in LCM through calls to the new TOPRANK procedure.

