3rd October 2017

Murray-Darling Vegetation Monitor User's Guide

Introduction

The Murray-Darling Vegetation Monitor (MDVM) is designed to allow accurate and efficient assessment of native tree stand condition along river reaches across the entire Murray Darling Basin and maps of estimated condition. The software utilises models of tree stand condition derived using machine learning techniques based upon extensive field surveys and corresponding Landsat satellite imagery. The model that has been developed and incorporated into MDVM can be applied to future Landsat satellite images to estimate the native tree stand condition at the time that the images were taken. These estimates can then be compared with field validation surveys from the same time period to assess the degree of model fit. Additionally these validation surveys may be used to adjust the model to better to fit the current observed conditions and thus produce more representative maps.

Modelling Technique

The vegetation condition model was derived using field observations of stand condition indices from 1754 surveys across seven years of observations (2009, 2010, 2012, 2014, 2015, 2016 and 2017). The original 175 sites surveyed in 2009 have repeated observations in four years. Most other sites do not have repeated observations. All surveys recorded the tree crown extent (CrownExt), percent area leaf index (PAI) and live basal area of the native trees (LiveBA). These measurements were then combined into a single index of Condition, which is an average score of the three components, scaled to 10.

The surveys were classed into six temporal epochs and matched with satellite images. The most recent surveys from 2016 and 2017 were combined as if these were observed within one epoch. Satellite images were used to model the observed indices. Median Landsat images computed over the year of the observations were produced by Geosciences Australia, along with upper and lower quartile images. The Landsat images have been chosen to provide the primary temporal component of the model, and are used to detect changes occurring from year to year. It is by the provision of temporal Landsat images that the MDVM can estimate the stand conditions and produce current maps of condition.

The models were trained using each year's observations, three constant images of tree presence, type and a base Autumn Landsat image constant, and by varying the temporal Landsat images (median, upper and lower quartile) to match the year in question. The model itself is a bagged ensemble of twenty multi-objective regression trees. Each tree estimates the four objectives (CrownExt, PAI, LiveBA and Condition) and these estimates are averaged across the ensemble for each pixel of the output map. MDVM used multi-threaded computing to efficiently sample each image, pass all the values through the bagged-regression trees. These determine the four condition estimates. These values are then re-assembled into a coherent multi-layer map.

Post-processing of condition maps may be performed if validation sites are provided. The post-processing step may re-align the model output with observed conditions that will further improve model performance. This is achieved by providing a simple inverse-linear regression model overlay onto the current model output. The validation data must supply site positions and observations of the four condition values. These are supplied to the MDVM as a CSV file which compares these values to the current modelled values. A least-squares inverse-linear regression ($x = m*(y + c)$) is then calculated for each condition index and applied to the current model map resulting in a corrected model map.

The MDVM tool is shipped with three models. The 'Veg Monitor Map All Years' requires 3 inputs, the median Landsat image (PCT_50), the lower quartile (PCT_25) and the upper quartile (PCT_75). Both of the other model types, 'Quick Veg Map All Years' and 'Veg 2016 and 2017 Only', require only a single Landsat image, whether a median image (PCT_50), or some other 6-band Landsat image. The 'Veg Monitor Map All Years' is the more complex model and produces slightly better results regardless of the year of the image. The 'Quick Veg Map All Years' may be used when only a single image for any year is available. The 'Veg 2016 and 2017 Only' is a single image model for 2016 and 2017. This model performs very closely to that of the 'Veg Monitor Map All Years' model for the 2016/17 years and perhaps should not be used to create maps from other years.

All three shipped models are supplied with an inverse-linear regression correction applied for optimal performance. Thus, they are in effect sequential models consisting of an ensemble of multi-objective regression trees subsequently corrected by an inverse-linear regression to provide the final map output. This does not preclude the tool from applying another correction should sufficient validation surveys be conducted to warrant further improvements to the outputs.

The models are not initially restricted in geographical scope and will be applied across the entire satellite image/s supplied provided these overlap with the other model input grids. Since the models were not trained by data from outside the regions of the native vegetation monitoring sites, the models cannot identify non-native vegetation, such as orchards, urban areas and irrigated fields, as sites of zero value. For this reason, estimates of vegetation condition for non-native vegetation types will be irrelevant and wildly inaccurate. To minimise this distraction, the user may opt to apply both native vegetation and woody masks to the modeling process. These mask grids will set the vegetation condition estimates to zero at that pixel if the native vegetation probability estimate and woody probability estimate requirements set by the user are not met. The mask grids themselves are models and do have some inadequacies but these are the best estimates of these two parameters at the current time.

Disk Contents

The disk contains five directories. The content of each is discussed below.

MDBALandsat

This directory contains all of the data provided by Geosciences Australia to DELWP and Ecoinformatics Pty. Ltd. to support this project. This directory contains a metadata directory and several Landsat image directories.

- **Metadata** details the processes required by Geosciences Australia to create the Landsat tiles required for the project.
- **20080901** contains the median images (**PCT_50**) for the 2009 model, as well as the upper (**PCT_75**) and lower quartile (PCT_25) images. Each of these subdirectories contains a **North** and **South** subdirectories which contain individual 1-degree tiles for the period for the north half of the Murray-Darling Basin (-26 to -31 degrees latitude) and the south half of the Basin (-32 to -38 degrees latitude).
- **20090901** contains the median images (**PCT_50**) for the 2010 model, as well as the upper (**PCT_75**) and lower quartile (PCT_25) images. Each of these subdirectories contains a **North** and **South** subdirectories which contain individual 1-degree tiles for the period for the north half of the Murray-Darling Basin (-26 to -31 degrees latitude) and the south half of the Basin (-32 to -38 degrees latitude).
- **20130401** contains the median images (**PCT_50**) for the 2012 model, as well as the upper (**PCT_75**) and lower quartile (PCT_25) images. Each of these subdirectories contains a **North** and **South** subdirectories which contain individual 1-degree tiles for the period for the north half of the Murray-Darling Basin (-26 to -31 degrees latitude) and the south half of the Basin (-32 to -38 degrees latitude).
- **20130901** contains the median images (**PCT_50**) for the 2014 model, as well as the upper (**PCT_75**) and lower quartile (PCT_25) images. Each of these subdirectories contains a **North** and **South** subdirectories which contain individual 1-degree tiles for the period for the north half of the Murray-Darling Basin (-26 to -31 degrees latitude) and the south half of the Basin (-32 to -38 degrees latitude).
- **20140401** contains the median images (**PCT_50**) for the 2015 model, as well as the upper (**PCT_75**) and lower quartile (PCT_25) images. Each of these subdirectories contains a **North** and **South** subdirectories which contain individual 1-degree tiles for the period for the north half of the Murray-Darling Basin (-26 to -31 degrees latitude) and the south half of the Basin (-32 to -38 degrees latitude).
- **20150401** contains the median images (**PCT_50**) for the 2016/2017 model, as well as the upper (**PCT_75**) and lower quartile (PCT_25) images. Each of these subdirectories contains a **North** and **South** subdirectories which contain individual 1-degree tiles for the period for the north half of the Murray-Darling Basin (-26 to -31 degrees latitude) and the south half of the Basin (-32 to -38 degrees latitude).

ModelInputs

This directory contains the database used to control the modelling process, the models used by the software and the static input grids used by the models to create the output maps.

- *MDVegMonitor.mdb* is the controlling database which stores how models are created and the list of models which have been created. When this file is opened for the first time by the MDVM software, the directories table are adjusted to point to the containing directory, such that the software knows where to find the model and static input files required for modelling. A copy of the database is included (*MDVegMonitor - Copy.mdb*) in case the original becomes corrupted by moving files/directories while running the software for the first time. If this happens, delete the *MDVegMonitor.mdb* file and rename the copy to *MDVegMonitor.mdb.*

- Three sets of .out, .py and .arff files are present, each set defining a different vegetation monitor model in Python script.
- ***AutumnLandsat*** is a directory containing a long-term median Autumnal Landsat Images of the entire Murray-Darling Basin. These two images form a base layer from which change in condition may be detected. These files may also be viewed with the image viewer either individually or as a mosaic.
- ***MDBAVegNative*** is directory containing native/non-native, created in 2013 and is used by the model as an input. It is split into North and South similarly to the Landsat tiles. This model contains four layers VegClass, WaterProb, NativeProb, NonNativeProb. The individual tiles can be viewed with the image viewer. Alternatively, the whole Basin model can be viewed with the image viewer by selecting the directory, which instructs the viewer to mosaic the two tiles together.
- ***SppModels*** is a directory containing the basic tree species model create of the Basin in 2013. These models have nine layers (SppWoody, Wetland, Dryland, RiverRedGum, BlackBox, Lignum, RiverCooba, Coolibah, RiverOak) and can also be viewed with the image viewer.

ModelOuptuts

This directory contains outputs of the MDVM software. Twelve models of the vegetation condition across the Basin are supplied. These use the 'Veg Monitor Map All Years' model provided in the software for the years 2009, 2010, 2012, 2014, 2015, and 2016/7 for the North and South sections of the Basin. In addition, a standalone version using the 'Veg 2016 and 2017 Only' model applied to the 2016 satellite images are supplied for the North and South of the Basin. Each of these images may be viewed with the image viewer in the software.

ModelAnalysis

This directory contains some of the analysis files that will be used to create the model report. These CSV files contain statistics on the fit of the overall model for each of the year's observations for the Test data (20% of the field observations).

This directory also contains a ***RawData*** directory which lists all field data used to create the models.

ModelSoftware

This directory contains the MDVM software and files required to be present for the software to run. AccessRuntime_x64_en-us.exe must be run to install Mircrosoft Office Libraries on the machine before MDVegMonitor.exe is run. If this is not done, the MDVM will not run successfully.

Software Installation

The Murray-Darling Vegetation Monitor tool was designed to be operated in a variety of environments, even those where the user does not have Administrative rights to install software on their machine. In this case, it can be run from the supplied disk.

The software package consists of three parts:

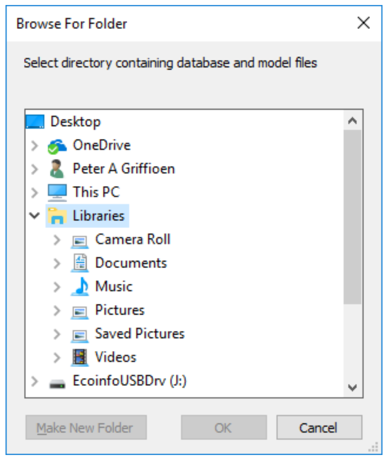
1. The MDVegMonitor.exe program. This program and the files required by it are in the ***ModelSoftware*** directory of the supplied disk. This software may be run from the disk by clicking on the MDVegMonitor.exe file. Alternatively, the entire contents of this directory may be copied onto the local machine to another directory you have created. Eg. ***C:\Program Files\MDVegMonitor\.*** From here right click on the program to add a shortcut to your desktop.
2. The MDVM input files and database. These files reside on the supplied disk in the ***ModelInputs*** directory. You may leave the directory there if the disk is always going to be present when the tool is run. However, it is recommended that you copy this entire directory (approx. 70GB) to your local machine of server. Note that this directory contains the software's controlling database which will be modified as the program is run. Thus you must have write privileges over the directory it is copied to.
3. The Landsat files to convert to vegetation condition model grids. These numerous and large files are located in the ***MDBALandsat*** directory of the supplied disk. These files were used to create the models used by the tool and also to create the original vegetation model grids for each of the epochs of field observations. These girds are not required for the software to operate unless you want to use them as examples to run the tool and create outputs.

Prerequisites

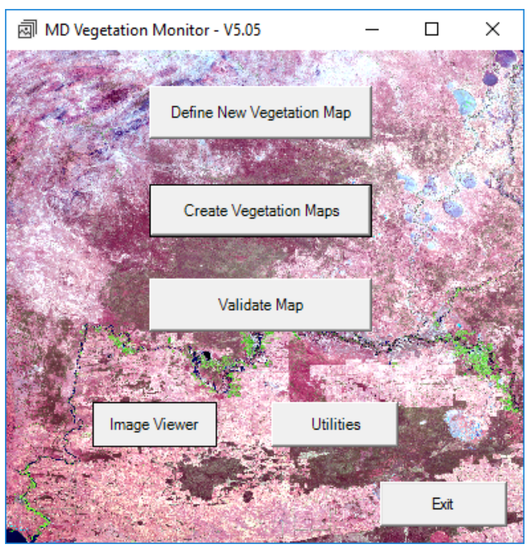
The MDVM tool requires a MS Windows 64-bit operating system is present and that the Microsoft Access Data Engine is installed. In the ***ModelSoftware*** directory of the supplied disk, run the AccessRuntime_x64_en-us.exe program. Alternatively you can download this program from Microsoft at https://www.microsoft.com/en-au/download/details.aspx?id=39358.

Starting the software

After installing the *AccessDatabaseEngine_x64.exe* software, the MDVM program may be run directly from the supplied disk, or from the location it has been copied to. On first running, the software will ask for the location of the ***ModelInputs*** directory. Once this is located, the software will locate the controlling database and input files it needs. This location is stored in the registry and will not be asked for again unless the database is not found. If the ***ModelInputs*** directory is subsequently moved, the software will again ask for its location.



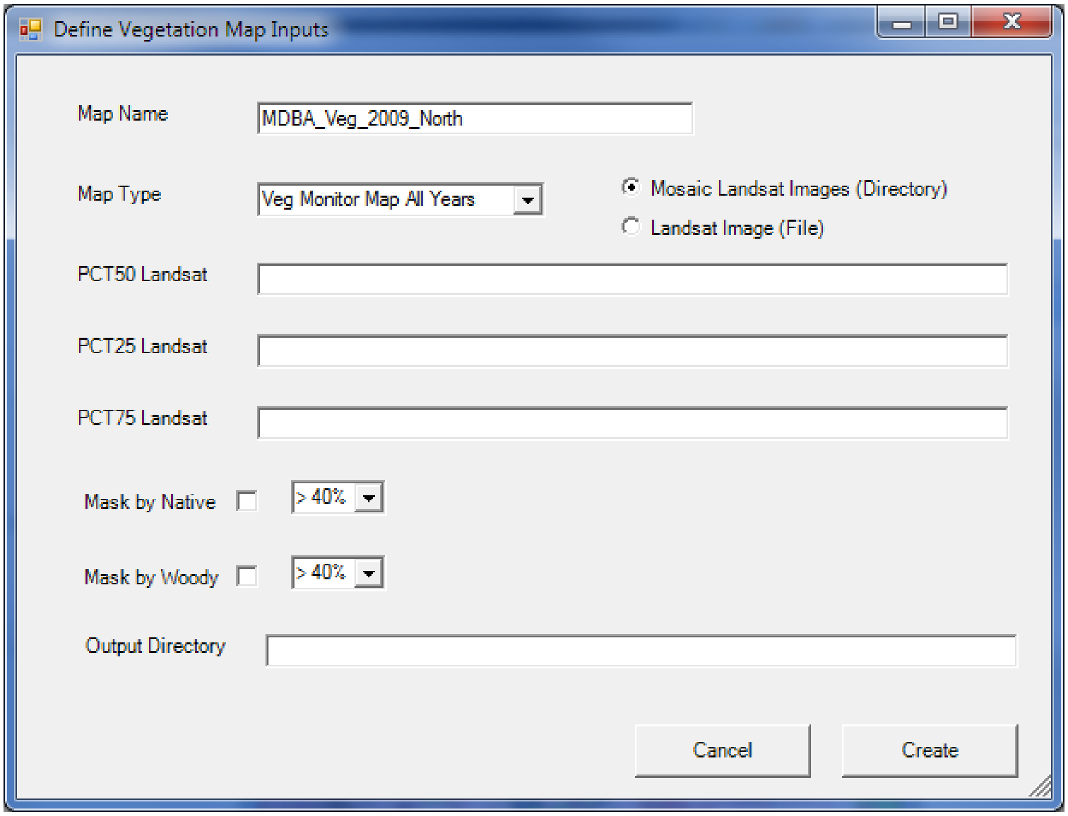
The main menu of the application is shown below:



Create Vegetation Quality Map

Making models of vegetation condition is a two-step process. Firstly, the input Landsat image/s must be identified and the options of the model set. These instructions are then stored to the controller database *(\ModelInputs\MDVegMonitor.mdb*). Secondly, the model calculator is started running on the selected batched models.

From the main screen, select 'Define New Vegetation Map'.



On this screen you are required to provide a map name, select the model type, define the one or three input Landsat files/mosaics, set some masks if required and define an output directory.

The Map name will become the filename. Only input valid filename characters should be used and best avoid spaces.

Three types of models are supported. The 'Veg Monitor Map All Years' requires 3 inputs, the median Landsat image (PCT_50), the lower quartile (PCT_25) and the upper quartile (PCT_75). Both of the other model types, 'Quick Veg Map All Years' and 'Veg 2016 and 2017 Only', require only a single Landsat image, be it the median image (PCT_50) or some other 6-band Landsat image in ENVI BIL format. The 'Veg Monitor Map All Years' is the more complex model and produces slightly better results regardless of the year of the image.

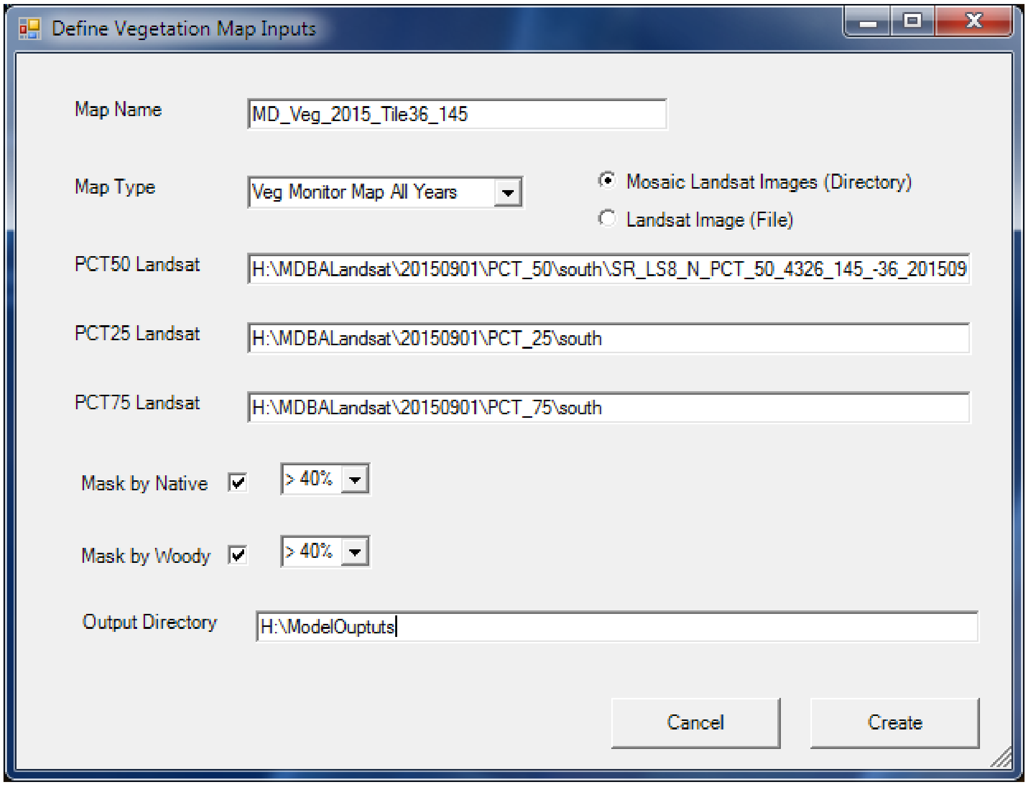
The 'Quick Veg Map All Years' may be used where you have a single image for any year. Note that the bands are expected to be labeled:

band names = {Band 1,Band 2,Band 3,Band 4,Band 5,Band 6}

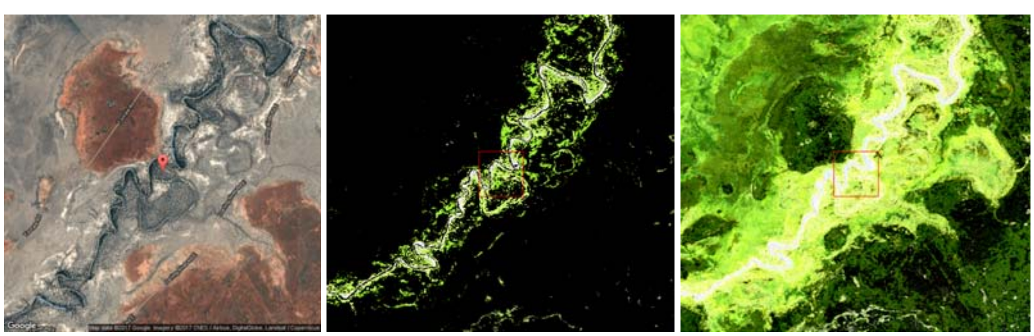
in the header file. The projection must be Geographic (GDA_94).

The 'Veg 2016 and 2017 Only' is a single image model for 2016 and 2017. It performs very closely to that of the 'Veg Monitor Map All Years' model.

When defining the Landsat images to be used, you may select an individual tile or a complete mosaic. A mosaic is a directory of Landsat tiles to be stitched together into a single image. The toggle switch controls the directory/file navigation to select either a mosaic or tile. You may even mix the two. In the example below, the PCT_50 input is a 1-degree tile while the PCT_25 and PCT_75 images are left as mosaics. The size of the PCT_50 image controls the modelling process such that the model output will be the size of the PCT_50 tile.



Masking of the output map may be set by selecting the mask options. This option is supplied to mask out areas of the map where the model should not be applied, such as on irrigated pastures or in urban areas. The masks will zero-out the map where the Native estimate, as per the *ModelInputs\MDBAVegNative* map, does not reach the required probability. Similarly for the woody estimate and the *ModelInputs\SppModels* input map. Note these input maps are only models and are not perfect, and are limited by the field data that was used to develop these models at an earlier stage.

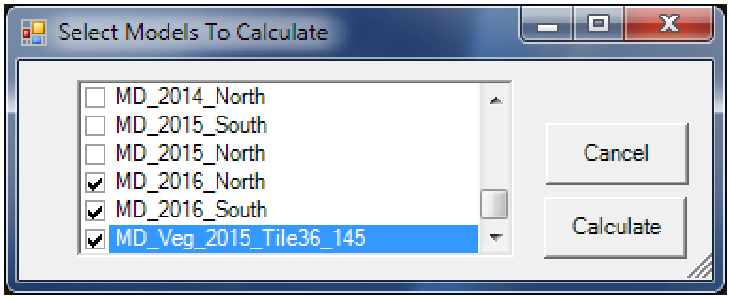


Google image of the Darling River and masked output of the model compared to unmasked output.

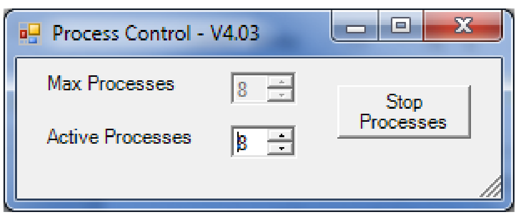
The final step requires the user to define the output directory for the results and to select the Create button. The model is then written to the database. It is recommended to define as few models at a time, say the North and South MDBA maps, and run all of these in a single batch in the next step.

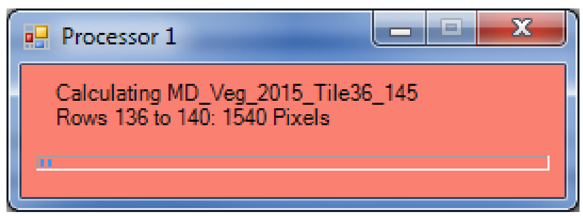
Create Vegetation Maps

The maps may be calculated by selecting the ‘Create Vegetation Maps’ button on the main screen.



Select from the previously defined models which ones that you would like to calculate and then click calculate button. Depending on the power of your PC, up to eight processors will then run to compute the output.





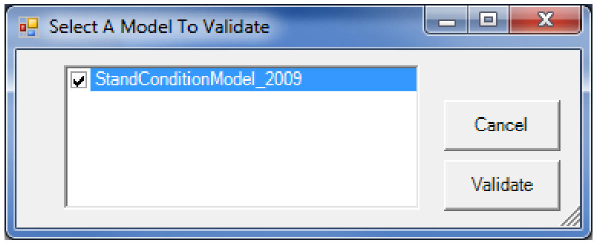
Processor 1 displays the progress bar of the overall calculation which may take between one and four hours. Once all of the Processors have finished, click on the Stop Processes button on the Process Control form to return to the main form.

If you need to use the computer for other purposes while the model is still calculating, the processor load of the Stand Condition Tool may be reduced by reducing the number of Active Processors on the Process Control form. If the Stop Processes button is clicked before all of the processors have completed, upon restarting the same model you will be given to option to continue from the current progress point.

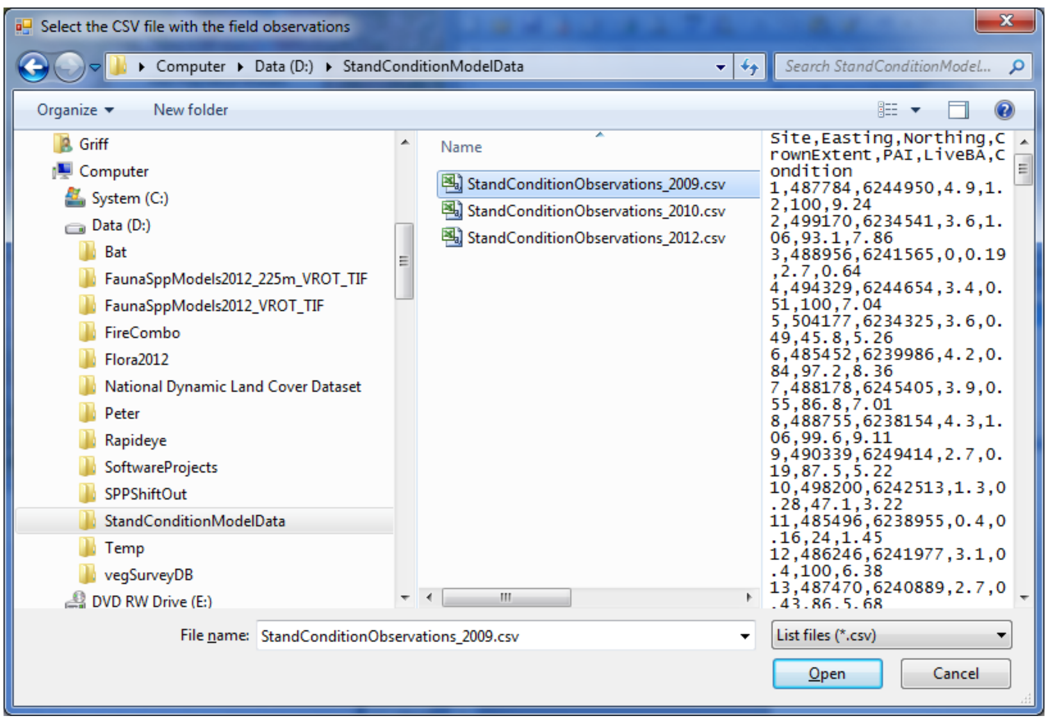
Validate Map

Stand Condition Maps created with the Murray Stand Condition Tool can be validated and improved with the addition of field observations. The validation tool will create two reports quantifying the fit of the model to the observations. Depending on the results, the validation data may be used to improve the model fit and generate a new map of condition indices. This may be useful if the tool is subsequently applied to a year with substantial field observations and with conditions that vary from the years of data supplied to train the tool. Thus the validation process can be used as a post-processing step applied to the models supplied with the tool to further enhance the maps produced by the tool.

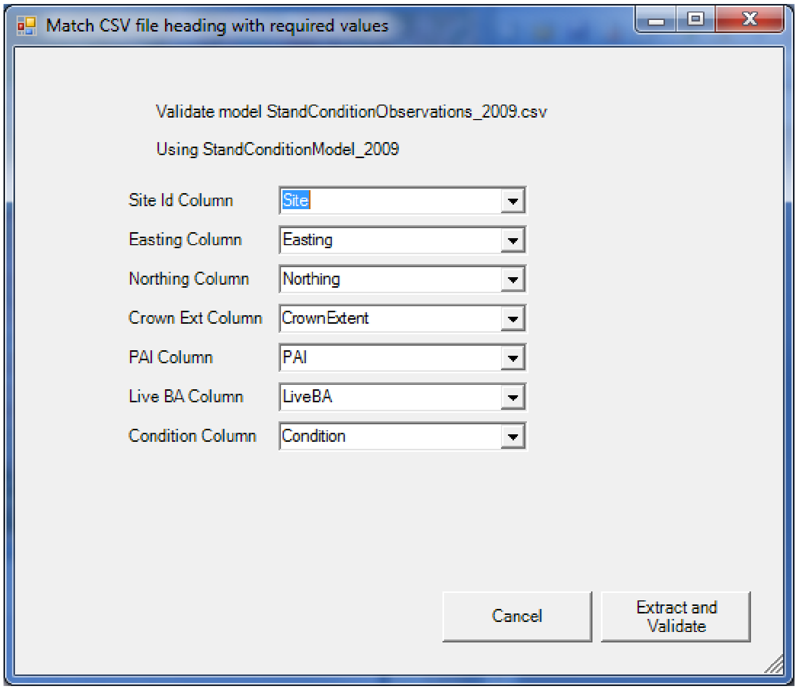
Start the validation process by selecting a previously created condition map to validate.



This map must be matched with the CSV file of field observations which are used to validate their respective maps to generate statistics on overall fit.



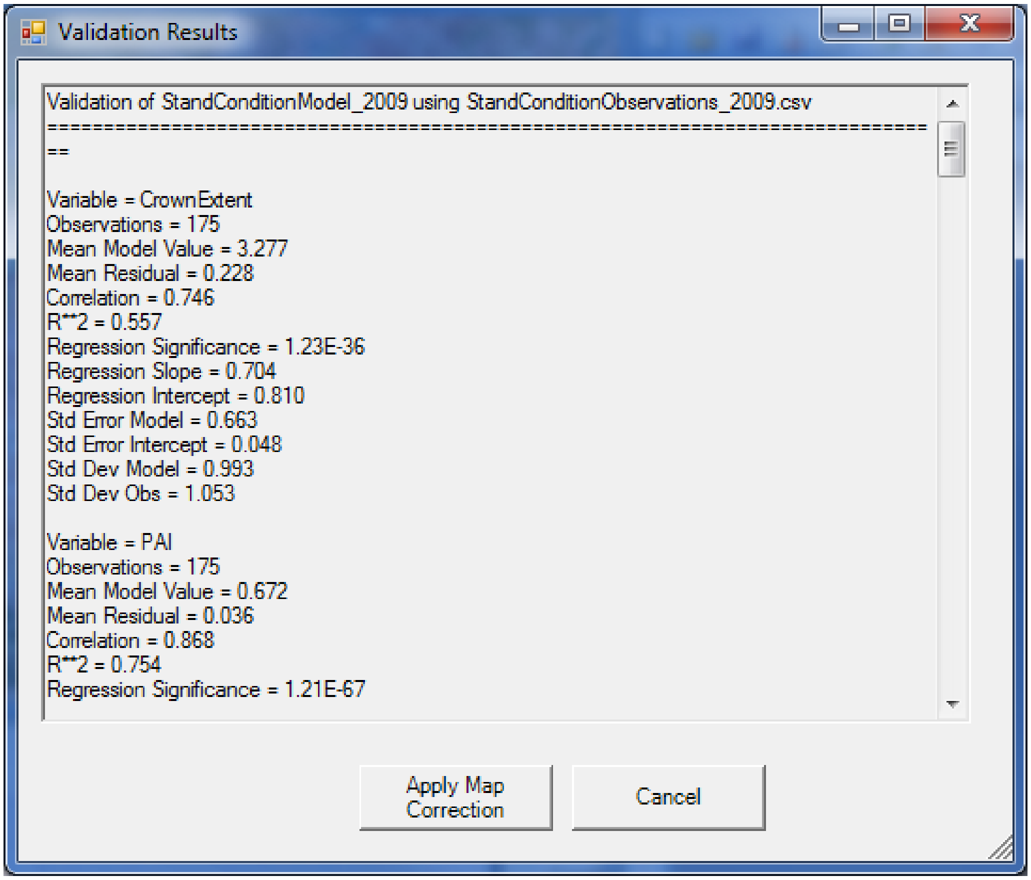
It is essential that each of the field measurement be matched up column for column with the data expected by the validation process. The next screen allows this to be checked and adjustments to be made if necessary.



Once the corresponding values have been matched the coordinates of the validation sites are used to extract the condition values from the map. These values are then compared to those of the validation file and two reports are generated in the same directory as the CSV files; '*ConditionModelName*_Report.csv' and '*ConditionModelName*_Predictions.csv'. These files contain the model fit statistic report and the model predictions of the supplied field observations compared to the map predictions.

Note that is it possible to validate maps which do not cover the spatial extent spanned by the field observations. For example, if a file of the 2016/7 field observations is used to validate the southern half of the Murray-Darling basin map, observations in the northern half could not be extracted from the map to compute the statistics. In this situation, only matched values are used to compute the statistics.

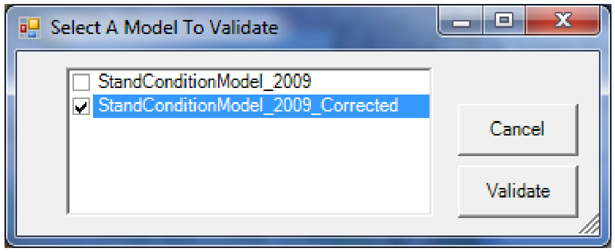
A similar situation may also arise if masked maps are used to validate the tool. The woody and native masks applied may have zeroed out some of the sites on the map used for the validation. For this reason, the validation process should only be applied to unmasked versions of previously run models.



For each of the condition indices, CrownExt, PAI, LiveBA and Condition, statistics of the map versus validation data fit are listed in the Report file. A good fit is indicated by correlation value above 0.75. The offset (regression intercept) and scalar (regression slope) indicate how well the model is performing in realistic terms to the observed values. A large offset or a slope below 0.7 indicates the predicted values would likely benefit from a post-processed inverse-linear regression being applied. An inverse linear regression will generally have the effect of stretching the upper and lower boundaries of the predicted values towards the corresponding field observed values while not substantially changing the intermediate predicted values. Being a class of linear regression, the correlation coefficients of the model versus observed values will not change greatly but the utility of the mapped values will be enhanced.

To create the map with the inverse-linear regression correction applied, simply click the 'Apply Map Correction' button. This will add the corrected map to the current model list and run the new model. You will then be prompted for the location to store the corrected map, which will be given the name '*ConditionModelName*_Corrected'.

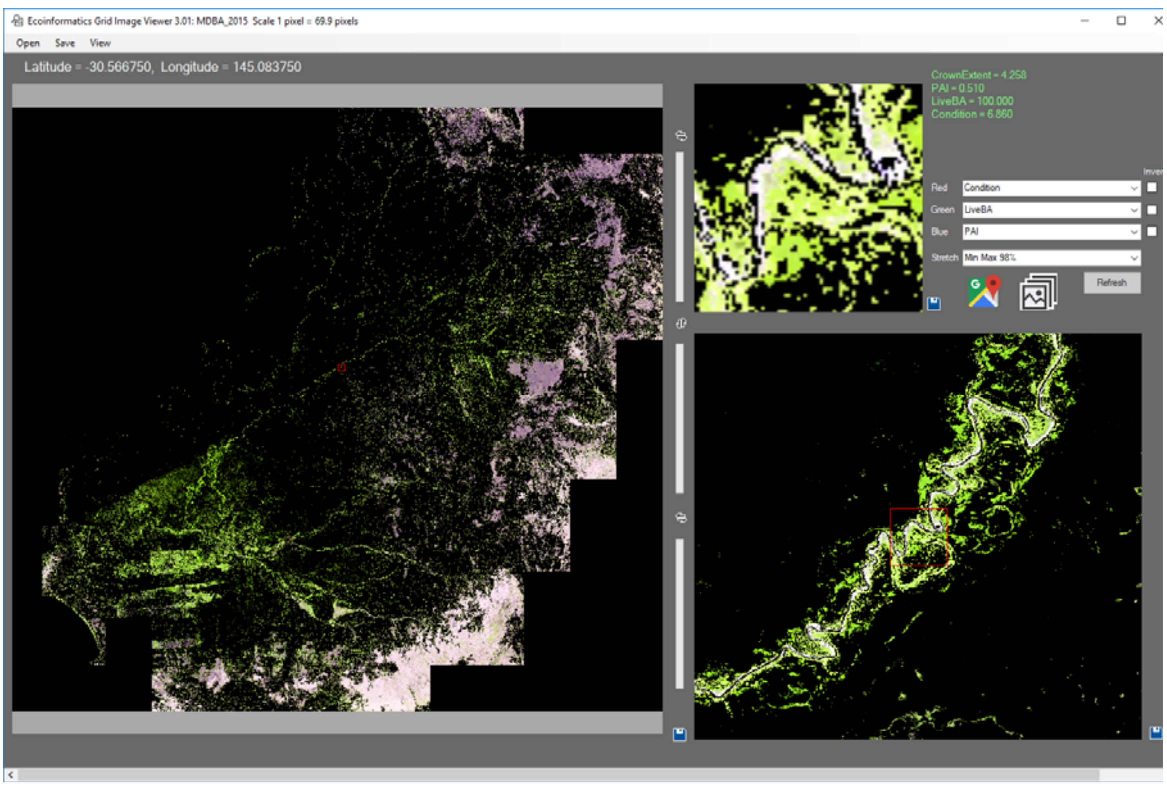
The corrected map may itself be validated against the same validation CSV file.



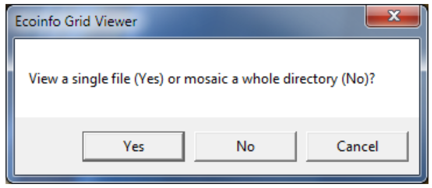
The resultant validation report will then show that each of the corrected map indices has an offset close to zero and a slope close to one indicating that no further improvement is possible.

Image Viewer

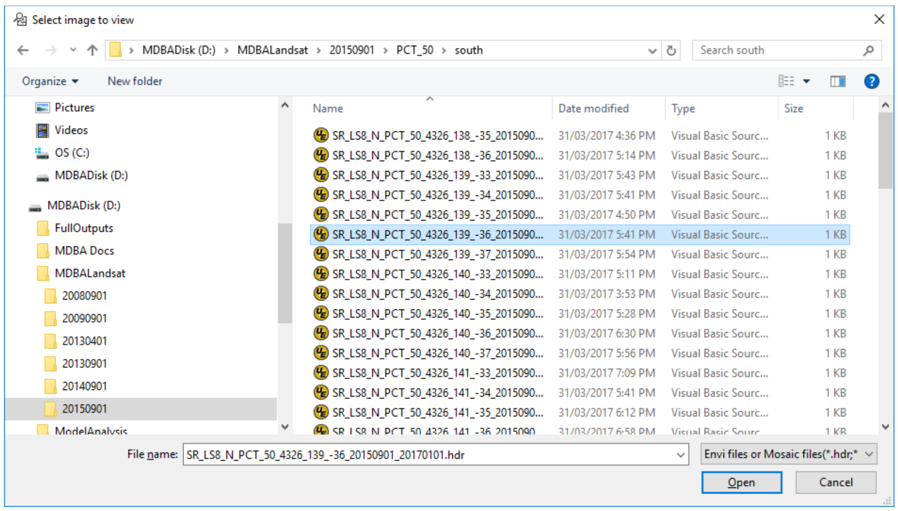
Satellite images and mosaics may be viewed using the image viewer. The viewer is handy to review the supplied data on the disk and model outputs and will display all of the image values at the current location of the cursor. The viewer can display a single ENVI file such as a 1-Degree Landsat tile, or mosaic whole a whole directory of files on-the-fly to make a single image. The viewer can display single layer files and multilayer files. For example, the ***ModelInputs/SppModels*** Directory contains mosaic of grids of nine layers. These files must be in ENVI format which is the default input and output format of the Murray Darling Vegetation Monitor tool.

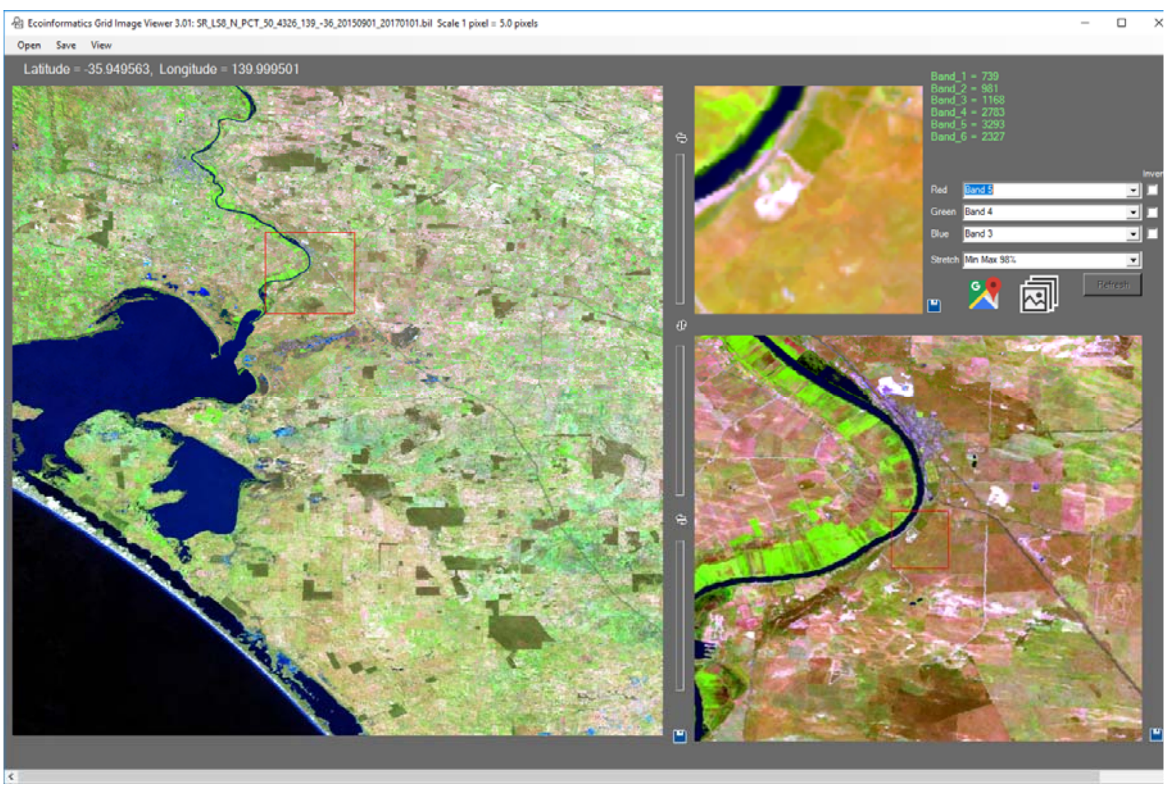


To run the viewer, click on the Image Viewer button on the main screen. The software will ask if you want to view an individual file (Yes) or mosaic a whole directory of tiles together to form one image.



Either select a .hdr file or a directory containing Envi files and view the image





The image viewer displays an overall navigation image which reports its scale in the header, the main 1-to-1 pixel view and 4-to-1 zoom view. Click on the overall map to move both the main and zoom views. Click on the main view to move the zoom window. Right mouse click on any map will save that location to the clipboard.

You may swap the three images between the three displays by clicking on the arrow icons between the displays. For example, you can swap the overall and main image such that the main image is the largest image.

The linked images button will open another image viewer which will allow you to open another image geographically linked to that already open. In this way you can compare two model outputs from different years or view a satellite image input file against it model output file. Each window will report all image values below the current cursor location allowing easy comparison of numerical values.

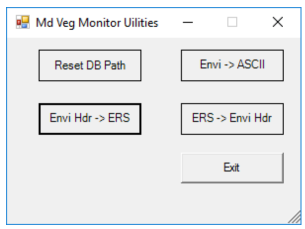
The Google button will open another window showing the zoom area in Google Maps. This may help you navigate as this map contains road and place names. This window has its own zoom bar allowing you to zoom in and out. Clicking on the Google map will show a corresponding cursor location on the main or zoom image. Shift click on the Google map will move the zoom and main image to be centered on that location.

You may adjust which bands are displayed as the RGB colours of the images. With Landsat images, RGB values of bands 5, 4 and 3 produce easily interpretable images. Selecting different bands enables the refresh button, which causes a refresh of the images to reflect your choices. You may also invert the colours with the Invert checkboxes. On-the-fly satellite indices may be switched on from the view menu if a Landsat image is displayed. Selecting this option forces a complete reload of the image. The indices can then be selected as the bands for display. For example, TCBRIGHT, TCGREEN and TCWET make for an interesting RGB display.

Bitmap, Jpeg or PNG format pictures of each of the images made be made by clicking on the save to disk icon associated with each image or from the View menu. To facilitate output of high-resolution images for publishing, the image size of the large image may be changed in the View menu. Image sizes ranging up to 3840x2160 pixels, which are suitable for a 4K display, may be made.

Utilities

The MDVM tool contains some simple utilities, the most important being the Reset DB Path and the others providing header files and ENVI file to ASCII file conversion.



Reset DB Path

The path to the Access database used by the software is set the first time the software is run on a computer. This path is stored in the computer's registry such that the software knows the location of the database in the future. When the database is located, it is opened and the location of the ***ModelInputs*** directory is recorded in the database.

If the location of the ***ModelInput*** directory is changed or the directory is renamed or a new database is supplied, then it will be necessary to reset the database path. Simply select this button on the Utilities form and then chose the new location of the ***ModelInput*** or equivalent directory. The registry and database will both then be updated. Note that if models have been added to the database prior to it being moved, these models **will not work**. Only the paths to the files required to make new models are updated in the database. Paths to Landsat files required by models and supplied by the user are not updated. Advanced users are welcome to open the MDVegMonitor.mdb and manually edit the entries in the Directories table to get existing models working again.

Envi Hdr -> ERS

This simple utility takes an ENVI .hdr file and creates the equivalent .ers file which may be useful in ERSMapper or ArcGIS.

ERS -> Envi Hdr

This utility takes an ERS header file and creates the equivalent ENVI .hdr file which may be useful in making available BIL formatted ERS satellite image files to the tool.

Envi -> ASCII

This utility takes an ENVI file and converts it to ArcGIS ASCII file/s, one file per layer.