A Brief History of MFworks and Raster Based Mapping

In 1983, C. Dana Tomlin, at Yale University, published a dissertation that outlined the Map Analysis Package (M.A.P.) and proposed a software application based on a series of raster map processing operations known collectively as map algebra. Tomlin later went on to produce the M.A.P. software while at Harvard University. One of the direct descendants of Tomlin’s M.A.P. is MAP II, a map processor produced by the creators of MFworks. MAP II was designed both as an instructional tool and as a multi-purpose research tool. MAP II is a good educational tool that will run on the complete family of Macintosh machines from Mac Plus on up. MAP II is currently available free of charge from ThinkSpace Inc. from our web site at:

www.thinkspace.com

Map•Factory succeeded MAP II with significant design changes that avoided the limitations of its predecessor. Map•Factory built on the strengths of MAP II, but far surpassed it in flexibility, speed, functionality, quality of output, and range of applications. Map•Factory had built-in scripting capabilities that allowed you to combine several operations in sequence, whereas MAP II could only perform one operation at a time. Map•Factory scripts could be saved and used repeatedly for different applications. Map•Factory scripts could be considered map processing models.

Map•Factory 2.0 introduced a new, slick dialog box interface that allowed you to quickly and easily perform operations one at a time or to interactively build scripts and spatial processing models. Map•Factory 2.0 also expanded the operations and translators suite to give you even greater map processing power and data interchange flexibility.
MFworks brings the power, flexibility, and ease of use of Map•Factory 2.0 to Windows 95/98 and Windows NT users. MFworks is positioned to become the dominant raster analysis engine on the Intel Pentium platform.

The purpose of MFworks is to provide low cost, high-tech tools for people who have both simple and complex spatial problems to solve. On Motorola based processors, MFworks is designed for higher end Macintosh machines (Macintosh II or newer). MFworks runs in native mode on the Power PC. On Intel based processors MFworks is designed to run efficiently on Pentium II machines, but will also operate on 486DX machines under Windows 95.

**The Map Layer**

A Geographic Information System (GIS) is a database of attributes, known as spatially referenced data, that have been measured or recorded as they occur over the area being mapped. The purpose of a GIS is to locate and analyze these attributes and present them in map form. Maps are a representation of the distribution of attributes over a specific geographic area. Spatially referenced data are values that represent the nature, form, presence, absence, condition, size, elevation, or relative position of attributes at specific points or within specific parcels of geographic space.

For example, attributes may be the direction that a slope faces, the number of oil wells per hectare, the suitability of land for agriculture, the presence of an archaeological site, the steepness of a road grade, or the degree of flood risk. The surface being represented is typically the Earth’s surface however, it can also be the surface of another planet, the interior of a building, a Cartesian graph, the surface of a microscope slide or any other surface from which information can be collected.

Each attribute is stored in its own file along with information on how that attribute will be displayed in map form. Attributes in a map database are normally measured or recorded over the same geographical space. Together they become a stack of attributes that describe a particular rectangle of geographical space. An attribute map can be thought of as a layer within a stack of related maps, or simply, a map layer.

The map layer is the file that MFworks uses to store spatial data along with a set of information that describes and defines the map. Each map layer is stored as a self contained unit it is linked to other map layers by a project,
which is a map layer reference list of related map layers (refer to Building a Map Database).

Maps are stored and displayed as a two-dimensional rectangular matrix, or grid, of square cells. Each cell in the map contains a value that describes an attribute of a parcel of real world space.

**Cell values: The third dimension**  
The cell is the fundamental unit of the map layer; it contains a single value that represents a real world attribute.

The map layer is composed of a rectangular grid of cells measured out by a two-dimensional (x, y) coordinate system giving each cell a location relative to an origin and to all other cells. The value contained in each cell of the map layer lends a third, often non-spatial, dimension to the map. The cell’s value, z, can describe any attribute within the real world parcel represented by the cell. For example, the cell’s value might indicate a river or a road, the number of households voting for a particular political party, the distance to the nearest hydro-electric generating station, or the elevation above or below sealevel.
There is also a special data value called “VOID”. “VOID” is the absence of data. It is different from zero in that zero is a value which can be used in computations, whereas “VOID” indicates the absence of a value.

**Raster data characteristics**

A raster based GIS imposes a grid structure over a surface. A raster map is a matrix of values, describing attributes, that are stored in cells that represent equal sized parcels of the real world. A raster based GIS is more powerful than this simple model implies. The cell can have an internal structure as well. Raster based GIS is not limited to mapping area characteristics such as forest cover, land use, or water bodies. It can also map objects with linear characteristics, objects with point characteristics, and objects with surficial characteristics.

**Area Data**

Raster based GIS lends itself well to portraying “area” characteristics as area is what is implied by the data structure. Area is defined by a length and a width and is given as a number of units square. In MFworks, the cell resolution, or real world distance represented by a single side of a cell, can be set to millimetres, centimetres, metres, or kilometres. A cell resolution of 22 metres means that each cell represents 484 square metres (22m x 22m) of real world area.
Linear Data

In the real world a road is a continuous linear feature, but in raster based GIS a road is composed of a series of adjacent cells all with a value that represents “road”.

Cell structure of a linear feature

Each black cell in the diagram above represents a segment of the road. Hence, there is an implied underlying structure to the road cells that looks like this:

The implied linear characteristic of “road” cells

Spatial Accuracy

In this example, the road passes through the centre of each cell. This does not necessarily reflect reality because roads are rarely completely straight over their entire course. What is implied by the linear characteristic of the cell is that somewhere within the cell there is a segment of road the whole cell is not a segment of road. The same is true for attributes with point characteristics. It is implied that somewhere within the cell the point data occurs.

When measurements are performed in raster based GIS, the assumption is that line and point data fall in the centre of the cell; this has ramifications for accuracy of measurement: it is possible for significant measurement errors in distance and direction to be made if the scale of the map is small or if the point is located far from the cell centroid. Tomlin does an excellent
job of defining all the possible measurement concerns that result from raster
data structure in his book Geographic Information Systems and
Cartographic Modeling, published by Prentice-Hall, Englewood Cliffs, NJ,
1990, pages 24 through 39.

**Point Data**
Point data cells could represent bus stops in Paris, archaeological sites in
Southwestern Ontario, or towns in Maine. The point data can be located
anywhere within the parcel of real world space represented by the cell, but
for measurement purposes MFworks assumes that the data points occur in
the centre of the cell (called the centroid). Therefore, when measuring the
distance between two points on a map of point data, the distance will be
measured from centroid to centroid. This, of course, affects the accuracy of
the distance measurement. The precision of the cell resolution affects the
precision of the distance measurement (refer to [Accuracy versus Precision](#)).

**Surficial Data**
A more complex and less intuitive concept is that of surficial characteristics.
Surficial data can represent either a three dimensional vertical component
such as elevation above sea level, or non-spatial components such as surface roughness, downhill direction, pollutant concentration, and so on.

Data type
The value contained in a cell must be a number. MFworks supports both fixed point (long integer) and floating point (single precision) data types, but a single map layer can contain only one or the other, not both. Map layers in the same project can be of mixed data type. Refer to Mixed Mode Arithmetic in the document Algebraic Statements.

The value range for fixed point data is ±2 147 483 647. The value range for floating point data is ±1.0E38 with 6 decimal points of comparative precision. Please refer to the document Scripting for an explanation of the way that fixed point and floating point data behave in MFworks operations.

Data value types
There are four types of data measurement that can be used for cell values: nominal, ordinal, interval, and ratio. MFworks uses these four types of measurement to represent different kinds of data.

Nominal Values
A nominal value is a number that represents a quality rather than a quantity. The numbers have no linear or mathematical relationship to one another. However, they do have a Boolean relationship to one another.

Nominal values are labels or identifiers that distinguish one observation from another. Telephone numbers and U.S. postal codes are examples of nominal values. They identify a location uniquely with a tag or label. Nominal values can be used to distinguish farms with corn versus farms
with alfalfa cities with more than 100 days of sunshine per year versus those with less, or high tide conditions versus low tide conditions.

Nominal values can be used in Boolean statements, cross tabulations, overlay, and summary statistics operations, but not mathematical operations.

Ordinal Values

Ordinal values indicate a sequence of ranked values on a linear scale however, rank gives no indication of a measured quantity, that is: how much higher the value 1 is than the value 2. Ordinal values do not permit mathematical comparison. However, you can perform Boolean logic, cross tabulation, overlay, and summary statistical analysis and comparisons. The neighbourhood with the highest school tax might be ranked “1” and the neighbourhood with the second highest might be ranked “2” on a map of the distribution of school taxation in a city, but there is no indication of precisely how much more tax one neighbourhood pays than another.
Interval Values

Interval values use a linear scale where each value’s position is relative to all other values, but is not relative to a fixed point. For example, when time is measured in intervals of years, there is no absolute year zero. Someone born in the year 1960 A.D. is not twice as old as someone born in the year 980 A.D. Temperature is also measured on an interval scale. 20°C is not five times as hot as 4°C, it is, however, 16°C greater. Interval measurement is additive.

Ratio Values

Ratio values are based on a linear scale with an absolute zero. Travel times between the downtown core and different neighbourhoods, the distance between fragments of forest, and elevation above sea level are all examples of ratio measurement. Ratio data can be used to perform computations such as multiplication and division. Any mathematical function can be applied to ratio data in a meaningful way. For example, increasing the speed limit on a highway from 100 km/h to 110 km/h will reduce travel time by 10%. Age, monetary value, income, and frequency of occurrence (i.e., of a particular plant species) are other examples of ratio measurement.
This figure demonstrates highway distance from a city centre. The lighter the colour, the greater the distance from the city centre.

**Cell coordinates**  
Cells in a map layer have a position relative to one another and to the origin of the map (refer to *The origin* below). Each cell’s position is referenced by row and column coordinates like the x, y coordinates on a Cartesian plane. By default, rows are numbered sequential from top to bottom, and columns are numbered sequential from left to right.

Keep this in mind when creating subscenes of map layers or when mosaicking map layers.

A secondary, or real world, coordinate system can be imposed on top of the cell coordinates by using the MFworks Georeferencing commands. The secondary coordinate system will affect distance measurements made with the Tracker and the grid system that can be overlaid in the Layout window.
The origin

The origin is the starting point on a map; it is the cell from which rows and columns are counted. In MFworks, the top left-hand corner of the map is the origin. Rows are counted from top to bottom beginning, by default, with “0” and columns are counted from left to right beginning, by default, with “0”.

However, the initial row and column numbers of the origin can be reset to any value between -32 000 and 32 000 by using the Information window.

The MFworks Cover operation allows you to overlay one map layer on top of another. Sometimes only a portion of each map layer will overlap the other. For the two or more map layers to overlap and align with each other properly, ensure that the relative origins are set correctly in the Information window.
You may also want to analyze only a portion, or subscene, of a larger map. Use the Subscene operation to extract the subscene. The origin of the subscene will be offset relative to the origin of the full map scene.

Cell resolution and map resolution

Every cell in the map layer has the same dimensions therefore, every cell represents an equal sized parcel of space in the real world. The only time that this is not true is if a small scale map is being used or if the map is not an equal area projection (refer to Accuracy versus Precision). As far as MFworks is concerned, map projection is ignored. **MFworks assumes that all cells have equal area.** It is important to keep this fact in mind when choosing the scale and projection of the map that is to be used for analysis.

Cell resolution is a ratio determined by the spatial dimensions of the cell relative to the spatial dimensions of the corresponding parcel on the ground. Resolution is defined by the number of units of distance (*e.g.*, metres) that are represented by a single side of a cell.

For example, if the parcel of land represented by the cell is 75m by 75m, then the cell resolution of the map layer is 75m.

Map resolution is determined by counting the number of cells there are *in a given distance* on the map. The easiest way to do this is to measure a known horizontal or vertical distance in the real world, then use the selection tool to determine how many cells there are in that distance. For example, if there...
are 75 cells in a 2.3 kilometre known distance, then the map resolution is 32.6 cells/km:

\[
\text{75 cells/2.3 kilometres} = 32.6 \text{ cells/km.}
\]

The corresponding cell resolution would be:

\[
\frac{2.3\text{km}}{75 \text{ cells}} = 0.0307\text{km/cell} = 30.7\text{m}.
\]

Note: cell resolution is the inverse of map resolution.

**The zone** Cells that represent the same attribute are assigned identical values; cells with identical values are called zones. For example, on a land cover map, all the cells whose values represent coniferous forest are members of the same zone. **A cell can only be a member of one zone since a cell can contain only one value.** A zone can be as small as one cell or as large as a whole map. **A map layer has as many zones as it has unique values.**

**The legend** All the zones that exist within a map layer are listed in the **legend**. The legend can be thought of as an attribute table. Each zone entry contains the four characteristics associated with a zone: the zone colour, the zone value, the number of cells (or area) in the zone, and text describing the zone. Zones are listed in numerical order by value in the legend. The legend is one of the views, or windows, associated with the MFworks map layer. Like the map itself, it is a way of portraying and summarizing the map data. It provides a
visual tie between the colour information in the Map window and the underlying numerical data.

Azimuth

By cartographic convention, maps of the Earth’s surface are oriented according to a system of lines of latitude and longitude called a graticule (these are the lines that are seen on many maps and most globes). The graticule makes it possible to determine direction on the map. Lines of latitude and longitude are aligned perpendicular and parallel to an arc drawn from pole to pole. The intersection of lines of latitude and longitude create a grid of coordinate points by which position can be determined.

The North Pole of the Earth is called Geographic North (also called True North) and the South Pole is called Geographic South. When facing Geographic North the azimuth is said to be 0° when facing Geographic South, azimuth is said to be 180°.

MFworks map layers are oriented relative to Geographic North by azimuth (Note: refer to the Terminology document for the distinction between Geographic North and Magnetic North.). The azimuth of a map layer is the direction that the top of the map faces relative to Geographic North. North is 0°/360° on a compass. Direction on a compass is measured clockwise. If the top of the map faces 330° (or -30° off North), then the azimuth of the map is 330°. The azimuth of a map layer is set in the Information window.
Building a Map Database

The map layer stack

The map database is a collection of associated map layers. Each map layer in the database contains data for one attribute or a set of attributes (e.g., road network, water bodies, surface elevation, populated places, etc.). Normally the map layers in a map database have a common geographic coverage.

The map database can be thought of as a stack of map layers, all with the same azimuth, cell resolution, and at least partially overlapping areal coverage.

In traditional cartography and map analysis, attribute layers were created on sheets of transparent acetate (plastic film). The layers were placed on top of an opaque basemap that depicted geographic features such as road networks, topography, and property boundaries. The layers of acetate had attributes placed on them in coloured translucent plastic film. The analyst could then see through every layer to the basemap allowing him or her to see how each layer related to the underlying Geographic features and to each other. The analysis that could be performed on this stack of acetate attributes was limited to visualization of the relationships among layers.

In raster GIS, this layer concept is carried over to the digital map where much more powerful analyses can be carried out on the map overlays. MFworks comes with a set of operations and functions that can be performed on one or more of the map layers in the map stack. These operations and functions, discussed in detail in the Scripting, Operations, and Algebraic Statements documents, are used to explore mathematically the interactions of zones within and among map layers.

Database organization: The project

In MFworks, the map database is organized into projects. The project is a reference list of map layers. It is stored as a separate file that contains the names of all the maps in the database and pointers to their location on disk. Normally the map layers in a project are all of the same or overlapping
geographical area. A map layer can be associated with any number of projects at any given time, or not be a member of any project – although the latter case is rare.

Because the project does not physically contain map layers, the project file can be stored in any folder or disk on the system. (Note: Map layer references will be dropped from the map layer reference list if a project is opened that references map layers stored on a removable or external disk that is not connected to the computer running MFworks). MFworks must be able to locate the map layers listed in an open project using Windows Explorer to reference them.

It is a good idea to use a hierarchical structure of folders to organize map layers rather than keep them all in one unorganized folder. For example, keep all basemap layers in one folder, all intermediate maps in another folder, and all the analysis results in still another folder. However, you should choose an organization system that works best for you.
Map Sources: Importing & Creating Maps

Before map analysis can be performed with MFworks, maps have to be obtained and converted to a form that can be used by the application. Maps can be obtained from a wide variety of sources, some of which are listed below. Many government agencies, university map libraries, and private institutions have digital maps, aerial photographs, satellite images, and paper maps available for purchase or use.

Most of the common digital map file formats can be imported directly to MFworks. Paper maps and airphotos can easily be scanned and imported as TIFF (Tagged Image File Format) files. Map data can also come from field survey data. The data can be entered into a word processor or spreadsheet as a coordinate pair and a value or as a matrix of points. You can then import this information into MFworks as a tab delimited text file.

Scanning and tracing

Access to a scanner or a digitizer is important if you want to make use of existing drawn or printed maps. Use a scanner to generate EMF (Enhanced Metafile Format) and TIFF files from hand-drawn maps, printed maps, airphotos, and printouts of satellite imagery. A digitizer can generate Boundary files (BNA) or XYZ files which MFworks can import and rasterize.

The MFworks EMF and TIFF file translators can import files that were created using drawing or image processing applications.

When scanning a map, choose a scanning resolution that is appropriate for the job. Most maps do not need to be scanned at 300+ dots per inch (dpi). The resolution of most paper maps is far below 300 dpi. Thus, it is usually sufficient to scan at 72 dpi to 150 dpi. The higher the resolution of the scan, the more storage space the resulting image will occupy, and the longer the processing times for operations and functions will be. A file scanned at 150 dpi will be displayed at screen resolution (usually 72 dpi for MacIntosh and 96 dpi for Intel based hardware), making it appear approximately twice its original size. To make the map output at its original scale, set the cell resolution correctly then scale the map in the Layout window.

Scanning at 1200 dpi may cause you to overflow the coordinate range of MFworks. The highest dpi that you should consider is 300 for printed maps and 600 for a 10" x 10" airphoto. Do not attempt to go beyond the actual resolution of the scanner. Many scanners have a true scanning resolution and an interpolated resolution. Do not use interpolated resolution.

Other sources

The translator modules that are shipped with MFworks are capable of translating data from a number of common sources. Satellite imagery can be
imported using the Raw Binary translator. This means that SPOT and Landsat TM imagery can be used, as well as many of the digital products available from the National Aeronautics and Space Administration (NASA), to create basemaps.

Other GIS programs may generate file formats such as Shape, XYZ, DXF, MIF, BNA, DEMs, or raw binary files that can be imported into MFworks. MFworks can perform powerful raster based GIS operations that are not available in vector based GIS programs. MFworks can also import maps created using the MAP II map processor.

Maps can be imported from spreadsheet programs using either SYLK (SYmbolic LinK -- the standard spreadsheet and database interchange format) or a matrix of points saved as a text file. The Tab Delimited Grid (TEXT) translator can import files generated by word processors and spreadsheets as a tab delimited list of points specifying row, column, and values, saved as a text-only file.

Sometimes a file format will be encountered for which a MFworks file translator does not yet exist. To import a file created by another GIS for which there is no MFworks translator available consult the reference material that came with the GIS to see if it can write the file in a format that MFworks can translate.

A module development kit is available for MFworks that allows users with programming experience and knowledge of PC and Macintosh and non-PC and non-Macintosh file formats to write their own translator modules.

Accuracy

Many people assume that since a map or a set of data has been published, it is correct. People place great faith in the maps they use, but how many times have they been driving on a road that does not exist on a map? How many times have they looked for a non-existent rest stop because the map says that one is there?

An unfortunate fact about maps is that they are created by people, and people have an incredible capacity for error. People make assumptions, misjudge a situation, forget to include information, “fudge” data to fit, and generalize detail. Human error is a significant problem that needs to be accounted for.

When someone collects data to create a map they may misidentify, mislocate, omit, mislabel, or mismeasure features. Measurement devices
may be incorrectly set or applied. Even the most experienced surveyor can misalign her or his survey equipment.

If using maps or data from a variety of international sources, keep in mind that different countries have different standards of map accuracy and precision -- and some countries may not have standards at all. As well, map standards may vary depending on the scale of the map. Much care may have been taken to map property boundaries on a 1:2000 cadastral map, while a 1:250 000 regional map may misplace features by tens or even hundreds of metres.

It is important to know the difference between accuracy and precision and the difference between large scale and small scale. These are two areas where many people new to GIS and Cartography often have trouble.

Accuracy vs. precision

Accuracy is a measure of how close something is to the truth. Precision is an indication of the fineness of a measurement. One person might measure the distance between two benchmarks as 202.59385022 metres. This measurement is precise to eleven significant digits however, if the true distance between the benchmarks is 210.1 metres, then the measurement is not accurate. Although 210.1 metres is not as precise a measure as 202.59385022 metres is, it is far more accurate. To help teach the difference between precision and accuracy, Dr. Smart at The University of Western Ontario, has his students envision four dart boards. The first dart board presents a picture of high accuracy and high precision, the second depicts low accuracy and low precision, the third shows high accuracy but low precision, and the last shows low accuracy but high precision:
The other pair of terms that are commonly confused are large scale maps and small scale maps. One way to recall which is which is to remember that large scale means more detail while small scale means less detail.

Scale is a ratio. A ratio of 1:1000 means that one unit of measure on the map is equivalent to one thousand units on the ground therefore, a feature on the map is 1/1000th of its true size. A map scale of 1:1000 means that 1cm on the map represents 1000cm (10m) on the ground. A map scale of 1:1 000 000 means that 1cm on the map represents 1 000 000cm (10km) on the ground.

Expressing these ratios as a fraction yields 1/1000 and 1/1 000 000.

1/1000 is 1000 times larger than 1/1 000 000 therefore, 1:1000 is the larger scale.

Source problems
When using data supplied by someone else to perform mapping and analysis, you should have documentation from the data compiler on the accuracy of the measurements. This will provide an idea of the reliability of the data. This information should include how much rounding was used on the original measurements -- in other words, what precision was used. There should be an indication of how well calibrated the instruments were. If doubt exists, it is safer to assume that the given precision is too high and should be rounded. If data values are rounded, make sure to note this fact.
somewhere on the final map or in the report that accompanies the map so that others will know what assumptions have been made.

Keep in mind that data from separate sources might vary in precision and accuracy. Problems such as the positional accuracy of features may have to be rectified based on the most reliable source being used. Remember, errors are compounded when data is combined from different sources.

Other source related problems include data that are out of date, area coverage that is incomplete, or sparse data that have been over generalized (interpolated). The latter is usually caused by insufficient data sampling which has been extrapolated over large areas and does not reflect the true distribution of the data.

**Multiple Sources: Scale, Projection, and Resolution**

The source map might be at a scale that is too small for the intended application. Be careful about varying scale, map projection, and cell resolution when using multiple source maps.

With small scale maps consideration must be given to non-equal area projections. The Earth is a sphere while maps of the Earth are flat. To represent the surface of a sphere on a flat surface, the surface of the sphere has to be warped. This warping or flattening is called a map projection. Many map projections exist, each with unique characteristics. If you are not familiar with the characteristics of the various map projections, may want to consult the readings suggested in the Welcome document.

The resolution of the map may be too low to resolve the required level of detail. For instance, if each cell in the map represents an area two kilometres by two kilometres, then it will not be possible to resolve individual buildings, farmer’s fields, or even road ways. If the resolution is too high, there may be far more data than is required.

**Multiple Sources: Unique Errors**

Each of the following data sources come with their own unique set of potential errors: topographic maps, hand drawn maps, DEMs, satellite imagery, and aerial photography (be aware of parallax and scale variation). Before using these data become acquainted with some of the possible pitfalls (refer to the reading list in the Welcome document).

There are a number of errors associated with every GIS application. These include digitizing and input errors, reproduction of error from source map, compounded errors when combining multiple layers, and the application of inappropriate map operations (e.g., using Scan when Filter should have been used). Take care at every step to check that errors have not been made.
Make sure the data processing algorithms have been checked for errors of logic (e.g., ensure that operations are specified in the correct order and that they will produce the intended result).

**Data Structure Errors**

Raster based GIS has an inherent set of potential error due to the structure of the data which is stored in discrete compartments. It is assumed that the attribute being mapped falls at the centre of the cell in reality this may not be the case. Tomlin (Tomlin, C.D. (1990) Geographic Information Systems and Cartographic Modeling. Englewood Cliffs, New Jersey: Prentice-Hall.) states that the worst case scenario for positional imprecision between two points of punctual data would be the diagonal width of one cell and the angular discrepancy could be almost 90°. These problems also affect linear features. Measurements along a linear feature can be done either incrementally from centroid to centroid or by Euclidean distance between the start and end points. The measurement that the operation uses will affect the resultant distance and direction measures.