

Visão de Davis, B. E. em “GIS: a visual approach”

Dados Matriciais (*Raster*) e Dados Vetoriais (*Vector*) (*Raster and vector data*)

INTRODUÇÃO

Os dados espaciais, em SIG, têm dois formatos primários (arranjo de dados para armazenamento ou apresentação): *raster/matricial* e *vector/vetorial*. O formato matricial é baseado numa estrutura de grade de células, enquanto o formato vetorial é mais parecido com um mapa de linhas. Há apenas poucos anos atrás, os SIG eram dedicados a uma ou outra estrutura de dados, mas os sistemas atuais integram ambas. Cada formato de dado tem suas vantagens e desvantagens, e os profissionais de SIG reconhecem que em muitos projetos há a necessidade de ambas.

DADOS MATRICIAIS E DADOS VETORIAIS

Nas seções seguintes serão discutidos os dados matricial e vetorial, os dois formatos de dados importantes em SIG. As descrições de cada tipo realçam suas respectivas estruturas de dados e seus usos, assim como suas diferenças e suas similaridades.

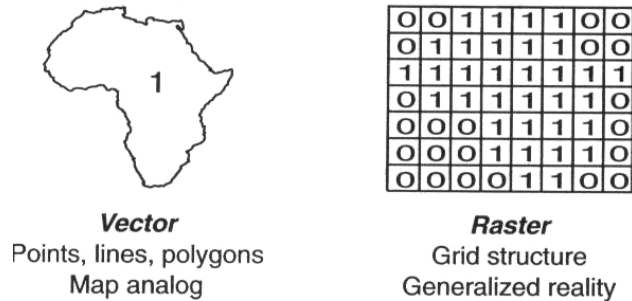
Como amostra, a imagem 1 da figura 3-1 demonstra que o formato vetorial tem polígonos que parecem normais, muito parecidos com um mapa. A África, código 1, é muito facilmente identificada. O formato matricial generaliza a cena para uma grade de células, cada uma com um código para indicar a feição apresentada. A África cuja representação ocorre por meio do código 1 circundada por células 0, é de difícil identificação.

No formato matricial, a cena tem uma estrutura de células de grade. A cada célula da grade uma identidade de feição única é atribuído, normalmente um número (p.ex., montante de chuva, ou um código numérico de uma categoria de uso do solo) ou um rótulo textual (um nome ou um código de letras). A célula é a unidade mínima de mapeamento, o que significa que é o menor tamanho com que qualquer feição da paisagem pode ser representada e mostrada.

Todas as feições na área de uma célula são reduzidas a uma simples identificação de célula. Isso significa que todos os objetos presentes na região coberta por uma célula é acumulada e combinada em uma única identificação; isso é uma generalização da paisagem e de suas feições. Por exemplo, se uma célula cobre uma região de 100x100m, toda informação da terra interna é codificada com um único valor (embora seja possível usar um código que denote duas ou mais feições específicas). Na imagem 2, apresenta-se uma única célula que cobre quatro regiões de uso da terra (códigos 1 a 3), mas pelo fato de que uma célula só pode ter um único valor, o uso da terra dominante, código 2, é atribuído (outras técnicas de codificação são discutidas mais adiante).

Vector and Raster Formats

Image 1: Vector and Raster Data Structure



**Image 2: Raster Cell Coding
Multiple land uses in one cell**

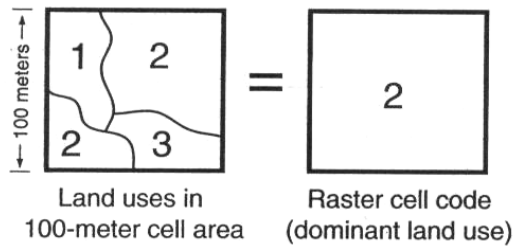


Fig. 3-1: Raster and vector formats.

MODELOS DE DADOS MATRICIAIS E VETORIAIS

Na imagem 1 da Figura 3-2 apresenta-se uma cena do mundo real, com as versões matricial à esquerda e vetorial à direita. As letras são códigos para as feições (R- rio, F- floresta, H- casa, todos sombreados por conveniência). As células que não contém feições recebem o código 0. Tal como na Figura 3-1, as diferenças entre dados matriciais e vetoriais são evidentes.

Modelo Matricial

Pelo fato de que o valor ou código de uma célula representar todas as feições dentro de uma grade, este não mantém corretos o tamanho, a forma ou a localização, para feições individuais. O rio, por exemplo, de fato é mais estreito do que uma célula, mas somente uma célula inteira pode ser codificada como rio, assim, o rio aparece mais largo do que realmente é. Também se deve notar, a mudança na forma do rio, que se torna mais geométrica do que sinuosa (curva) devido a sua representação por células quadradas. Outras feições são generalizadas.

Raster and Vector Data Models

Image 1: Raster and Vector Data Formats

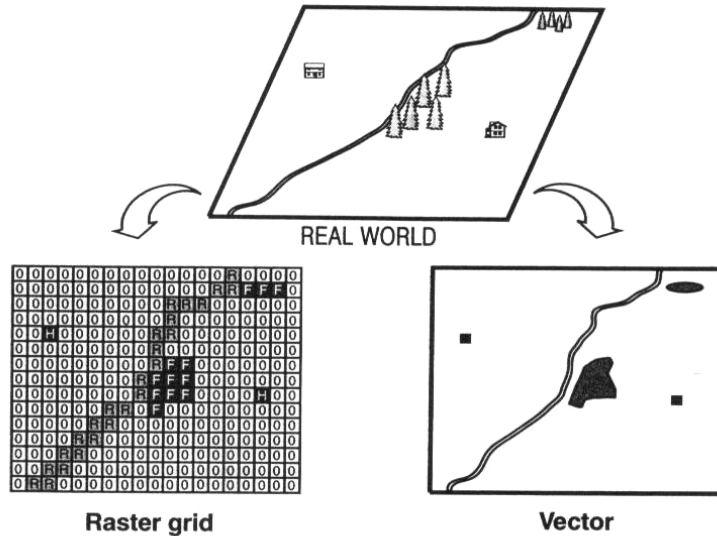


Image 2: Point Features

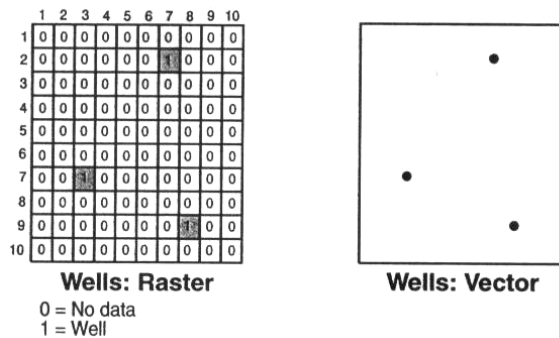


Fig. 3-2: Display effects of raster and vector.

Mesmo nos casos em que “*nada*” exista (não haja dados), a célula deve ser codificada. A maioria dos temas em SIG apresenta apenas as feições que são necessárias em uma região; apresentar tudo o que pertence a uma paisagem poderia ser uma confusão. Por exemplo, há apenas três poços na região da imagem 2, e no formato matricial aparecem três células codificadas como 1. No entanto, o resto da região deve ser codificada como 0, para indicar que não há dados, ou não há poços.

Em geral, um mapa SIG usa vários milhares de células (muitas pra serem vistas individualmente sem ampliação), para tornar as feições reconhecíveis. Para alguns projetos, a generalização espacial não é importante, mas outras demandam por acurácia em forma e em localização.

Modelo Vetorial

Por definição, vetores são elementos de dados que permitem descrever posição e direção. Em SIG, um vetor é a representação gráfica de feições como mapa, sem o efeito de generalização de uma grade matricial. As linhas são analógicas, isto é, não são quebradas em células ou em fragmentos, mas são contínuas do seu início ao seu final. Portanto, a forma representada é mais acurada, como um mapa real. De fato, os dados vetoriais são muito mais acurados do que os dados matriciais.

Na imagem 1 da Figura 3-2, o rio preserva suas curvas e as regiões de floresta tem forma realística. A única limitação é a espessura das linhas usadas para representar a feições. Também as casas não são normalmente mostradas em tamanho correto (exceto para escalas muito grandes), e são consideradas feições pontuais e devem ser representadas por meio de símbolos pontuais. Neste caso um símbolo quadrado.

Os poços, da imagem 2, são denotados por pontos na versão vetorial. Para os locais em que não há dados, nada é introduzido e presume-se que não haja dados ou poços. O modelo vetorial parece ser muito mais simples e mais fácil para a visualização do que o modelo matricial.

DADOS MATRICIAIS

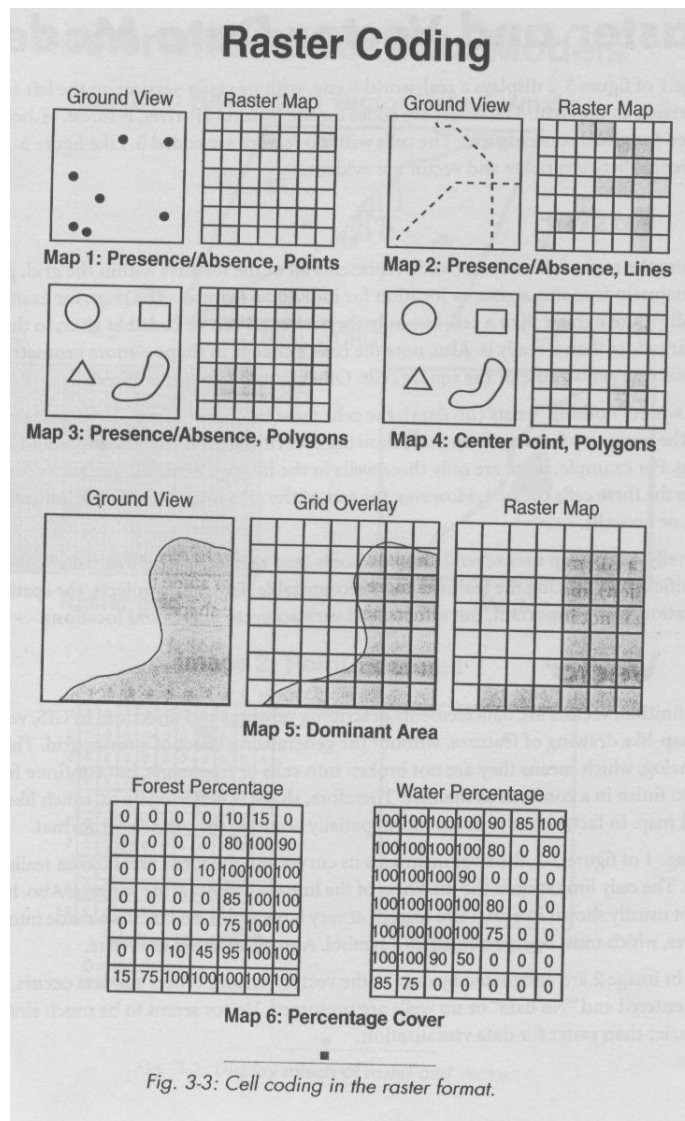
Nas seções que se seguem serão discutidos: codificação matricial; problemas de codificação matricial; mapeamento matricial; resolução; produção de grades e feições lineares; e acurácia e precisão matricial. Estas considerações e conceitos são importantes para o usuário SIG de modo que entenda as diferenças entre os formatos matricial e vetorial.

Codificação Matricial

No processo de entrada de dados, os mapas podem ser tornados digitais ou escanerizados a um tamanho de célula especificado e a cada célula é associado um código ou um valor. O tamanho de célula pode ser ajustado de acordo com uma estrutura de grade (100 x 200 células) ou por unidades de terreno (cada célula = 30m), também denominado *resolução*. A resolução necessária depende das necessidades de projeto, discutidas em material a seguir. Para definir códigos de células existem três esquemas elementares e um esquema avançado. Estes esquemas são discutidos nas seções que se seguem.

Presença/Ausência

O método mais elementar consiste em registrar uma feição se algo dela ocorre no espaço da célula. Este é o único modo prático de codificar feições de ponto e de linha, porque elas não tomam muito espaço de uma célula. Nos mapas de 1 a 3 na Figura 3-3, são apresentados a visão no terreno e então a codificação de um mapa matricial para feições pontuais, lineares e de polígono. Deve-se perceber generalização significativa e as mudanças em forma que se produzem. Por exemplo, linhas diagonais são codificadas em células com forma de escada (mapa 2).



Codificar a presença de uma feição é complicado nos casos em que duas ou mais feições ocorrem no espaço de uma mesma célula, como se pode ver no canto esquerdo do mapa de feições pontuais, em que dois pontos são tão próximos entre si que compartilham a mesma célula. A existência de somente um ponto pode ser mostrada na célula, e não de mais do que um. No terreno há sete pontos, mas o mapa matricial mostra somente seis. Complicações ainda maiores podem surgir nos casos em que as duas feições pontuais são de diferentes tipos. Por exemplo, se uma feição fosse um poço d'água e outra um poço de gás. Poderia ser necessário prever dois temas ou dois níveis de dados diferentes.

Centro da Célula

Para o segundo método de codificação, somente o que ocorre no centro da célula é considerado na codificação. Para o caso de haver múltiplas feições internas ao limite da célula, aquela que estiver no centro da célula “determina” o código. Este esquema não é

bom para pontos ou para linhas, pois as linhas não passam necessariamente exatamente no centro da célula. No mapa 4 apresenta-se a codificação de polígonos. Deve-se notar a diferença significativa entre os métodos *Presença/Ausência* e *Centro da Célula*. Há, obviamente, diferença em área total calculada entre os dois resultados, o que pode ser um problema maior do ponto de vista de acurácia.

Área Dominante

Um método comum consiste em atribuir um código a uma célula pela avaliação da feição que ocupa a maior parte da célula. Esta abordagem é apropriada para polígonos, embora para feições lineares pudessem ser atribuídos códigos de acordo com o maior comprimento linear que ocorre dentro da célula. O conjunto de mapas 5 da Figura 3-3 é uma ampliação de uma porção de uma cena de paisagem que mostra floresta e água. A grade sobreposta mostra como as feições ocupam as células e a visão do mapa mostra o resultado final. Há considerável diferença entre as visões de terreno e de mapa.

Cobertura Percentual

Em cada um dos métodos anteriores descritos resultam erros espaciais. A codificação por área dominante, por exemplo, pode adicionar mais área à feição dominante e reduzir área para aquelas feições não codificadas. Embora o erro em cada célula possa ser muito pequeno, o acúmulo deste erro pode ser significativo em acurácia. Um método mais avançado consiste em separar cada feição para a codificação em diferentes temas individuais, e então atribuir valores percentuais de cobertura em cada célula. No mapa 6 da Figura 3-3 se apresenta a floresta e a região de água do mapa 5, com as suas percentagens de ocorrência em cada célula. Uma soma de todas as células produz a área de cobertura de cada feição.

Os valores de cobertura percentual são associados às células. Esse modo é mais trabalhoso (demanda por mais tempo de processamento), mas proporciona melhor acurácia espacial. Este método é especialmente útil em células de transição que contém diversas feições. Entretanto, para um tema que tenha 15 atributos (classes de uso da terra, por exemplo) devem ser produzidos 15 novos temas separados.

Nenhum destes métodos é perfeito e em cada um deles podem ocorrer erros de acurácia. Normalmente umas poucas células são codificadas de modo incorreto ou “injusto” devido às suas próprias posições relativamente à grade sobreposta. Entretanto, estas são ilustrações exageradas de grandes cenas que normalmente compreendem vários milhares de células; a ocorrência de células equivocadas normalmente não representa um problema significativo.

Problemas de Codificação Matricial

A codificação matricial produz erro de acurácia espacial, embora a perda possa ser aceitável, como já se discutiu. Na Figura 3-4 se ilustram alguns dos possíveis problemas de conversão de uma paisagem em um formato de dados matricial. Os polígonos de uso da terra são codificados em nove células. Embora esta ilustração seja um exagero dos problemas espaciais, estes erros pequenos podem ser compostos em muitas células. O método de

cobertura dominante poderia ser aplicado para definir os códigos de uso da terra neste exemplo.

O mapa apresenta as categorias de uso do solo de A a E (divididas por linhas tracejadas e denotadas por letras inseridas em círculos). A célula 1 é claramente parte da classe A, mas a célula 2 está dividida entre as classes A, B e E. Em razão de uma célula ter apenas um código, a decisão de atribuir pelo método da classe dominante, faz com que a célula 2 seja atribuída à classe A, apesar de que A constitui menos do que 50% da célula. Esta decisão implica em considerável perda para as categorias B e E.

A célula 3 parece ter iguais proporções das classes B, C e E, o que a torna difícil de classificar. Terços exatos são provavelmente raros de ocorrer, mas C é uma fração maior. A célula 4 é quase totalmente A, de modo que a decisão é fácil (mas resulta em alguma perda de F). A célula 5 é predominantemente E, mas a pequena área de D será perdida. Este é outro problema especial: a classe D claramente existe na região e pode ser importante, mas o processo de escaneamento pode eliminá-la como categoria de dados. Por outro lado, se isso é o resultado de uma ampliação, D poderia ser muito pequeno para ser importante.

A célula 6 é certamente E. A célula 7 é classe A pois tem 55% de cobertura, mas isso significa que os 45% restantes de uso F serão perdidos. A célula 8 é classificada como F. Talvez parte de código E que é convertido para F na célula 8 ajude a compensar a perda de F nas células 4 e 7. A célula 9 é 50% E e 50% F, e então recebe a classe E, o que resulta em mais erro para F.

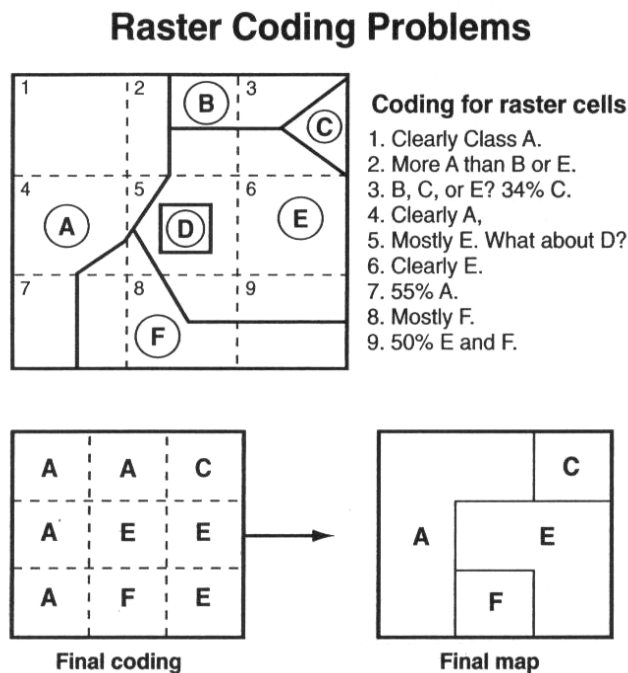


Fig. 3-4: Inaccuracies related to raster coding.

Os mapas codificados são mostrados na parte de baixo. Estes mapas parecem completamente diferentes do uso da terra de entrada. Deve-se notar que os usos da terra B e D foram eliminados, em grande parte por haver poucas posições em células; um leve

deslocamento da grade poderia produzir diferentes resultados. A perda de alguns usos da terra pode implicar em problemas para o projeto, tais como deturpar/prejudicar estatísticas. O uso do método do centro da célula, para o caso da célula 5 poderia determinar o código D, desse modo eliminaria os vizinhos E e A.

Os sistemas matriciais generalizam a paisagem e também conduzem a erros de acurácia espacial e de classificação. Isso poderia não ser importante para alguns propósitos, mas poderia ser essencial para outros. Uma solução possível seria incrementar a resolução pelo incremento do número de células, de modo a torná-las menores e assim, mais sensíveis à classificações acuradas.

Mapeamento Matricial

O maior problema relacionado às estruturas matriciais é o fato de que a forma das feições é forçada a ter uma forma artificial de grade de células. Para feições que naturalmente têm ângulos retos, tais como culturas de forma retangular ou para distritos retangulares, isso pode não representar maior problema. Entretanto, para o caso de muitas feições, tamanho e forma podem se tornar distorcidos de modo indesejável.

Na Figura 3-5 são mostrados dois conjuntos de dados cartográficos: polígonos e feições lineares. O conjunto de polígonos apresenta regiões codificadas de 1 a 4 (p.ex. quatro diferentes usos da terra). Os limites dos polígonos 1, 2 e 3 são não regulares ou com aparência retilínea artificial. A segunda figura, da primeira coluna, representa as células codificadas, que força linhas naturais a ter um comportamento regular de linhas retas que não é natural. Por isso, as regiões de usos da terra de 1 a 3 foram generalizadas para figuras geométricas de ângulos retos. Isso resulta em erros de acurácia espacial em termos de tamanho e localização.

As feições de uso de terra de código 4, têm forma regular retangular na imagem original, de modo que suas representações matriciais não são afetadas de significativamente em forma. Eventualmente as posições das células podem combinar com o mundo real e não haver perda de acurácia.

Os mapas sombreados (figuras de abaixo) mostram como os códigos sombreados ou coloridos são usados para tornar a visualização mais agradável (são chamados de *mapas coropléticos*). Certamente um mapa em que são mostrados apenas números em cada posição de célula não é legível, ou é pouco atraente, de modo que mudar os códigos para elementos gráficos que sejam agradáveis aos olhos pelo uso de cor, sombreado e símbolos resultará numa visualização mais apropriada.

As feições lineares podem ser bastante afetadas pela escanerização. Duas mudanças importantes são feitas nos mapas matriciais sombreados: (1) a forma de figuras diagonais muda e (2) a espessura de feições lineares aumenta devido ao tamanho mínimo das células de mapeamento. As árvores de estradas da ilustração de feições lineares são diagonais. As diagonais são transformadas em degraus de escada pelo efeito de ligação de uma sucessão de células. A espessura de todas as feições lineares são generalizadas para o tamanho da célula pois não existe unidade de mapeamento menor do que a célula (não esquecer da “unidade mínima de mapeamento”).

Como foi ressaltado, a maioria dos conjuntos de dados matriciais é composta por milhares de células. Cada célula é muito pequena para ser vista sem ampliação, de modo que as feições podem parecer realísticas par olhos desarmados, tal como em fotografias de jornal

feita de *pixels* muito pequenos. A perda de acurácia espacial, frequentemente, não é significativa naquela escala. Para cada projeto deve-se decidir qual é a perda aceitável.

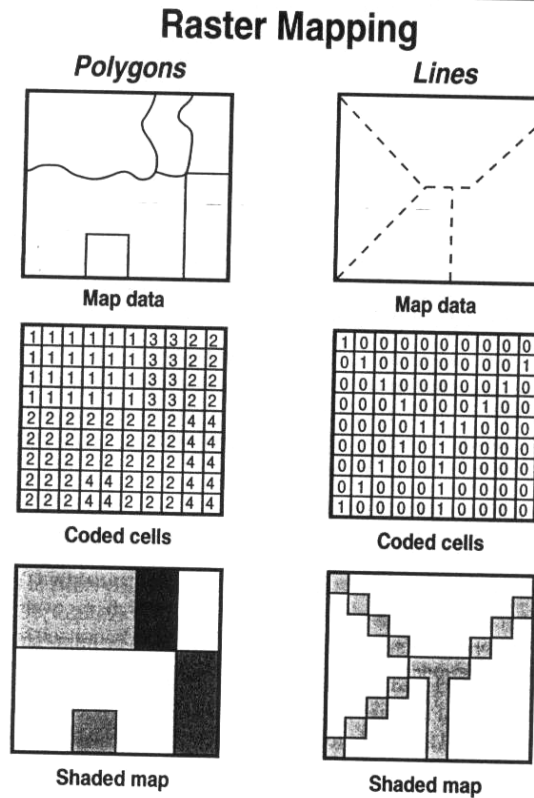


Fig. 3-5: Gridded nature of raster mapping.

Resolução

O incremento do número de células em um conjunto de dados incrementa por sua vez a resolução espacial, o que auxilia no incremento da acurácia espacial. Isso normalmente é feito como parte do estágio início da entrada de dados de cada mapa. O número de células (100 x 200) ou o tamanho de cada célula (30m) pode ser definido. Na Figura 3-6 são mostradas grades de baixa densidade/ baixa resolução à esquerda e de alta densidade/alta resolução à direita. O mapa de eco-sistema que aparece entre as grades, no meio da figura, mostra a versão vetorial de entrada de dados.

A grade 8 x 7 à esquerda é mais esparsa, com somente 56 células. Ao se utilizá-la para representar o continente africano, o resultado uma forma quase que irreconhecível (esquerda embaixo). Uma vantagem de usar relativamente poucas células é se ter um tempo de processamento baixo e facilidade de análise. A forma correta pode não ter importância num projeto em que a ênfase está na possibilidade de realizar análises rapidamente, e uma visão geral seja suficiente. A grade muito mais densa, à direita, com 990 células, torna a forma reconhecível, o que poderia ser desejável para mapeamento e análise. Esta grade demanda por mais armazenamento e tem um tempo de processamento maior. A tecnologia

moderna, entretanto, tem tornado menos importantes as questões relativas ao armazenamento e ao tempo de processamento.

Sugere-se comparar a resolução de célula de cada grade (1.210.000 versus 90.000 quilômetros quadrados por célula). A depender da aplicação, a célula de 300 km pode ser muito detalhada, muito geral, ou justamente o contrário. Muitos arquivos matriciais em SIG contém muito mais células e definição espacial melhor dos que estes exemplos.

Raster Resolution

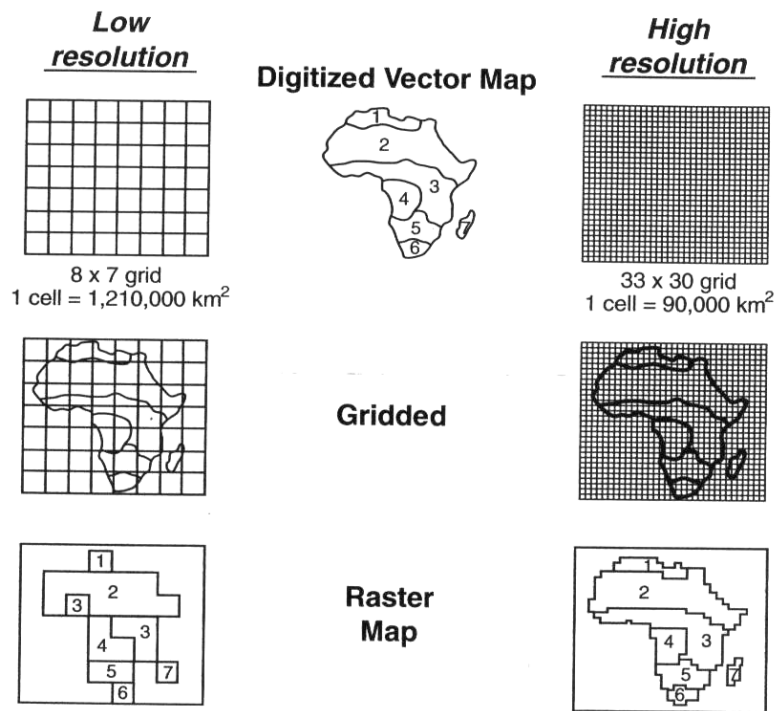


Fig. 3-6: Resolution and the raster format.

Geração de Grades e Feições Lineares

Para perceber como o formato matricial pode criar erros de acurácia (*inacuracies*), deve-se notar a forma real do rio à esquerda da Figura 3-7 e então compará-la às duas versões matriciais (mapa 1 e mapa 2, que foram produzidos pela aplicação do método presença/ausência). Cada célula tocada pelo rio foi codificada (sombreada) como “rio”.

No mapa 1 usou-se uma grade de baixa resolução 8 x 12 células, o que resultou numa representação com uma forma mais generalizada e crua. No mapa 2 usou-se o dobro de células (16 x 24), e a nova forma parece mais realística, embora isso esteja ainda muito longe da acurácia em forma e posição proporcionada pelo formato vetorial.

Para efeito de medidas espaciais, suponha-se que o lado de cada célula no mapa represente 1 quilometro. O tamanho real do rio é 24 km, mas há 27 células rio na nova forma, o que resulta num erro de 3 km se o número de células for usado para estimar o seu

comprimento. O contato com uma célula neste caso é uma questão de sorte; um pequeno deslocamento inicial da grade, neste caso, poderia resultar num número maior ou menor de células. Entretanto, deslocar a grade não é necessariamente uma opção, pois esta grade tem início no canto superior esquerdo de uma área escaneada de um mapa original, e que é difícil de ajustar uma vez que já foi definido.

As células da grade de alta resolução no mapa 2 tem comprimento de lado de 1/2 km, e as células rio perfazem um total de 52, o que conduz a um comprimento de 26 km - uma medida melhor, mas não perfeita. Maior densidade de células numa grade normalmente implica em medidas mais acuradas (mas não tão boas como no caso vetorial). Portanto o tamanho da célula (resolução) é importante. Entretanto, o propósito de um projeto normalmente determina os requisitos de acurácia e de resolução.

Raster Gridding Linear Features

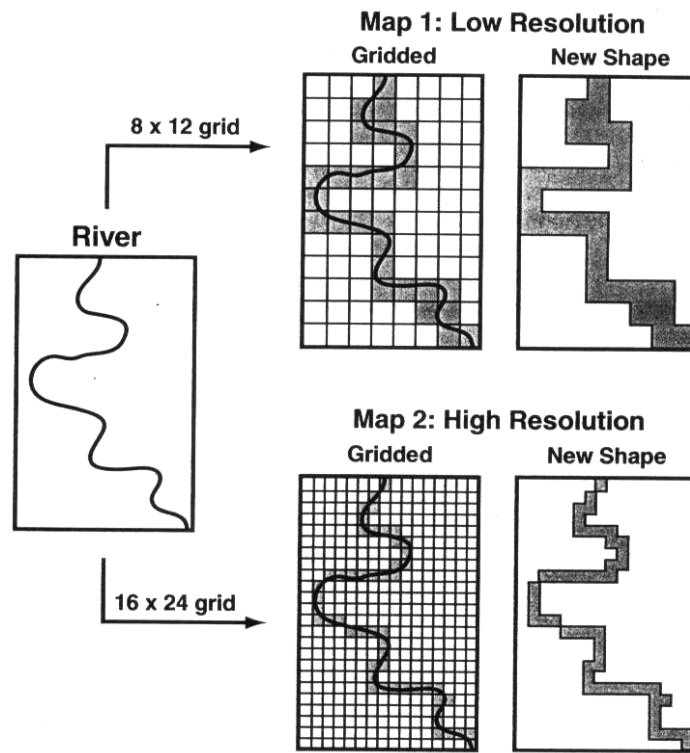


Fig. 3-7: Example of spatial inaccuracy due to the raster format and resolution.

Precisão e Acurácia Matricial

Os aspectos de precisão de dados matriciais (a posição exata) e de acurácia (verdade espacial) frequentemente são um problema. Pelo fato de uma célula da matriz se constituir na maior resolução e a menor unidade de mapeamento, não há meio de saber exatamente em que lugar da célula a feição ocorre.

A imagem 1 da Figura 3-8 mostra que uma feição X pode estar localizada em qualquer lugar dentro do limite da célula, mas a localização do ponto de vista do formato matricial é simplesmente dentro de toda a célula (em outras palavras, em qualquer lugar dentro da célula). Isso pode não ser muito importante para algumas aplicações, mas a incerteza posicional pode ser significativa em outras. Certamente, células menores têm erro espacial menor, pois a área de dúvida é menor.

A incerteza se torna maior nos casos em que se realiza medidas através de células (imagem 2). A distância verdadeira no mundo real entre duas feições que estão em células adjacentes pode variar consideravelmente, tanto dentro das células (dentro dos limites), até fora dos limites das células, o mesmo em cantos opostos da célula. Apesar disso, a distância matricial é duas células pois a célula é a mínima unidade de medida.

As medidas de área (imagem 3) também são generalizadas. Uma medida de demonstração de uma unidade de célula é mostrada no centro. O triângulo padrão mede 30 unidades quadradas, mas demanda por 40 células neste formato matricial de baixa resolução. O círculo tem 28,26 unidades de área, mas usa 32 células matriciais para representá-lo. Certamente, o montante de erro pode ser reduzido por meio da adoção de uma resolução maior, mas é verdade que o formato matricial pode criar problemas em alguns projetos.

Raster Accuracy and Precision

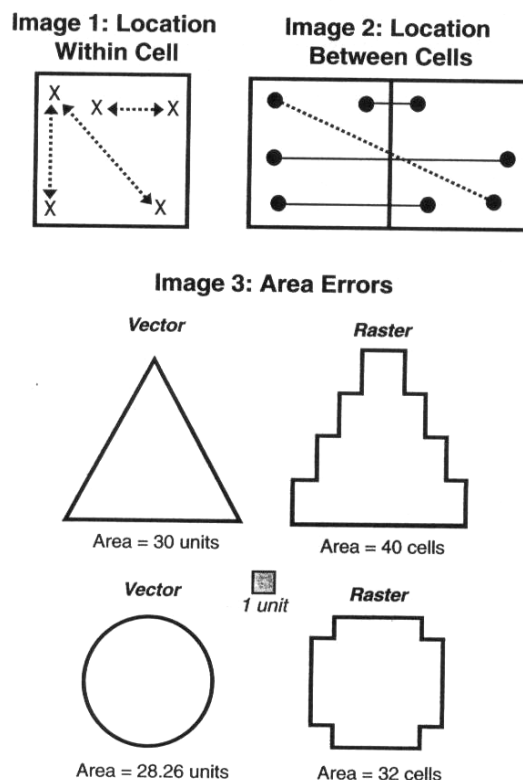


Fig. 3-8: Raster data precision and accuracy are not the same thing.

DADOS VETORIAIS

Feições representadas de acordo com o modelo de dados vetorial parecem ser mais realísticas do que com o modelo matricial, e têm acurácia espacial superior. A imagem 1 da Figura 3-9 mostra os três tipos de feições vetoriais: ponto, linha e polígono. Como esperado, o ponto é representado com um símbolo (um X neste caso), e a linha e o polígono aparecem como representações gráficas simples, sem as complicações do formato matricial.

A imagem 2 mostra que os dados vetoriais são estruturados em vários elementos, que não aparecem na imagem 1. As feições vetoriais são definidas primariamente pela sua forma, mais especificamente pela linha de contorno de sua forma (preferivelmente a um grupo de células do formato matricial). Em SIG, o sistema vetorial é uma *estrutura de dados baseada em coordenadas*, isso significa que cada ponto é localizado por suas coordenadas (X, Y); por exemplo, a localização da feição ponto por meio da linha 2 e coluna 2, na imagem 2.

Pontos de forma são os extremos e as dobras que definem o contorno da feição. No início e no final de cada feição de linha ou polígono há um *nó*. A cada dobra (mudança de direção) há um *vértice*. Os nós são pontos extremos e os vértices são pontos intermediários, para definir a forma. Feições pontuais são nós isolados. Uma forma é registrada por meio das coordenadas de seus pontos de forma. Observar os números linha-coluna próximo a cada nó e cada vértice na imagem 2.

As *cadeias* interligam pontos de forma para representar o contorno da feição. As cadeias são vetores (daí o termo *sistema vetorial*), ou caminhos de estruturas de dados que não são parte dos elementos de dados armazenados; as cadeias não são linhas reais, mesmo que apareçam no monitor, mas definem e apresentam a conexão entre pontos de forma (também são chamadas arcos, limites ou ligações). Os arquivos de dados de sistema vetorial armazenam somente as coordenadas de cada nó e de cada vértice; o hardware representa graficamente as cadeias de conexão. Em um monitor de visualização, somente as cadeias são vistas, como a definição da feição.

Os nós e os vértices são os elementos que são realmente armazenados, ao passo que as cadeias são um componente virtual (existem em efeito mas não em forma real). A apresentação visual é produzida por conveniência ou necessidade humana mas não é necessária para o computador. Este é um formato de armazenamento de dados eficiente. Os nós e os vértices podem ser inspecionados a partir de condições de visualização especiais para edição. A estrutura de dados vetoriais também são conhecidas como *modelo arco-nó* por usar cadeias (arcos) e pontos extremos (nós).

Pelo fato de não haver cadeias presas às feições pontuais, a estrutura de dados o percebe/considera como um nó único, tal como na linha 2 coluna 2, da imagem 2. Uma feição linear reta é uma cadeia única (dois nós conectados por uma cadeia), enquanto que uma linha complexa em diversas cadeias que determinam segmentos individuais. A feição de linha ilustrada inicia-se no nó 1 e conecta-se ao vértice 1 e este a vértice 2, e termina no nó 2.

Os polígonos consistem em três ou mais cadeias de envolvem uma região. deve-se notar que o nó de polígono ilustrado é tanto nó de início como de final (comum é polígonos únicos). Polígonos conectados normalmente compartilham diversos nós. Avalie a imagem 2 para entender o posicionamento e a função dos diversos elementos do sistema vetorial e como estes elementos definem pontos, linhas e polígonos.

Vector Data

Image 1: GIS Features

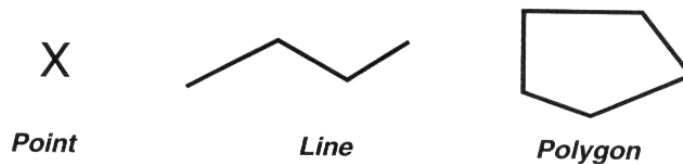


Image 2: Vector Data Structure

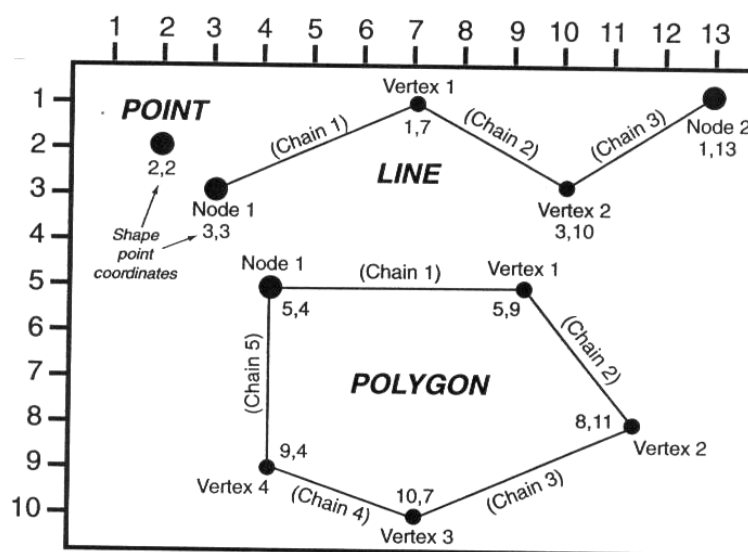


Fig. 3-9: The components and structure of vector data.

ESTRUTURAS MATRICIAIS E VETORIAIS

As estruturas matriciais e as estruturas vetoriais têm diferentes métodos de armazenamento e de visualização de dados espaciais. Na Figura 3-10 estão resumidas as feições matriciais e vetoriais. As células matriciais são armazenadas e apresentadas como células, mas no formato vetorial somente os nós e os vértices são armazenados (as cadeias são elementos virtuais apresentados por conveniência). Isso resulta em considerável diferença em armazenamento.

Um ponto num sistema matricial é uma única célula, mas no sistema vetorial um ponto é apenas um nó representado por um símbolo com suas coordenadas de posição associadas. Uma linha simples no sistema matricial consiste em uma sequência de células, que erroneamente dão a impressão de sua largura. Num sistema vetorial uma linha simples consiste em dois nós (um em cada extremo) e uma cadeia que os conecta. Uma linha matricial mais complexa consiste em células conectadas, algumas vezes semelhantes a

degraus se diagonais. As linhas complexas no formato vetorial têm vértices para marcar a variação de direção e nós em cada extremo. (Na figura os pontos maiores são os nós e os vértices são os pontos menores).

Raster and Vector Data Structure

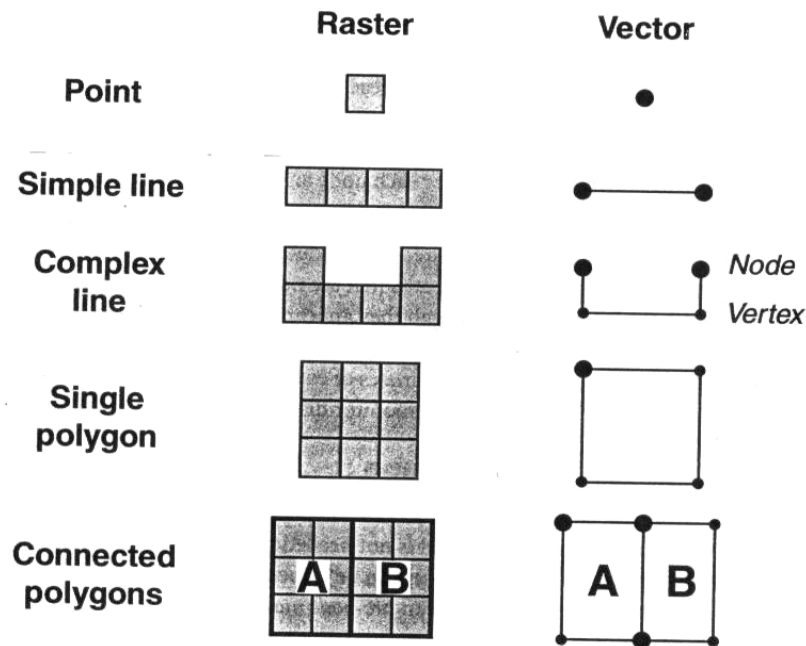


Fig. 3-10: The difference between raster and vector data structure.

Os polígonos matriciais são preenchidos com células. Para polígonos simples, o formato vetorial normalmente tem um nó único e vários vértices para marcar as mudanças de direção do limite. Os polígonos conectados são simplesmente blocos de células, lado a lado no formato matricial, mas no formato vetorial compartilham limites comuns e nós comuns. Estão ilustrados os polígonos adjacentes A e B, duas séries de células conectadas no formato matricial, mas uma cadeia os separando no sistema vetorial (com nós nos seus extremos).

Por razões de programação, normalmente somente um tipo de dado está num tema de um SIG (um arquivo ponto, linha ou polígono). Uma das razões para isso é evitar confusões entre feições que podem se parecer mas são diferentes, tais como quatro ruas que se conecta (feições de linha) que poderiam ser interpretadas como um polígono quadrado. Para a apresentação e para mapeamento, a visualização pode conter uma mistura de tipos de feições (i.e., ponto, linha e polígono), mas estas feições são oriundas de arquivos de dados diferentes.

Embora os modernos SIG possam administrar tanto dados matriciais quanto vetoriais, há motivos para realizar conversões entre os formatos. Há problemas especiais a considerar, e nesta seção se discute o processo, com ênfase na mudança matricial-vetorial.

Na Figura 3-11 está presente a evolução de um rio de sua forma vetorial original para uma versão matricial, e então a sua conversão para uma nova versão vetorial. O processo de mudança de feições matriciais para feições vetoriais normalmente envolve a realizar a conexão entre os centros das células. Pelo fato de haver incerteza de onde está de fato o rio, o centro das células é a localização lógica para um vértice vetorial.

No mapa 1 aparece o rio original tirado da Figura 3-7. Ao ser imageado por sensoriamento remoto para uma estrutura matricial, o rio é generalizado e sua forma é tornada artificialmente quadrada (mapa 2). (Esta visualização é uma ampliação exagerada e o arquivo de dados matriciais reais mostra o rio com uma forma razoável.) Para o caso de haver necessidade de uma versão vetorial, os centros das células são unidos numa tentativa de construir um melhor caminho (mapa 3). O resultado é mais um rio com contornos agudos, sem as curvas esperadas (mapa 4), o que cria dúvidas acerca da forma verdadeira do rio.

Raster to Vector Conversion

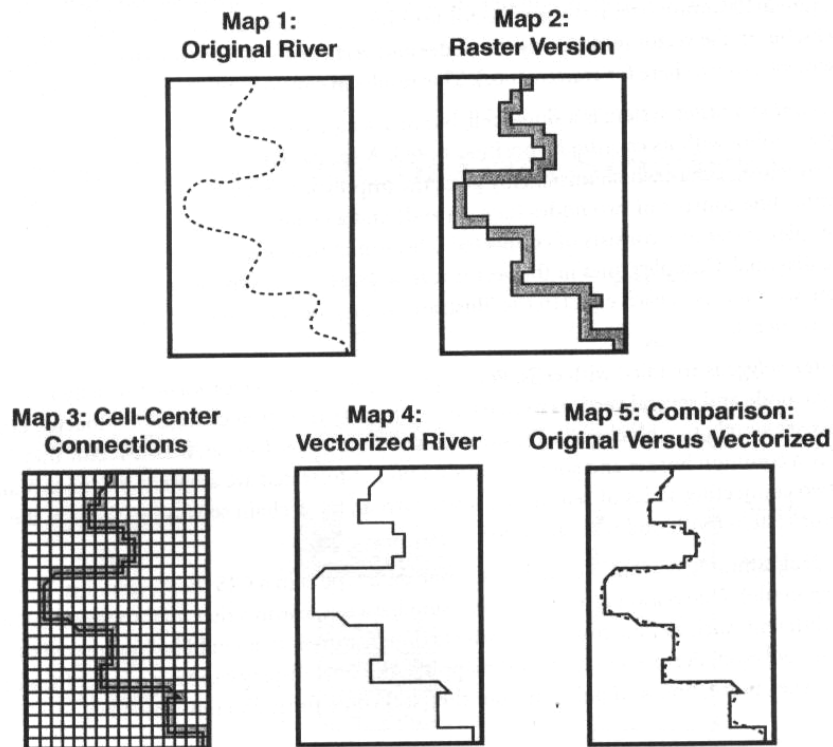


Fig. 3-11: Conversion of data from raster to vector format.

Esta incerteza causa problemas de acurácia, O produto final pode parecer ser bastante acurado pelo fato de ser vetorial (embora não nesta visualização ampliada), mas o

novo rio não é mais do que cadeias conectadas com vértices estimados espacialmente. No mapa 5 há uma comparação entre a forma do rio original e sua nova versão vetorizada. A diferença não é grande neste caso, mas pode ser significativa para os casos em que alta acurácia for demandada.

Alguns SIG podem fazer com a nova versão vetorial pareça mais suave por meio de suavização de vértices afiados ou algumas vezes pela criação de curvas artificiais entre os vértices (chamadas *splines*), mas novamente, o usuário pode ser induzido a acreditar que a boa aparência é mais acurada do ponto de vista espacial do que a versão matricial, em casos em que não é.

Há pelo menos quatro razões para converter dados de formato matricial para o formato vetorial:

- 1) apesar dos problemas de acuracidade, a aparência visual do formato vetorial é preferida;
- 2) alguns plotadores operam apenas com dados em formato vetorial (é mais fácil produzir linhas do que quadrados);
- 3) a comparação com dados vetoriais é melhor no caso em que ambos os arquivos de dados têm o mesmo formato; e
- 4) embora a maioria dos modernos SIG aceitem ambos os formatos de dados, alguns (especialmente os mais antigos) tem tanto o modelo matricial quanto o modelo vetorial como a estrutura de dados operacional central, e os formatos de dados deveriam ser idênticos ara facilitar o gerenciamento e as operações analíticas. Este requisito está se tornando desatualizado e pode não ser mais uma razão válida.

A produção de dados matriciais a partir de dados vetoriais frequentemente á chamada *gridding* devido à conversão de vetores suaves para uma estrutura de grade. O tamanho das células matriciais pode ser estabelecido pelo usuário, como o tamanho no terreno (grade de 100 m), pelo número de linhas e de colunas, ou de modo a combinar com outro tema matricial. Este processo está mostrado com os mapas 1 e 2, que são da Figura 3-7, usada na discussão de geração de grades matriciais.

VANTAGENS E DESVANTAGENS DE DADOS MATRICIAIS E DADOS VETORIAIS

Embora dados normalmente estejam em formato vetorial ou formato matricial, pode ser difícil para um usuário iniciante saber que tipo demandar ou em que situações realizar sua conversão. Nas seções seguintes são discutidas as vantagens e as desvantagens, associadas a ambos os formatos.

Formato Matricial - Vantagens

Apesar dos problemas relacionados à acurácia, os dados no formato matricial têm vantagens (mostradas na figura 3-12 e discutidas a seguir):

1- A estrutura d dados matricial é relativamente simples: uma grade com um único código em cada célula (imagem 1).

É uma estrutura fácil de entender e usar, mesmo por iniciantes.

2 - A estrutura simples de uma grade torna as análises mais fáceis (imagem 2).

Os computadores realizam a comparação de dados de modo rápido, e se há uma pilha de arquivos de dados para ser manipulada para análise, o computador lê cada posição da célula na grade uma a uma, e realiza a análise para aquela célula para cada arquivo de dados, com por exemplo, encontrar o maior valor. Por exemplo, a figura apresenta três temas (A, B e C), com a primeira célula de cada grade realçada (posição da célula 1). O programa compara os valores de todas as células 1 e insere o maior valor na célula 1 de uma nova grade, que tem um novo tema. Este tipo de operação pode ocorrer rapidamente num grande arquivo de dados de grade.

3 - Pelo fato das estruturas matriciais serem relativamente simples, a plataforma computacional pode ser “low-tech” (não deve necessariamente ser sofisticada, imagem 3).

Máquinas mais velhas, mais lentas e limitadas podem facilmente administrar dados matriciais, e o nível de tecnologia e custo pode ser baixo. Entretanto, mesmo equipamentos atuais razoáveis são capazes de operar dados vetoriais, e esta vantagem está se tornando menos relevante.

Raster Advantages

Image 1: Simple Data Structure

3	1	1	1	0	0
0	3	2	3	3	1

Image 3: Low-Tech Platforms



Image 4: Remote Sensing Data

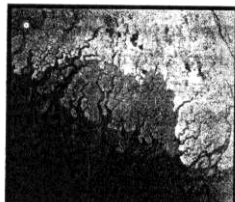


Image 2: Easy Analysis

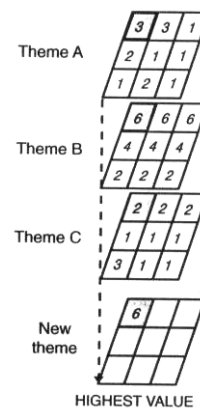


Image 5: Modeling

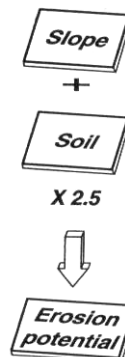


Fig. 3-12: The advantages of the raster format.

4 - As imagens de sensoriamento remoto (obtidas a partir de aeronaves ou de satélites) são naturalmente matriciais (imagem 4).

Isso permite integração e comparação entre imagens e dados SIG. A imagem é facilmente integrada a um SIG de formato dados matricial, pela igualdade dos formatos de dados (aspecto discutido adiante no Capítulo 5).

5 - A modelagem, neste contexto, consiste num conjunto de arquivos de dados gerais ou um conjunto de procedimentos universais voltados à realização de certas tarefas em SIG (imagem 5).

Se, a erosão de solo predita em uma região está expressa como uma fórmula (um modelo de erosão) que se utiliza a declividade e valores de solos, os números em um sistema de grade são mais fáceis de analisar do que o formato vetorial. Na figura mostra-se um exemplo simples de adição do valor de declividade ao valor de solo e então multiplicar o resultado da soma por 2,5 para obter um número que indica o potencial de erosão. O processo de cálculo é aplicado célula a célula progressivamente, como se mostra na imagem 2 da Figura 3-12. A tarefa é realizada para cada célula do conjunto de dados, o potencial de erosão de toda a região pode ser mapeado e analisado de modo fácil e rápida. Modelagem é discutida no Capítulo 10.

Formato Matricial - Desvantagens

As desvantagens dos dados no formato matricial são apresentadas, e ilustradas com apoio na Figura 3-13:

1 - A falta de acurácia é comum em dados matriciais (imagem 1).

Normalmente espera-se que perdas sejam compensadas por ganhos, e de modo geral, que a influência da falta de acurácia seja de algum modo anulada neste processo. Este pode ser um pensamento ou uma expectativa, no entanto, e em projetos nos quais haja demanda por alta acurácia deve-se usar mais células (maior resolução) ou usar o formato vetorial nos casos em que for possível.

2 - Pelo fato de que, em cada célula há uma representação generalizada da paisagem, os resultados de sua aplicação podem ser de baixa qualidade se comparados aos dados vetoriais (imagem 2).

Na imagem 2 se mostra o mapa original e uma comparação entre alta resolução e baixa resolução para o continente africano. Mesmo o uso de um número muito grande de células pode garantir apenas melhor resolução, e não necessariamente alcançar uma acurácia satisfatória. Mais células significa arquivos maiores, computação e apresentação de dados mais lenta.

3 - Pelo fato de haver generalização em cada célula, pode-se não saber ao certo o que ocorre num dado local (imagem 3).

Conseqüentemente, as células da matriz (sugerem) implicam em verdades – numa estrutura implícita (compreendido embora não expresso diretamente). Isso significa que aquilo que é mostrado não é necessariamente aquilo que existe na paisagem. O usuário deve julgar se o nível de generalização é, ou não, aceitável.

Raster Disadvantages

Image 1: Spatial Inaccuracy

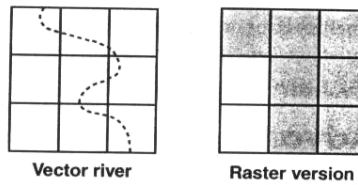


Image 2: Generalization

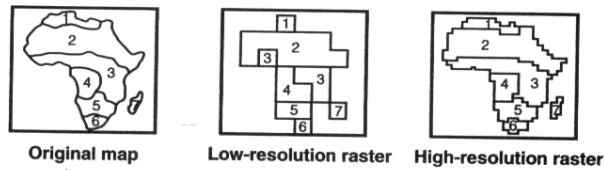


Image 3: Implicit Data

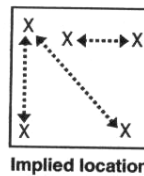


Image 4: Large Data Sets

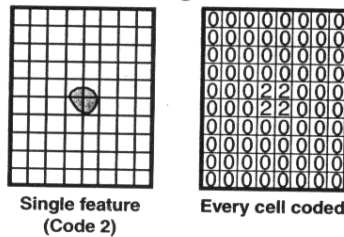


Fig. 3-13: Disadvantages of the raster format.

4 - Para cada célula deve haver um código, mesmo que não haja nada em seu interior (imagem 4).

Toda célula é codificada, o que faz com que seja necessário armazenar grandes quantidades de dados, especialmente para altas resoluções. Por isso, os sistemas matriciais podem conter conjuntos de dados muito grandes, mesmo que eventualmente, estes sejam de pouca utilidade.

Formato vetorial - Vantagens

As vantagens do formato vetorial são apresentadas, e ilustradas com apoio na Figura 3-14:

1 - De um modo geral os dados vetoriais são mais parecidos com mapas do que os dados no formato matricial (imagem 1). Os dados geográficos são mais acurados e mais confiáveis do

que os dados em formato matricial. As visualizações vetoriais também são mais agradáveis aos olhos.

2 - Os dados vetoriais têm alta resolução.

A realização de ampliação não produz danos à qualidade de apresentação. De fato, o detalhe e a acurácia podem ser maiores do que a resolução do dispositivo de apresentação; as limitações podem ser devidas mais a restrições de *hardware* do que dos dados. Na imagem 1 está apresentada uma visualização ampliada do Gabão (G) sem generalização dos dados espaciais.

Vector Advantages

Image 1: High-resolution Map Analog
Scale-independent Resolution

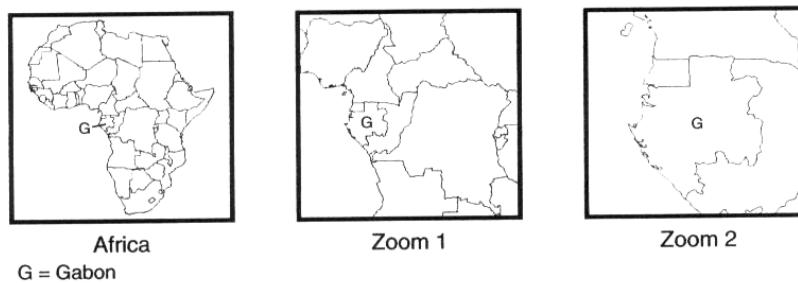


Image 2: Node-Vertex Storage

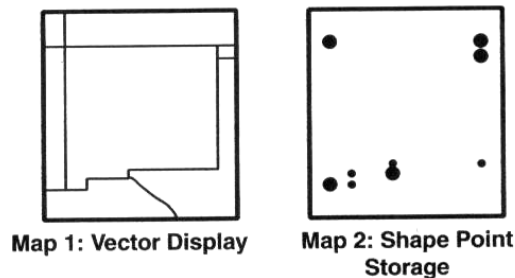


Image 3: Understandable

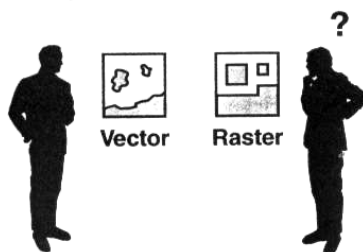


Image 4: Topology

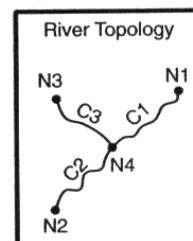


Fig. 3-14: Advantages of the vector format.

3 - A alta resolução suporta a alta acurácia.

Pelo fato dos sistemas vetoriais utilizarem-se e dados estruturados em coordenadas e de cadeias, a apresentação pode ser ampliada aos limites dos padrões de coleta de dados original. Se a coleta de dados em campo ocorreu a mais ou menos 10 m de acurácia posicional, a apresentação da tela pode ser ampliada até que este limite seja alcançado. Os limites ou contornos das feições não se tornarão mais espessos ou perderão sua acurácia posicional.

4 - Os dados em formato vetorial tem vantagens em termos de armazenamento (imagem 2).

Os sistemas e os administradores de dados estão sempre relacionados ao armazenamento de dados. Os formatos de dados vetoriais demandam por menos armazenamento e normalmente oferecem maiores potencialidades do que os administradores de formatos matriciais. Isso ocorre porque as feições são definidas e armazenadas apenas como nós e vértices, enquanto que para os conjuntos de dados matriciais cada célula é codificada. Isso significa que arquivos de dados vetoriais podem ser menores e mais rápidos do que arquivos de formato matricial. Para dados vetoriais, apenas os elementos de dados essenciais são armazenados.

5 - Em geral as pessoas entendem mais facilmente o que se representa, nos mapas, de modo vetorial (imagem 3).

As apresentações vetoriais podem ser mais aceitáveis do que apresentações generalizadas de dados matriciais para um não especialista. As estruturas de dados vetoriais parecem ser o sistema preferido por muitos usuários de SIG, especialmente para o caso dos projetos em que se queira controle sobre a acurácia dos dados e sobre a apresentação de dados. O público em geral, facilmente reconhece as apresentações vetoriais. Entretanto, atualmente a qualidade de apresentação é uma questão menor, dado o poder dos modernos sistemas para apresentar dados matriciais e vetoriais como visualizações agradáveis e compreensíveis.

6 - Os dados em formato vetorial podem ter topologia (imagem 4).

Como se apresentou anteriormente, no próximo capítulo, a topologia é uma estrutura de dados especial que tem vantagens também especiais, em SIG. Os dados matriciais não podem ser organizados ou se beneficiar da topologia, e por isso muitos usuários consideram os modelos de dados vetoriais como sua escolha adequada.

Formato vetorial - Desvantagens

Embora o formato de dados vetorial seja o melhor para muitos usuários e aplicações, este formato tem algumas desvantagens.

1 - Os formatos de dados vetoriais podem ser mais difíceis de manipular do que os dados matriciais.

Os dados vetoriais normalmente são armazenados como longas listas de coordenadas de nós e de vértices – de fácil representação computacional, mas de difícil edição por parte do usuário. Conhecer o modo de ler e de operar um arquivo de dados pode demandar tempo, e não ser trivial. Entretanto, a maioria dos usuários pode não estar interessada nestas

questões. A imagem 1 da Figura 3-15 apresenta uma idéia de dados de topologia tratados no próximo capítulo ao mostrar quatro tabelas que normalmente compõem sistemas topológicos vetoriais. Estes sistemas apresentam certo grau de dificuldade em sua administração.

Vector Disadvantages

Image 1: Complex Data Structure

Chain Topology			Node Topology	
Chain	Start Node	End Node	Node	Chains
1	1	2	1	1
2	2	3	2	1,2,3
3	3	5	3	2
4	4	5	4	4
5	5	6	5	3,4,5
			6	5

Chain Coordinates		Node Coordinates		
Chain	Coordinates	Node	X	Y
1	70,22 71,23	1	70	22
2	71,23 73,23	2	71	23
3	71,23 71,51 71,21	3	73	23
4	70,21 71,21	4	70	21
5	70,21 72,21 72,22 73,22	5	71	21
		6	73	22

Image 2: Demanding Teaching



Technical skill required

Image 3: Introductory Training



Decisions of cost, needed technical skill, applications, relevancy, and other considerations

Fig. 3-15: Disadvantages of the vector format.

2 - Ao contrário dos requisitos para a operação com dados matriciais, para eficiente manipulação de dados vetoriais há demanda por computadores mais robustos.

O gerenciamento de equipamentos de computação se torna mais do que um problema. Novamente, o poder de computação dos modernos resolve estas questões que passam a ser uma questão não tão importante nos dias atuais.

3 - O uso de computadores mais robustos, demandou mais por administração, e outras considerações que tornam mais caros os dados vetoriais.

Atualmente a maioria das pessoas, entretanto, possui equipamento relativamente barato, com poder de processamento para manipular a alta demanda dos SIG baseados em dados vetoriais.

4 - Aprender os aspectos técnicos dos sistemas com dados vetoriais é mais difícil do que entender a simplicidade da estrutura matricial, especialmente ao se considerar a topologia.

Isso torna o ensino mais difícil, especialmente para estudantes do ensino secundário ou do início da universidade. Na imagem 2 apresenta-se as demandas ao professor. Uma parte do conhecimento não é intuitiva e deve ser dita de modo cuidadoso.

Por causa destas considerações, alguns usuários proferem o formato matricial mais barato e mais simples. Por exemplo, os sistemas podem ser mais fáceis de usar em treinamento inicial para aplicações com poucos requisitos ou que têm limitado suporte financeiro e de recursos humano. A questão matricial versus vetorial deve ser decidida para estudantes iniciantes (imagem 3).

A tecnologia de computadores avançou muito nos últimos anos, de tal modo que, sistemas poderosos estão a preços acessíveis para a maioria dos usuários. Hoje há mais integração entre os formatos matricial e vetorial, e as diferenças entre vetorial e matricial são menos problemáticas. As desvantagens de ambos os formatos estão se tornando menos importantes. Talvez num futuro próximo matrizes e vetores não sejam obstáculos ao uso de dados em conjunto.