

Preenchimento de Células

Ana Paula Aguiar
Pedro Ribeiro de Andrade
Priscila Gregati Ferrari

Divisão de Processamento de Imagens (DPI)
Instituto Nacional de Pesquisa Espacial (INPE)

18 de fevereiro de 2008

O Plugin de Preenchimento de Células possibilita calcular valores para atributos de tabelas associadas a planos de informação do tipo celular. O objetivo é homogeneizar informações provenientes de diferentes fontes, em formatos distintos (dados vetoriais, matriciais e também outros planos celulares), agregando-os em uma mesma base espaço-temporal. Este plano celular servirá então como base para atividades de modelagem, em TerraME, ou outras aplicações, como, por exemplo, análise estatística, como ilustra a Figura 1.

Motivação: homogeneização de dados em formatos distintos

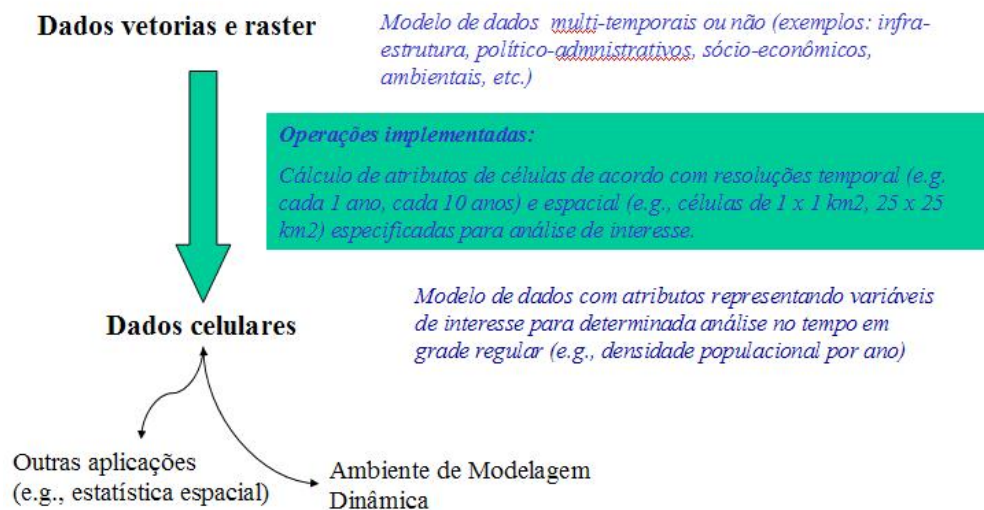


Figura 1: motivação para o uso de dados celulares

Este documento está organizado em três subitens:

- Modelo de dados celular – descreve brevemente os objetivos e os tipos de tabelas (dinâmica e estática) que podem estar associadas a um plano celular, e apresenta alguns exemplos.
- Operações – descreve conceitualmente as diversas operações que podem ser usadas para calcular os atributos, de acordo com o tipo e semântica do mesmo¹, e também apresenta alguns exemplos.
- Modo de utilização da ferramenta – explica como usar a ferramenta para preencher os atributos das células.

Todos os testes executados durante a elaboração deste documento usam o TerraView versão 3.2RC3.

Modelo de dados celular

Na versão atual da biblioteca Terralib, usada como base no desenvolvimento do TerraView e do plugin, células são retangulares. Por exemplo, elas podem ter resolução de 1m x 1m, 500m x 500m, 100km x 200km, de acordo com a necessidade da aplicação. A um mesmo plano celular podem estar associadas diversas tabelas estáticas (para atributos que não variam no tempo) e dinâmicas (para atributos que variam no tempo). O número total de tabelas e o seu conteúdo devem ser decididos de acordo com a aplicação final do usuário, e o modo como ele decide organizar seus dados. Tabelas estáticas são criadas por funções do TerraView externas ao plugin, e seus atributos calculados usando o plugin. Tabelas dinâmicas são criadas apenas no contexto do plugin (ver seção Modo de Uso).

A Tabela 1 apresenta um exemplo ilustrativo de especificação dos atributos de interesse para uma determinada aplicação. Recomenda-se que a montagem de uma especificação semelhante seja a primeira etapa na organização dos dados em formato celular. Após esta especificação, a segunda etapa é a criação de um banco de dados (caso não exista) e importação/organização dos dados vetoriais e matriciais. Uma vez organizados os dados de entrada, inicia-se a etapa de criação dos planos celulares e preenchimento dos atributos com o plugin (como descrito na seção Modo de Uso).

¹ Semântica x tipo do dado: por **semântica** dos atributos estamos nos referindo ao seu significado, não seu **tipo** (inteiro, real, string). Isto é, semanticamente um dado **categorico** representa uma classe temática (por exemplo, de uso de solo); um dado **numérico** representa uma quantidade (inteiro ou real); um dado **nominal**, representa um nome (ex., lugar, pessoa, etc.). Em Terralib/Terraview não existe uma definição da semântica de determinado atributo, apenas do seu tipo. Mas é importante que o usuário tenha em mente esta diferenciação para o entendimento de quais operações podem ser aplicadas a um determinado dado de entrada. Em relação ao tipo do dado, um dado categorico pode ser uma string ou um inteiro; um numérico, um inteiro ou real; um nominal, normalmente uma string.

Tabela 1: Exemplo de especificação de atributos

Atributo celular		Operação		Dados de entrada			
Atributo de saída	Valor		Representação	Layer	Tabela de atributos	Atributo	Valores
Atributos estáticos							
<i>object_id</i>							
dist_nucleo_urbanos	contínuo	Mínima distância	Vetorial (pontos)	Layer Centralidades do IBGE	n/a	n/a	n/a
fertilidade_solo	contínuo	Porcentagem de classe majoritária em área	Vetorial	Layer agro-ecológico (poligonal) com atributo indicando classe de fertilidade (10 classes)	agro_ecologico	FERTILIDAD	Catégorico (Alto, médio/baixo - reclassificar)
declividade	contínuo	Média	Raster	layer (grade) de declividade a ser gerado com base no SRTM	n/a	n/a	Contínuo
altitude	contínuo	Média	Raster	layer (grade) de declividade a ser gerado no SRTM	n/a	n/a	Contínuo
dist_rios	contínuo	Mínima distância	Vetorial (linhas)	Layer de rios duplos e simples	n/a	n/a	n/a
Atributos dinâmicos (97 - 01 - 04 - 06)							
<i>attr_id</i>							
<i>initial_time</i>							
<i>final_time</i>							
porc_floresta, porc_desflorestado, porc_outros	contínuo	Porcentagem por classe em área	Raster	Dados raster do PRODES	n/a	n/a	Catégorico (floresta, desflorestado, outros)
dist_estradas	contínuo	Mínima distância a linhas	Vetorial (Linhas)	Layers agregando todas as estradas	n/a	n/a	n/a
presenca_area_protegida	contínuo	TerraView:Presença	Vetorial (Polígonos)	Layers de áreas protegidas no tempo	n/a	n/a	n/a

No caso de tabelas estáticas, cada linha armazena os atributos de uma célula, como ilustra a Figura 2. Cada célula possui obrigatoriamente um identificador único denominado *object_id*. No caso de tabelas dinâmicas, a mesma tabela pode apresentar diversas linhas referentes a uma mesma célula, para tempos distintos. O identificador único nesse caso é o campo *attr_id*, que combina o *object_id* da célula, o tempo inicial e final da linha.

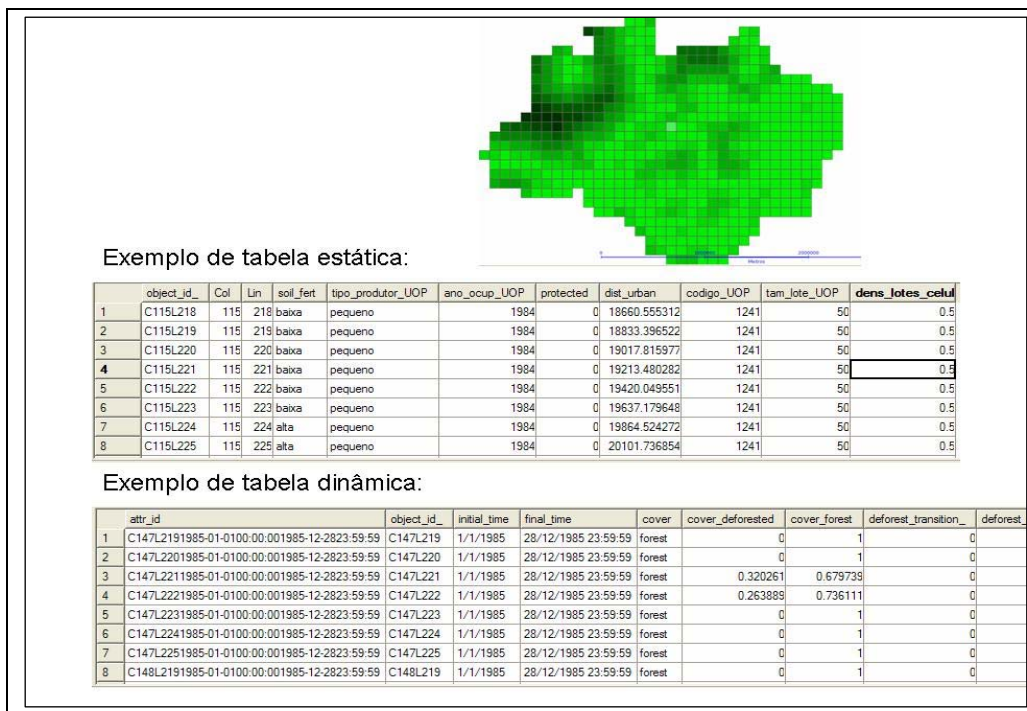


Figura 2: Exemplos de tabelas estática e dinâmica

Operadores para cálculo dos atributos

Existem vários operadores para cálculo de atributos das células. Cada um deles usa como entrada dados armazenados em temas, abrangendo todos os diferentes tipos de representação geométrica disponíveis no TerraView. A Figura 3 ilustra graficamente alguns desses casos para diferentes representações.

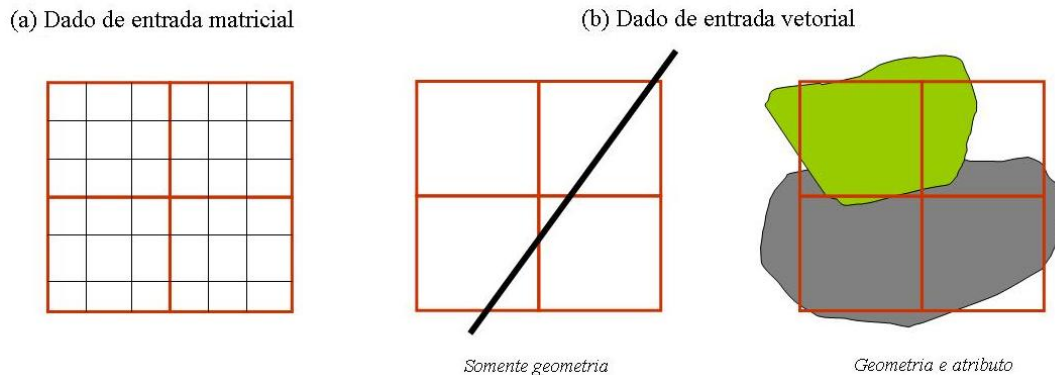


Figura 3: exemplo das diferentes representações

A ferramenta de preenchimento de células permite que os atributos tanto das tabelas dinâmicas quanto estáticas sejam calculados. De acordo com a representação geométrica e semântica dos atributos dos dados de entrada, diferentes operadores podem ser aplicados. Estes operadores são descritos a seguir, e estão sumarizados na Tabela 2.

(a) Dados de entrada em representação matricial ou celular:

- Caso os valores dos pixels/células de entrada representem valores *numéricos* (por exemplo, uma grade de temperatura ou altitude), pode-se calcular atributos com base em operadores de média, desvio padrão, valor máximo, mínimo e soma. Os valores dos atributos das células são calculados com base nos valores dos elementos (pixels ou células) que as interceptam.
- Caso os valores dos pixels/células representem valores *categóricos* (por exemplo, uma classe temática de solo), pode-se calcular atributos com base em operadores categoria majoritária, minoritária e porcentagem de cada categoria encontrada (nesse caso, mais de um atributo de saída é calculado, um para cada classe temática do dado de entrada). Do mesmo modo, os valores dos atributos das células são calculados com base nos valores dos elementos que interceptam a célula. Atualmente o plugin não possui suporte a dados categóricos, portanto dados neste formato devem ser descritos em tabelas usando atributos inteiros ou *string*.

(b) Dados de entrada com representação vetorial:

- O tipo mais simples de operação é baseado *apenas na geometria dos dados de entrada*, não considerando o valor de qualquer atributo. Entre estes operadores, encontram-se: distância mínima (do centro de cada célula ao vetor mais próximo),

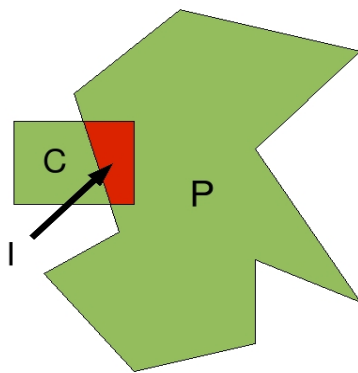
contagem de elementos vetoriais, presença de qualquer elemento vetorial. Todos estes operadores funcionam com dados pontuais, lineares e poligonais.

- Utilizando *os atributos dos dados de entrada*, existem dois tipos de operadores:
 - Operadores mais simples, similares aos descritos acima para dados matriciais, para valores de atributos numéricos (média, valor máximo, mínimo e soma) e categóricos (categoria majoritária, minoritária e porcentagem de cada categoria encontrada), considerando todos os elementos vetoriais que interceptam a célula.
 - Operadores mais complexos, envolvendo os valores dos atributos, mas também a área de intersecção de polígonos. Para atributos categóricos: categoria majoritária em área, porcentagem de cada categoria em área. Para atributos numéricos: média ponderada pela área de intersecção, soma ponderada pela área de intersecção.

Todas as operações listadas anteriormente são intuitivas, excluindo-se as funções “Soma ponderada pela área de intersecção” (SP) e “Média ponderada pela área de intersecção” (MP). A explicação do funcionamento destes dois operadores é pertinente. A SP conserva a soma total dos valores do atributo em todos os polígonos, redistribuindo os valores para as células. Portanto, ela é útil principalmente para a geração de atributos baseados em valores absolutos, como por exemplo, número de habitantes ou renda *per capita*. Para tanto, esta operação calcula a soma dos valores proporcionalmente à área de intersecção, e inversamente proporcional à área do polígono. Dado que:

$f(C)$ é a o atributo gerado para a célula C ,
 $f(P)$ é o atributo do polígono P ,
 $a(I)$ é a área de intersecção entre o polígono P e a célula C ,
 $a(P)$ é a área do polígono P ,

a fórmula superior da Figura 4 mostra o cálculo de SP. A MP calcula a média a partir dos valores de todos os polígonos, sendo bastante útil para valores relativos (como taxas e porcentagens), por exemplo, densidade populacional. Desta forma, ela é proporcional à área de intersecção, e inversamente proporcional à área da célula. A fórmula inferior da Figura 4 descreve o cálculo de MP.



$$f(C) = \sum_{I=P \cap C \neq \emptyset} f(P) \frac{a(I)}{a(P)}$$

$$f(C) = \sum_{I=P \cap C \neq \emptyset} f(P) \frac{a(I)}{a(C)}$$

Figura 4: Fórmulas para o cálculo de SP e MP.

Tabela 2: sumário das operações

Operador		
Tipo 1: Matricial, Celular e Vetorial (somente valor de atributo)	Tipo do atributo	Repres.
Valor mínimo	Numérico (inteiro ou real)	matricial
		vetorial
		celular
Valor máximo	Numérico (inteiro ou real)	matricial
		vetorial
		celular
Valor médio	Numérico (inteiro ou real)	matricial
		vetorial
		celular
Soma dos valores	Numérico (inteiro ou real)	matricial
		vetorial
		celular
Desvio padrão	Numérico (inteiro ou real)	matricial
		vetorial
		celular
Classe majoritária (número de elementos)	Categórico, nominal (inteiro ou string)	matricial
		vetorial
		celular
Porcentagem de cada classe (quantidade)	Categórico, nominal (inteiro, string)	matricial
		vetorial
		celular
Tipo 2: Vetorial poligonal (área de intersecção e valor do atributo)		
Média ponderada pela intersecção	Numérico (inteiro ou real)	polígonos
Soma ponderada pela intersecção	Numérico (inteiro ou real)	polígonos
Classe majoritária (em área)	Categórico, nominal (inteiro, string)	polígonos
Porcentagem da classe majoritária (área)	Categórico, nominal (inteiro, string)	polígonos
Tipo 3: Vetorial (apenas geometria)		
Distância Mínima	N/A	pontos
		linhas
		polígonos
Presença	N/A	pontos
		linhas
		polígonos
Contagem	N/A	pontos
		linhas
		polígonos
Porcentagem total de intersecção	N/A	polígonos

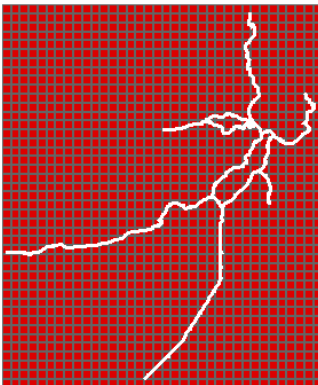
Modo de uso

Esta seção está dividida em três tópicos: (1) como criar um plano de células; (2) como criar novas tabelas estáticas em um plano celular existente; e (3) como usar o Plugin de Preenchimento de células propriamente dito. Os itens (1) e (2) são externos ao Plugin; o item (2) é opcional, porém o item (1) é necessário para iniciar o uso do mesmo.

Criação de plano de células

Um plano de células é criado a partir de um outro plano de informação. O TerraView possui duas estratégias para a criar células, dependendo do tipo do dado espacial do plano de informação de entrada. Caso seja usado um dado matricial, ou alguma geometria que não seja poligonal, as células serão criadas a partir do retângulo envolvente do plano de informação. Desta forma, as células cobrirão toda a extensão do plano de informação, de forma semelhante a um dado matricial, com um número constante de células em cada linha e em cada coluna do plano. Caso seja selecionado um conjunto de dados poligonal, as células serão criadas de forma a preencher todo o espaço dos polígonos, mas serão criadas células apenas onde houver interseção espacial entre algum polígono e a célula a ser criada. A Figura 5 mostra a diferença entre o uso de dados poligonais e não-poligonais na criação de um plano de células.

(a) Células a partir de linhas



(b) Células a partir de polígonos

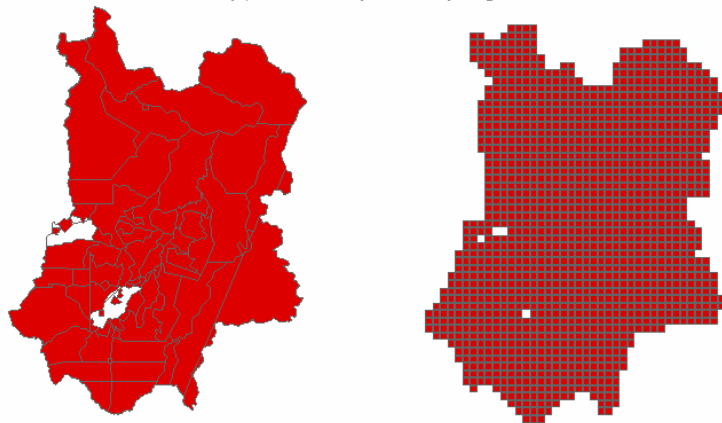


Figura 5: Estratégias para criação de planos de informação celulares

Para criar um plano de informação de células, clique com o botão direito no plano de informação desejado, como mostra a Figura 6, e então clique na opção Criar Células. Uma nova janela aparecerá, como mostrado na Figura 7. Preencha a resoluções para os eixos X e Y (que podem ser diferentes), escolha um nome para o plano de informação a ser criado, e

então pressione “Executar”. Ao criar um plano de informação de células, o TerraView cria automaticamente uma tabela estática, com o mesmo nome do plano de informação.

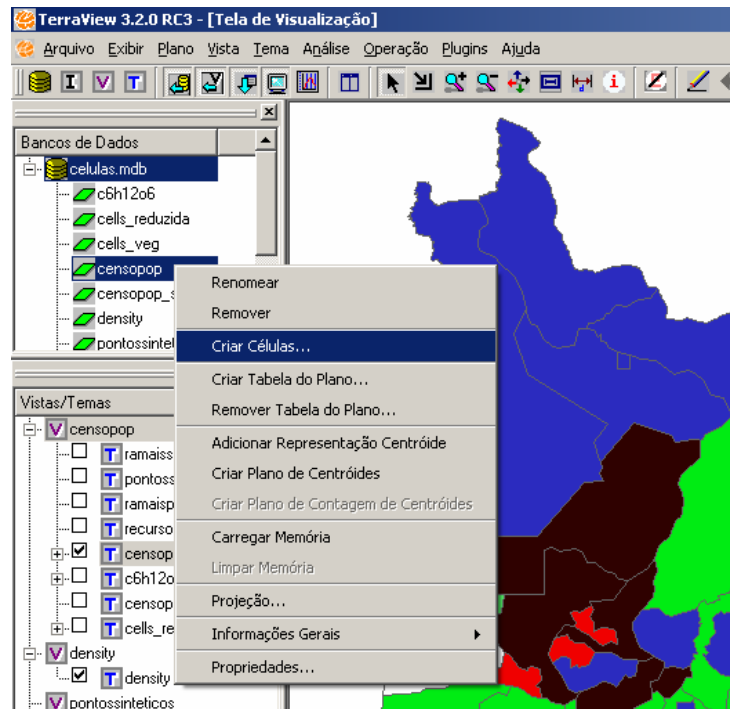


Figura 6: Criação de um plano de células no TerraView

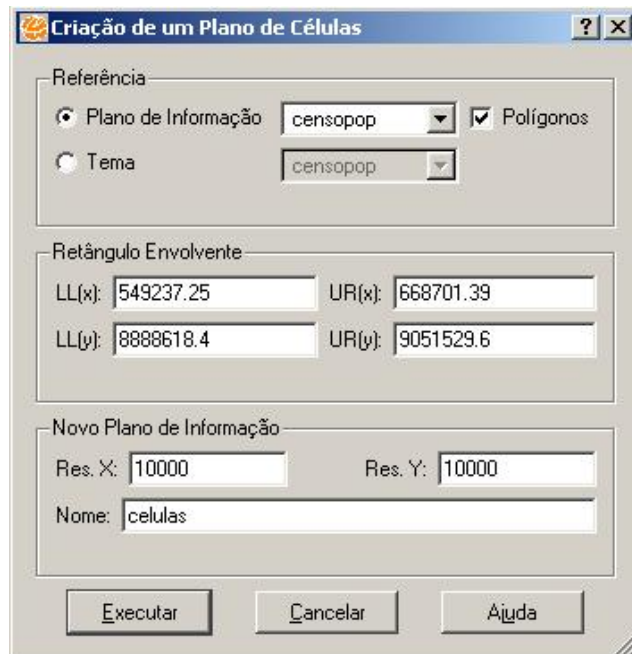


Figura 7: Parâmetros para criação de um plano de células

Criação de tabelas estáticas adicionais

Para criar tabelas estáticas adicionais, deve-se clicar com o botão direito sobre o plano de informação no qual se deseja adicionar a tabela, tal como na criação de um plano de informação celular, e então clicar na opção “Criar Tabela do Plano”. Então aparecerá a janela mostrada na Figura. Para criar a tabela, basta escolher o seu tipo (estático), e colocar no nome da tabela. Note que o nome desta tabela deve ser diferente de qualquer outro nome de tabela do banco de dados, não apenas do plano de informação. A opção inserir objetos (*default*) cria uma linha na tabela para cada geometria existente. Caso esta opção não seja escolhida, a tabela será criada vazia.

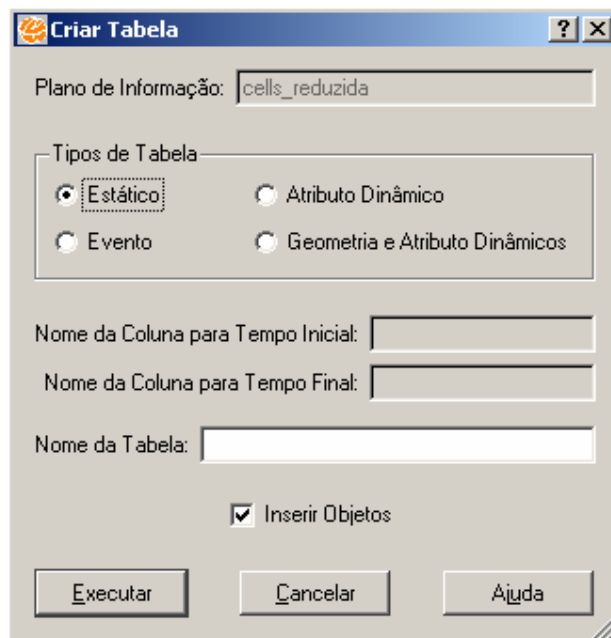
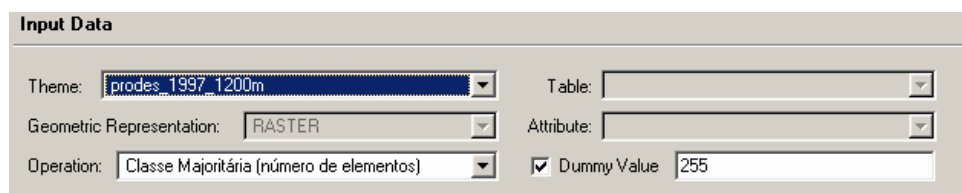


Figura 7: criação de tabelas adicionais

Plugin de Células

Inicialmente devem ser informados alguns dados referentes ao tema de entrada. Primeiro, escolha qual o tema de entrada e a representação geométrica e a ser usada. Em seguida, para os temas de dados vetoriais, a representação geométrica (pontos, linhas ou polígonos) deve ser indicada, caso exista mais de uma para o tema indicado. Se o tema for de dado matricial, ou se ele possuir apenas uma geometria, esta opção será desabilitada. De acordo com a geometria selecionada, serão habilitadas diferentes opções de operação. Selecione uma delas. Depois, também para os temas de dados vetoriais, e de acordo com a operação escolhida, um atributo pode ser necessário. Em caso positivo, uma tabela e um atributo devem ser selecionados para serem usados na operação. Diferentes operações aceitam diferentes tipos de atributo como parâmetro (por exemplo, a operação de Valor Mínimo aceita somente os atributos do tipo inteiro e real), portanto é necessário escolher

primeiramente a operação, de forma que o plugin habilitará apenas os atributos válidos. Se o tema for de dado matricial, esta opção será desabilitada. A opção de valor dummy pode ou não ser usada, e serve para ignorar um determinado valor do processamento, toda vez que este for encontrado. Este valor está tipicamente associado a valores de transparência em imagens, embora ele também possa ser usado em outras situações. Um exemplo típico de uso do valor dummy é o cálculo de valores mínimos, quando existem pixels na imagem cujo valor é maior que o valor usado como transparência. A Figura 5 mostra um exemplo de preenchimento dos dados de entrada.

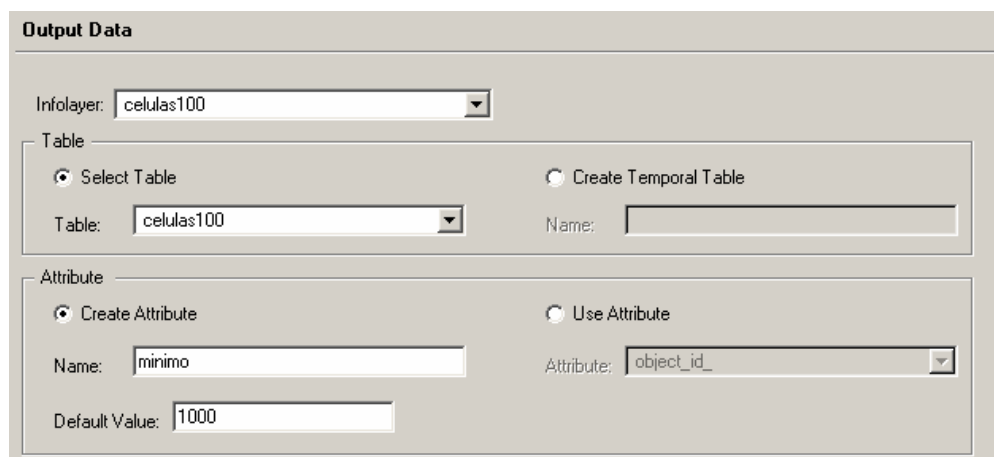


The 'Input Data' dialog box is shown with the following settings:

- Theme: prodes_1997_1200m
- Table: (empty)
- Geometric Representation: RASTER
- Attribute: (empty)
- Operation: Classe Majoritária (número de elementos)
- Dummy Value: 255

Figura 5: Exemplo de uso do dado de entrada.

Após selecionar os dados de entrada, o próximo passo é definir o plano de informação onde o resultado será armazenado. Este PI deve ser obrigatoriamente do tipo celular, e ter sido criado anteriormente. Em seguida, deve ser indicado se será criada uma nova tabela temporal (dinâmica) para armazenar o resultado, ou se será usada uma tabela já existente (estática ou dinâmica). Se optar por criar uma nova tabela, o nome desta deve ser indicado. Se optar por usar uma tabela existente, ela deve ser selecionada de uma lista. Após isto, deve ser indicado se será criado um novo atributo (para a tabela a ser criada ou tabela indicada), ou se será usado um atributo já existente, no caso da tabela também já existir. Se optar por criar um novo atributo, deve indicar o nome que será utilizado para o mesmo. Se optar por utilizar um atributo já existente, deve selecionar o atributo desejado na lista de atributos existentes. Note que usar um atributo existente implica em sobrescrever os seus valores, a menos que o dado seja temporal e um novo tempo esteja sendo adicionado. A Figura 6 mostra um exemplo de escolha do nome do atributo, e de onde este atributo será armazenado.



The 'Output Data' dialog box is shown with the following settings:

- Infolayer: celulas100
- Table:
 - Select Table
 - Create Temporal Table
 - Table: celulas100
 - Name: (empty)
- Attribute:
 - Create Attribute
 - Use Attribute
 - Name: minimo
 - Attribute: object_id_
 - Default Value: 1000

Figura 6: Exemplo de atributo a ser criado

O próximo passo é necessário somente para tabelas temporais. Se uma tabela temporal for usada, é preciso indicar se será criado um novo tempo ou se será usado um tempo existente (somente nos casos em que algum atributo já exista). Se optar por criar um novo atributo, o seu nome deve ser indicado, além dos tempos inicial e final, o formata do dado, o índice AM-PM e o separador do dado. Se optar por utilizar um atributo já existente, deve selecionar o atributo desejado na lista de atributos existentes.

Ao final da tela, encontram-se os botões:

Executar: efetua a operação de preenchimento de células

Cancelar: fecha a janela

Ajuda: exibe a janela de ajuda

Sobre...: exibe a janela de “Sobre o Plugin”

Sumário das Operações

A lista abaixo descreve os tipos de operações disponíveis.

Operação	Atributo*	Matricial	Celular	Ponto	Linha	Polígono
Valor Mínimo	N - IR	X	X	X	X	X
Valor Máximo	N - IR	X	X	X	X	X
Valor Médio	N - IR	X	X	X	X	X
Soma	N - IR	X	X	X	X	X
Desvio Padrão	N - IR	X	X	X	X	X
Classe Majoritária (número de elementos)	C - IS	X	X	X	X	X
Porcentagem de Cada Classe (quantidade)	C - IS	X	X	X	X	X
Média Ponderada Pela Interseção	N - IR					X
Soma Ponderada Pela Interseção	N - IR					X
Classe Majoritária (área)	C - IS					X
Porcentagem da Classe Majoritária (área)	C - IS					X
Distância Mínima				X	X	X
Presença				X	X	X
Contagem				X	X	X
Porcentagem Total de Interseção						X

* *tipo*: S = string, I = inteiro, R = Real; *categoria*: N: numérico (IR); C: categórico (IS).