

Automatização do processo de geração dos resultados do projeto Zarc

Daniel Meneguim Barbosa¹
Renato José Santos Maciel²

Resumo: Este trabalho trata sobre a alteração de ferramentas do *workflow* de simulações do zoneamento agrícola de risco climático do software *Galaxy Project* para que estas sejam executadas de forma mais dinâmica e apresentem resultados mais acessíveis ao estudo. Foram feitas diversas mudanças no funcionamento e na interface de cada uma das ferramentas e constatou-se que estas facilitaram o acesso do usuário aos dados da simulação, disponíveis agora em um banco de dados relacional e a execução de cada uma das ferramentas ficou mais simples, uma vez que os parâmetros são previamente carregados do banco.

Palavras-chave: Zoneamento Agrícola de Risco Climático, *workflow*, Galaxy project.

¹Estudante de Engenharia de Computação da Universidade Estadual de Campinas (Unicamp), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

²Cientista da Computação, analista da Informação da Embrapa Informática Agropecuária, Campinas, SP.

Introdução

O Zoneamento Agrícola de Risco Climático (Zarc) é um instrumento de política agrícola e gestão de riscos na agricultura (BRASIL, 2017). São feitos trabalhos de identificação, quantificação e mapeamento de áreas mais favoráveis ao plantio das diversas culturas, levando-se em conta o clima e a distribuição pluviométrica. Esses dados são processados e geram-se mapas que mostram regiões e períodos do ano que apresentam as condições mais favoráveis à semeadura.

Este trabalho aborda a atualização do *workflow* desenvolvido no *Galaxy Project* para executar simulações do Zarc. O software *Galaxy Project* (GALAXY PROJECT, 2017) é uma plataforma web de código aberto que permite o desenvolvimento e a execução de ferramentas e *workflows* por meio de uma interface simples e robusta.

A versão anterior do *workflow* apresenta o mesmo conjunto de ferramentas, porém estas consultam e gravam arquivos do disco com parâmetros e resultados das simulações. Este método é pouco escalável pois é muito dependente da manipulação e gerenciamento do disco, além de tornar o estudo das simulações complexo por necessitar do acesso a múltiplos arquivos gerados em processos intermediários.

Assim, o objetivo desse trabalho é introduzir melhorias na integração de ferramentas do software *Galaxy Project*, com o intuito de reduzir o tempo de acesso a disco, melhorar as funcionalidades do software, equipando-o melhor para o estudo de simulações no âmbito do projeto Zarc.

Materiais e Métodos

O software *Galaxy Project* permite o desenvolvimento de ferramentas de funções específicas dentro de um ambiente autogerenciado. O software oferece um mecanismo próprio de gerenciamento dos arquivos de dados e apresenta interface web com diversos componentes gráficos que podem ser personalizados para cada ferramenta. As ferramentas podem ser inseridas em um *workflow*, onde os dados de saída de uma ferramenta alimentam a entrada de outra.

A versão anterior do *workflow* apresenta quatro ferramentas: Sarrazon, Gerador de CSV, Interpolador e Calculador de área apta. O Sarrazon recebe como entrada um arquivo com datas (decêndios), estações, parâmetros do cenário e coeficientes da cultura. Esses dados são processados e geram-se arquivos com valores de risco por estação, fase fenológica da cultura e decêndio. Esses arquivos são lidos pelo gerador de CSV, que insere os dados dos arquivos no banco de dados. O Interpolador consulta os dados gravados pelo gerador de CSV, recebe os valores de risco do usuário e faz interpolações, gerando um *raster* por fase, frequência, decêndio e cenário. Esses *rasters* são lidos pelo Calculador de área apta, que aplica cortes por fase (só considera áreas aptas com valor de risco maior que o corte), filtros, seleção de zonas e seleção de municípios e grava no banco de dados a aptidão de cada município em cada decêndio.

Cada uma das ferramentas foi alterada para que houvesse menos interações com arquivos gravados em disco e apresentasse acesso mais simples e direto ao usuário. Substituiu-se a manipulação de arquivos por métodos de escrita e consulta a um banco de dados relacional e alteraram-se as interfaces utilizando-se componentes gráficos que permitiram-se mostrar dados gravados no banco, evitando-se a necessidade de *upload* de arquivos com dados pelo usuário.

Dividiu-se o desenvolvimento em quatro etapas, cada uma composta pela alteração de uma das ferramentas, seguindo a ordem de execução do *workflow*.

Resultados e Discussão

Na primeira etapa alterou-se a ferramenta Sarrazon, responsável por executar a simulação do Sistema de Análise Regional dos Riscos Agroclimáticos (Sarra). A Figura 1 apresenta a ferramenta Sarrazon.

Simulador Sarrazon Executa um lote de simulações (Galaxy Version 1.2) Options

Arquivo com datas

Arquivo com estacoes

Arquivo com parametros

Selecao de cultura

Figura 1. Ferramenta Sarrazon.

Na Figura 1, observam-se quatro componentes gráficos contendo listas de itens, sendo estes o arquivo de datas, estações, parâmetros e culturas. Na seleção de cultura, criou-se outro método para listar os nomes das culturas disponíveis no banco de dados. Ao executar esta ferramenta, os dados da cultura selecionada são carregados diretamente do banco.

A segunda etapa constituiu-se de alterar a ferramenta geradora de CSV, responsável por abrir os arquivos gerados pela ferramenta Sarrazon e organizar o conteúdo deles em tabelas contendo o valor de risco por estação e decêndio. A Figura 2 apresenta a ferramenta Gerador de CSV do Sarra.

Gerador de csv do sarra Cria os arquivos csv do sarra (Galaxy Version 1.0) Options

Arquivo com id do job sarra

Arquivo com parametros da simulacao

Arquivo com decendios

id da simulacao com codigo do cenario

Fase
☐ Select/Unselect all
☐ Fase 1
☐ Fase 2
☐ Fase 3
☐ Fase 4

Figura 2. Ferramenta Gerador de CSV do Sarra.

Na Figura 2 observam-se cinco componentes gráficos, contendo listas de identificadores de execução da ferramenta Sarra, parâmetros de cada simulação, decêndios, arquivos com identificadores das simulações e cenários correspondentes e o último componente gráfico com a seleção de fases. Desenvolveu-se um método que grava as tabelas geradas pela ferramenta diretamente no banco. A terceira etapa consistiu em alterar a ferramenta de interpolação. A Figura 3 apresenta a ferramenta de interpolação.

Interpolador Interpolação (Galaxy Version 1.0) Options

Arquivo com id do job csv

2188: Gerador de csv do sarra on data 2187, data 2186, and others

Isna

☐ Select/Unselect all

☐ 90%

☐ 80%

☐ 70%

☐ 60%

☐ 50%

☐ 40%

☐ 30%

☐ 20%

☐ 10%

Arquivo com fases

2189: Gerador de csv do sarra on data 2187, data 2186, and others

Arquivo com id das simulações

2190: Gerador de csv do sarra on data 2187, data 2186, and others

Arquivo com decendios

2172: Gerador de csv do sarra on data 2168, data 2167, and others

✓ Execute

Figura 3. Ferramenta de interpolação.

Na Figura 3 observa-se que a ferramenta possui cinco elementos gráficos, quatro elementos para a seleção de arquivos e um para a seleção do risco. Criou-se um método para carregar a tabela de riscos do banco por meio do identificador selecionado no primeiro elemento gráfico (arquivo com id do job csv).

A quarta etapa consistiu em alterar a ferramenta Cálculo de área apta, apresentada na Figura 4. Na Figura 4 observa-se que a ferramenta possui dez elementos gráficos. Para a seleção de filtros, utilizou-se um elemento gráfico que permite a seleção de múltiplos itens da lista, que foi populada com os filtros presentes no banco de dados. Para a inserção de valores de corte, há quatro elementos gráficos que permitem valores reais. Criou-se um método de tratamento da execução desta ferramenta, que carrega os mapas gerados pela interpolação, aplica cortes por fase e os filtros selecionados.

Nos campos de seleção de arquivos de todas as ferramentas, desenvolveu-se um método que filtra o tipo correspondente da seleção, evitando que arquivos incorretos sejam selecionados para a simulação.

Criou-se um novo workflow que permite a execução em série de todas as ferramentas desenvolvidas.

Calcula área apta de municípios Calcula área apta sobre os rasters gerados pelo interpolador (Galaxy Version 1.0) Options

id da simulacao com codigo do cenario
 2190: Gerador de csv do sarra on data 2187, data 2186, and others

Arquivo com decendios
 2172: Gerador de csv do sarra on data 2168, data 2167, and others

Arquivo com fases
 2189: Gerador de csv do sarra on data 2187, data 2186, and others

Arquivo com frequencias
 2245: Interpolador on data 2172, data 2190, and others

ID do job da interpolação
 2244: Interpolador on data 2172, data 2190, and others

Seleção de filtros
☐ Select/Unselect all

Filtro Teste

Tmed 21 - Geada - Temp. Solo
 Correção do Xavier - Milho Safrinha v2 - Tmin <= 12
 Correção do Xavier - Milho Safrinha v2 - Tmin combinada: 10, 13, 14
 Corte temperatura média geada milhosoja v1 - Tmed <= 21.5
 Correção do Xavier - Milho Safrinha
 sem filtro

1.0

corte fase 2
 1.0

corte fase 3
 1.0

corte fase 4
 1.0

Figura 4. Ferramenta Cálculo de área apta.

Considerações finais

Observa-se que o novo workflow substituiu em grande parte a consulta a dados gravados em arquivos em disco pela consulta ao banco de dados. Essa característica permite escalar a execução de simulações mais eficientemente, evitando a necessidade de gerenciamento de disco e apresentando uma interface mais completa, que evita a manipulação manual de código e de arquivos de entrada. Os dados gravados no banco de dados também permitem maior acessibilidade aos estudos das simulações, uma vez que para acessar os parâmetros e resultados, bastam consultas ao banco de dados em vez de uma grande quantidade de arquivos intermediários gerados durante os processos.

Para maior eficiência do workflow, é necessário que as ferramentas Sarrazon e Interpolador insiram diretamente os dados de saída no banco de dados em vez de gerar arquivos intermediários. Além disso, todos os arquivos que contém parâmetros da simulação passam pelas ferramentas do workflow, sendo copiados da entrada para a saída. Este método é pouco escalável para o desenvolvimento de ferramentas novas, pois para cada ferramenta nova é necessário modificar todas as outras. Uma possível solução é utilizar a capacidade do workflow do Galaxy Project de direcionar um arquivo para múltiplas entradas, porém esta solução deixa o workflow mais complexo e torna mais difícil a interpretação.

Referências

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Zoneamento agrícola**. Brasília, DF, 8 mar. 2017. Disponível em: <<http://www.agricultura.gov.br/assuntos/riscos-seguro/risco-agropecuário/zoneamento-agricola>> . Acesso em: 14 set. 2017.

GALAXY Project. Disponível em: <<https://galaxyproject.org>> . Acesso em: 14 set. 2017.