

Manual do Optis

Índice

1 Sobre o programa.....	3
2 Como compilar e executar o Optis.....	3
3 Como configurar o Optis.....	4
4 Exemplo de configuração.....	5
4.1 Script.....	6
4.2 CONF.H.....	7
5 Referências.....	9

1 Sobre o programa

O Optis é um software de calibração hierárquica multiobjetivo escrito na linguagem C. A paralelização do código foi feita com OpenMP. Para maiores informações sobre a metodologia consultar [Varejão *et al.*, 2013]. O software foi baseado no algoritmo genético NSGA-II [Deb *et al.*, 2002].

De maneira resumida, o Optis funciona da seguinte maneira: o algoritmo recebe uma lista de parâmetros a serem calibrados; utilizando técnicas baseadas na evolução genética cria várias configurações de parâmetros diferentes; executa o modelo com cada uma dessas configurações; e, por fim, verifica qual configuração apresenta o melhor resultado. A Figura 1 exemplifica este funcionamento.

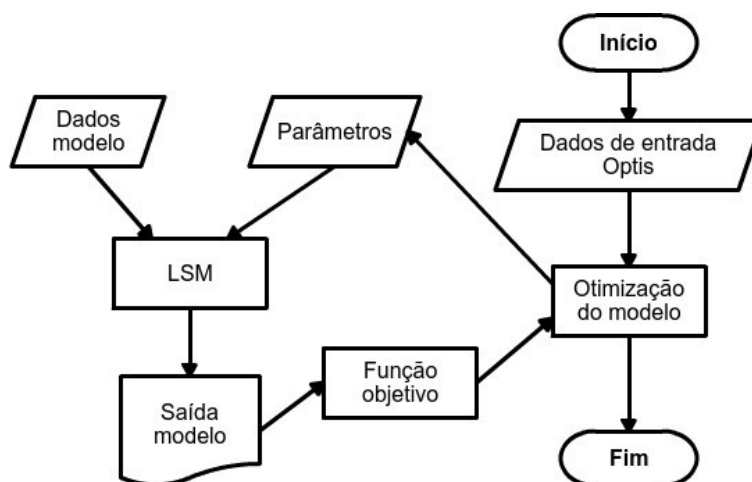


Figura 1: Estrutura do Optis.

O programa foi idealizado para calibrar o modelo INLAND, que é o pacote de superfície do Modelo Brasileiro do Sistema Climático Global, representado pela sigla LSM na Figura 1. Porém, o Optis pode ser adaptado para calibrar outros modelos.

2 Como compilar e executar o Optis

Para compilar o Optis, execute os seguintes comandos no diretório raiz do programa:

'./configure' (Para fazer a instalação do ambiente de compilação).

'make' (Para compilar).

Para executar o Optis, execute o script ./runoptis.sh na pasta raiz do programa.

3 Como configurar o Optis

Para executar o Optis é necessário que o arquivo de configuração (conf.h) esteja dentro do diretório input ('data/input'). Este arquivo (conf.h) possui a seguinte estrutura:

- Sete parâmetros referentes ao algoritmo NSGA-II [Deb *et al.*, 2002]:
 - popsize: Total de gerações do algoritmo genético (Inteiro);
 - ngen: Tamanho de cada população (Inteiro múltiplo de 4);
 - pcross_real: A probabilidade de crossover (Real, entre 0 e 1);
 - pmut_real: A probabilidade de mutação (Real, entre 0 e 1);
 - eta_c: Índice de distribuição de crossover (Inteiro);
 - eta_m: Índice de distribuição de mutação (Inteiro);
 - seed: Uma semente para o algoritmo de geração de número aleatório (Real);
- Três parâmetros para execução do modelo:
 - Um script que define uma estrutura temporária do modelo a ser calibrado (string);
 - O nome do executável do modelo que foi definido pelo script anterior (string);
 - Nome do arquivo de saída do modelo (string);
- Dados relativos aos parâmetros do modelo que serão calibrados:
 - Número de arquivos do modelo que possuem parâmetros para serem calibrados (Inteiro);
 - Nome(s) do(s) arquivo(s) de parâmetros (string);
- Dados relativos aos sítios de calibração:
 - Número de sítios que serão utilizados (Inteiro);
 - Para cada sítio:
 - Nome do sítio (string);
 - O total de dados (linhas) que serão usados na comparação entre observado e simulado (Inteiro);
 - Número de colunas que existe no arquivo de dados observados (Inteiro);

- Número de colunas que existe no arquivo de output do modelo (Inteiro);
 - Número de linhas que serão excluídas na leitura dos dados observados (Inteiro);
 - Número de linhas que serão excluídas na leitura do output do modelo (Inteiro);
 - Caminho do arquivo de dados observados (string);
 - Número de links simbólicos necessários para cada sítio (Inteiro);
 - O alvo (string) e o nome (string) de cada link;
- Dados relativos aos níveis hierárquicos:
- Total de níveis hierárquicos (Inteiro);
 - Para cada nível:
 - Modo de calibração (Inteiro): 0 (para que todas as funções objetivo sejam independentes) e 1 (para que funções objetivo de mesmo nome sejam agrupadas);
 - Nome do nível de calibração (string);
 - Total de parâmetros calibrados (Inteiro);
 - Para cada parâmetro: Nome do parâmetro (String), índice do arquivo onde ele se encontra (Inteiro), a linha (Inteiro), a coluna de início (Inteiro), o espaço a ser alterado (Inteiro), limite mínimo (Real) e limite máximo (Real);
 - Total de funções objetivo (Inteiro);
 - Para cada F.O.: Nome (String), índice do sítio (Inteiro), medida de erro (Inteiro), coluna no arquivo de dados observados (Inteiro), coluna no arquivo de output do modelo (Inteiro);

É importante ressaltar que a ordem dos dados no arquivo de configuração (conf.h) precisa ser respeitada para que ele possa ser lido corretamente pelo algoritmo.

4 Exemplo de configuração

Exemplos do script de estrutura (Seção 4.1) e do arquivo conf.h (Seção 4.2) para uma calibragem do modelo INLAND (no modo pontual).

4.1 Script

```
1  #!/bin/bash
2
3  mkdir -p data/offline/single_point;
4
5  mkdir data/offline/single_point/output;
6  mkdir data/offline/single_point/params;
7
8  ln -s data/offline/single_point/conf      conf;
9  ln -s data/offline/single_point/params    params;
10 ln -s data/offline/single_point/input      input;
11 ln -s data/offline/single_point/output     output;
12
13 ln -s /home/user/Programas/inland/inland-single_point    inland;
14
15 cp /home/user/Documentos/parametros/*    params;
```

Esse arquivo é um shell script que gera toda a estrutura fixa (que não varia de acordo com o sítio) que o modelo INLAND precisa pra ser executado no modo pontual. Esse script será usado para criar a seguinte estrutura (temporariamente) para cada indivíduo em cada sítio:

```
|----- inland
|----- input
|----- conf
|----- params
|----- output
|----- data
|         |----- offline
|         |         |----- single_point
|         |         |         |----- params
|         |         |         |----- output
```

Na linha 13 é feito um link para o binário do modelo com o nome “inland”. Esse nome precisa ser inserido no arquivo conf.h (na linha 12, Seção 4.2).

Na linha 15 uma cópia dos arquivos de parâmetros iniciais é feita para a pasta params. É importante que todos os arquivos de parâmetros sejam copiados pois o Optis irá alterar o valor dos parâmetros paralelamente para cada indivíduo da população.

4.2 CONF.H

```
1  /* CONF.H */
2
3  10          /* Número de gerações */
4  4           /* Tamanho da população */
5  0.9         /* pcross_real */
6  0.5         /* pmut_real */
7  10          /* eta_c */
8  10          /* eta_m */
9  0.5         /* seed */
10
11 /home/user/Programas//optis/data/input/struct.sh
12 inland
13 output/single_point-output.csv
14
15 4            /* Total de arquivos de parâmetros do modelo */
16 params/canopy
17 params/fire
18 params/soil
19 params/vegetation
20
21 2            /* Total de sítios */
22
23 km67
24 26280
25 7
26 13
27 1
28 78841
29 /home/user/Documentos/km67/obs_hourly.csv
30 3
31 /home/user/Documentos/km67/conf      data/offline/single_point/conf
32 /home/user/Documentos/km67/input    data/offline/single_point/input
33 /home/user/Documentos/km67/single_point_parameters
   data/offline/single_point/params/single_point_parameters
34
35 FNS
36 26280
37 7
38 13
39 1
40 78841
41 /home/user/Documentos/fns/obs_hourly_fns.csv
42 3
43 /home/user/Documentos/fns/conf_fns   data/offline/single_point/conf
44 /home/user/Documentos/fns/input_fns  data/offline/single_point/input
45 /home/user/Documentos/fns/single_point_parameters_fns
   data/offline/single_point/params/single_point_parameters
46
47 2            /* Total de níveis hierárquicos */
48
49 1
50 Rn-H
51 2
```

```

52 tau15 1 38 10 8 4000 5000
53 ko15 1 40 10 8 0.1 0.5
54
55 4 /* Total de funções objetivas */
56 Rn 1 0 4 5
57 H 1 0 5 8
58 Rn 2 0 4 5
59 H 2 0 5 8
60
61 0
62
63 LE-NEE
64 1
65 kc15 1 39 10 8 1.5e-05 2.5e-04
66
67 4 /* Total de funções objetivas */
68 LE 1 0 6 7
69 NEE 1 0 7 10
70 LE 2 0 6 7
71 NEE 2 0 7 10

```

Esse arquivo aceita blocos de comentários, sendo que ‘/*’ marca o início e ‘*/’ o fim do bloco.

Na linha 11 é definido o caminho para o script da Seção 4.1. Na linha 12 é indicado o nome do executável que foi definido no script. A linha 13 indica o local e o nome do output do modelo, dentro da estrutura criada.

Nas linhas 16 a 19 estão indicado os arquivos de parâmetros dentro da estrutura que foi criada pelo script.

Nessa configuração dois sítios são definidos (linha 21). Cada sítio precisa definir o arquivo de dados observados (linhas 29 e 41) que será usado na comparação com o output do modelo. Além disso, cada sítio precisa definir os links (linhas 31, 32, 33, 43, 44, 45) que serão criados pra completar a estrutura do INLAND. Esses links não podem ser criados pelo script anterior porque eles variam de sítio para sítio.

Nessa configuração são definidos dois níveis hierárquicos (linha 47). Cada nível precisa definir os parâmetros que serão calibrados (linhas 52, 53 e 65). Cada parâmetro precisa definir em qual arquivo (definido entre as linhas 16 a 19) ele está, sendo que 1 indica o primeiro arquivo da lista (canopy), 2 o segundo (fire) e assim sucessivamente. Além disso, é preciso definir a linha, a coluna (onde começa o parâmetro) e o espaço (que será substituído pelo Optis com os valores testados na calibração). O Optis escreve os parâmetros em notação científica por isso é preciso de

pelo menos 7 espaços: 1 para o sinal (+ ou -), 1 para a parte inteira do número, 1 para o ponto, 1 para o símbolo E (elevado), 1 para o sinal do expoente e 2 para o expoente.

Cada nível também precisar definir as suas funções objetivo (linhas 56 a 59 e 68 a 71). Cada F.O. tem um nome, sítio (segue a ordem apresentada no arquivo sendo, nesse caso, 1: km67 e 2: FNS), medida de erro (estão implementadas duas medidas de erro, 0: MAE e 1:RMSE), posição no arquivo de dados observados e posição no output do modelo .

5 Referências

- Deb, K., Pratap, A., Agarwal, S., e Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6 (2), 182-197, doi: 10.1109/4235.996017.
- Varejão, C. G., Costa, M. H., e Camargos, C. C. S., 2013. A multi-objective hierarchical calibration procedure for land surface/ecosystem models, Inverse Problems in Science and Engineering, 21 (3), 357-386, doi: 10.1080/17415977.2011.639453.