

Otimização Computacional do Modelo BRASIL-SR

SOUZA, Jefferson Gonçalves; MENDES, Celso Luiz; SANTOS, Rodrigo Costa

Mestrando em Computação Aplicada – Instituto Nacional de Pesquisas Espaciais
(INPE) – São José dos Campos – SP – Brasil

Jefferson.souza, celso.mendes, rodrigo.costa@inpe.br

***Abstract.** This paper describes the improvement of computational performance of the BRASIL-SR model. This computational code, which is used for the estimate of incident surface solar radiation, was optimized with an application of OpenMP directives and code modifications for the implementation of data output (I/O) operations using NetCDF format files. After these implementations, it was verified that the speedup of the code was further improved with the I/O optimizations, with a more stable efficiency, which shows a better use of the Santos Dumont supercomputer.*

***Resumo.** Este artigo apresenta a melhoria de desempenho computacional do modelo BRASIL-SR. Este código computacional, utilizado para a estimativa da irradiação solar incidente na superfície, foi otimizado com a aplicação de diretivas de OpenMP e modificações do código para implementação de operações de entrada e saída de dados (I/O) utilizando arquivos em formato NetCDF. Após estas implementações, verificou-se que o speedup do código foi ainda maior com um novo tratamento de I/O, gerando uma eficiência mais estável, o que mostra uma melhor utilização do supercomputador Santos Dumont.*

1. Introdução

Diante da preocupação em atender a demanda energética mundial sem aumentar as emissões de gases do efeito estufa, diversos países têm buscado soluções relacionadas com a inserção de fontes limpas em suas matrizes energéticas. A geração solar tem crescido vertiginosamente no Brasil, principalmente depois dos déficits hídricos registrados nos últimos anos em função da variabilidade climática. Condições de estiagem afetam a geração de energia elétrica, já que os reservatórios podem vir a atingir níveis muito baixos; isso também levanta questões relativas ao conflito do uso da água, que também é utilizada no abastecimento e na agricultura. Há uma necessidade iminente da diversificação da matriz elétrica brasileira, o que minimizaria problemas relacionados à Segurança Hídrica, Energética e Alimentar. Também é importante que esta diversificação ocorra no sentido da inserção de fontes renováveis, com baixa emissão de gases.

Antes da implantação de empreendimentos solares, é necessário conduzir estudos de verificação, viabilidade e disponibilidade do recurso. Com o mercado de energia em alta, a obtenção de dados confiáveis de irradiação solar acaba sendo um problema, já que são poucas as estações de medição com a qualidade e com séries históricas disponíveis para este fim. Além disso, a utilização de dados medidos para o

mapeamento fica inviável, já que a interpolação dos valores fica comprometida pela baixa cobertura espacial em países como o Brasil. Assim, a utilização de códigos computacionais, tais como o BRASIL-SR, que realiza estimativas precisas a respeito do potencial solar no Brasil é o melhor modo de se obter dados confiáveis.

Neste trabalho, busca-se a melhoria do desempenho computacional relacionado ao tempo de processamento do modelo de transferência radiativa BRASIL-SR, que devido ao seu alto custo computacional, possui um elevado tempo de processamento em processadores convencionais. Na primeira fase deste estudo, está sendo realizada uma análise do desempenho da versão original do modelo – que é executada em modo serial – e utilizando a ferramenta GNU Profiler (GPROF) com o objetivo de indicar quais são os processos mais custosos computacionalmente. Posteriormente, está sendo utilizada a inserção das diretivas de OpenMP, permitindo a criação e o gerenciamento automático de threads, nas regiões críticas, e a alteração dos dados de entrada e saída (I/O), para o formato Netcdf.

2. Metodologia

Neste trabalho está sendo utilizado o modelo BRASIL-SR, que foi adaptado para as condições brasileiras pelo LABREN/CCST/INPE e a Universidade Federal de Santa Catarina – UFSC (Pereira et al., 1996), sendo o principal recurso deste grupo a quantificação do recurso energético solar a partir de estimativas de irradiação solar incidente na superfície terrestre. Ele combina a aproximação de dois fluxos na solução da equação de transferência radiativa com o uso de parâmetros determinados de forma estatística a partir de imagens de satélite. Neste trabalho, foram utilizadas 665 imagens do mês de abril de 2016, com resolução temporal de 30 minutos e resolução espacial de 0.03° em longitude e 0.05° de latitude, o que corresponde aproximadamente a 3km x 5km do ponto nadir do satélite.

O código está escrito em linguagem Fortran 90 e foi separado em 7 módulos a fim de facilitar qualquer tipo de aperfeiçoamento ou alteração de algum cálculo do modelo. No módulo 1 é calculada a transmitância em céu claro, no módulo 2 a transmitância em céu parcialmente nublado, no módulo 3 é realizado o cálculo das Irradiações Global, Difusa e Direta, no módulo 4 da Irradiação Par, no módulo 5 a integral diária da irradiação Par, no módulo 6 a integral diária das Irradiações Global, Direta, Difusa e Plano Inclinado e no módulo 7 as médias mensais das irradiações Global, Direta, Difusa, Par e Plano Inclinado. Este modelo está sendo executado no supercomputador Santos Dumont, localizado em Petrópolis, RJ, no Laboratório Nacional de Computação Científica – LNCC. Para a compilação dos códigos do modelo é necessário carregar as variáveis de ambiente e a inclusão no path dos programas e bibliotecas de forma modular utilizando o aplicativo “module”. Foi utilizada a biblioteca Netcdf 4.4.1 e o compilador Intel ifort 16.0.2 em ambiente computacional RedHat Linux 6.4 (LNCC, 2017).

Este supercomputador conta com processadores Intel Xeon com 12 núcleos por unidade de processamento e cada nó computacional têm duas unidades de processamento tendo assim um total de 24 núcleos por nó. Há um total de 12.096 núcleos multi-core, não sendo contabilizados os nós computacionais com os processadores Intel Xeon Phi e GPU Nvidia. A capacidade de armazenamento principal é de 1,7 Pbytes, com sistema

secundário de 640 Tbytes, e sistema operacional Linux (LNCC, 2018). Para a paralelização está sendo utilizado o Open Multi-Processing (OpenMP), que é uma interface de programação multi-processada de memória compartilhada em múltiplas plataformas.

O modelo, em sua forma original, realiza suas operações de I/O em formato binário. Também está sendo avaliada a implementação destas operações a partir de arquivos em formato NetCDF (Network Common Data Form) que é uma especificação desenvolvida pela fundação UNIDATA e oferecida sob a forma de bibliotecas utilitárias, visando o armazenamento de grandes volumes de dados. Além disso, um dos objetivos do formato é o acesso eficiente a pequenos subconjuntos de grandes conjuntos de dados, fazendo uso do acesso direto em vez de acesso sequencial, sendo muito mais eficiente quando a ordem que os dados são lidos é diferente da ordem que foram gravados (UNIDATA, 2018).

3. Resultados Preliminares e Discussões

Após utilizar a ferramenta GPROF, foi observado que as funções LINGAU e MATBAU, que são responsáveis por calcular a irradiação difusa para cada intervalo espectral, são as que mais consomem tempo de processamento, 41% do tempo total da execução sequencial. Foi iniciado o processo de inserção das diretivas de OpenMP com o propósito de diminuir o tempo de execução do modelo. Inicialmente, nos dois primeiros módulos, foi utilizado a diretiva *!#omp parallel do private(I,J)* no loop responsável por chamar as funções custosas. Internamente ao loop, existia uma chamada de leitura de arquivo, onde foi inserida a diretiva *#!omp critical* para que aquela região fosse executada com apenas um thread por vez. Visto que a região estava ficando serial com a diretiva, o seu conteúdo foi alocado em um vetor e passado por parâmetro pelas funções. Nos 4 módulos restantes, foi inserido a diretiva *!#omp parallel do private(I,J)* no loop responsável pelos calculos de incidencia de radiação e das médias mensais e diárias.

Abaixo (veja tabela 1), podemos observar que o tempo de processamento melhorou à medida que a quantidade de threads foi aumentando no modelo original.

Tabela 1. Tempo (segundos), Speedup e Eficiência do Modelo

	1T	6T	12T	18T	24T
Tempo Total	96637	20540	12811	10036	8863
Speedup	1	4,70	7,54	9,63	10,90
Eficiência	1	0,78	0,62	0,53	0,45

Com a versão do modelo contendo os dados em NetCDF, o tempo total de processamento diminuiu ainda mais. Abaixo (veja tabela 2), podemos observar a melhora do desempenho e a eficiência, até 12 threads, se mantendo constante.

Tabela 2. Tempo (segundos), Speedup e Eficiência do Modelo

	1T	6T	12T	18T	24T
Tempo Total	90700	19469	9808	6935	5703
Speedup	1	4,66	9,25	13,08	15,90
Eficiência	1	0,78	0,77	0,73	0,66

Em comparação com o modelo original, podemos observar que o speedup e a eficiência no modelo alterado se mostraram melhores que na versão original do modelo.

Com os testes realizados, podemos afirmar que o modelo com I/O em formato NetCDF e utilizando 24 threads tem uma melhora de 37% no tempo, e melhora de 41% comparado com o código original. O tempo de processamento de I/O sem NetCDF é de 4444 segundos e com NetCDF é de 2728 segundos, mostrando uma melhora em 61% no tempo de I/O. O tempo de I/O em relação ao tempo de processamento sem Netcdf representa 50% do tempo e o processamento com I/O com NetCDF representa 48% do tempo. O speedup e eficiência também melhoraram com essa alteração. A possível resposta para esta melhora é que os arquivos em Netcdf tem acesso direto podendo ser recuperados diretamente sem passar por outros dados na sequência, sendo importante quando a execução do modelo é realizada utilizando com mais de um thread. Além das imagens de satélite e dos dados climatológicos de entrada, os dados de saída somam 4100 imagens.

4. Conclusão

A paralelização do modelo BRASIL-SR, utilizando OpenMP, mostrou uma redução satisfatória no seu tempo de processamento com 24 threads, posteriormente será utilizado mais threads afim de avaliar melhor o desempenho e eficiência do modelo. Além disso, a implementação relativa às operações de I/O em formato NetCDF mostrou uma melhora adicional, com uma melhor eficiência paralela. Esses testes iniciais são parte do trabalho que está sendo realizado para a redução do tempo de processamento do modelo, e a próxima implementação será a exploração de outras técnicas para otimização deste tempo de processamento, tais como blocagem e vetorização.

5. Referências

- Laboratório Nacional de Computação Científica (LNCC), “Manual Supercomputador Santos Dumont”, Disponível: http://sdumont.lncc.br/support_manual.php?pg=support# Acesso: Janeiro/2018
- Pereira, E.B., Martins, F.R, Gonçalves, A.R., Costa, R.S., Lima, J.L., Rütther, R., Abreu, S.L., Tiepolo, G.M., Pereira, S.V. and Souza, J.G., “Atlas Brasileiro de Energia Solar”, 2.ed. São José dos Campos: INPE, 2017. 88p. ISBN 978-85-17-00090-4, 2017
- Pereira, E.B., Abreu, S.L., Stuhlmann, R., Rieland, M. and Colle, S. “Survey of The Incident Solar Radiation In Brazil by use of Meteosat Satellite Data”, Solar Energy, Phoenix, v. 57, n.2, p.125-132, 1996.
- Stuhlmann, R., Rieland, M. and Raschke, E. “An Improvement of the IGMK model to derive total and difuse solar radiation at the surface from satellite data”, J. Applied Meteorology, v29, n. 7, p.586-603, 1990.
- UNIDATA, “NetCDF 4.6.1”, Disponível: https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_introduction.html Acesso: Janeiro/2018