

Mineração e Visualização Remota de Dados Coletados de uma Rede de Sensores Ambientais

Gerson da Penha Neto
Alessandra de França Santana
Nandamudi Lankalapalli Vijaykumar
Rafael Duarte Coelho dos Santos

Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{gerson.neto, Alessandra.santana, vijay, rafael.santos}@lac.inpe.br

Abstract. The atmosphere is in constant interaction with the terrestrial biosphere, forming a coupled system. This interaction plays a fundamental role in the climate system and hydrological and biogeochemical cycles, through the exchange of energy of mass (eg water and carbon) between the terrestrial vegetation and the atmospheric boundary layer. Aiming to understand and anticipate these changes and their influence on the climate system, the main focus of many studies of surface-atmosphere interaction is to quantify the surface fluxes on various terrestrial biomes. With this motivation, this paper presents a software tool developed to test a prototype of environmental sensors in an experimental area of rainforest in the Amazon.

Palavras-chave: sensoriamento remoto, mineração de dados, meteorologia.

1. Introdução

A atmosfera está em constante interação com a biosfera terrestre, formando um sistema acoplado. Esta interação exerce um papel fundamental no sistema climático e nos ciclos biogeoquímicos e hidrológico, através das trocas de energia e massa (por exemplo, água e carbono) entre a vegetação terrestre e a camada limite atmosférica. Com o objetivo de se compreender e prever estas trocas e sua influência no sistema climático, o foco principal de inúmeros estudos de interação superfície-atmosfera é quantificar os fluxos de superfície sobre vários biomas terrestres. Desde meados da década de 90, o principal método utilizado para se medir os fluxos superficiais tem sido o método de covariância de vórtices turbulentos (“eddy covariance method”), que fornece medições diretas e com alta resolução temporal das trocas integrais dos ecossistemas na escala do dossel, Dabbert et al, (1993).

Medições deste tipo vem sendo realizadas em mais de 400 torres micrometeorológicas em todo o planeta ,Baldocchi (2003, 2008), e aprimoraram significativamente o conhecimento científico sobre a interação entre os ecossistemas vegetais e as condições ambientais. Porém, estes experimentos também trouxeram à tona importantes limitações da aplicação desta técnica sobre superfícies consideradas ‘complexas’, como áreas de floresta tropical e terrenos heterogêneos.

A base do método de covariância de vórtices é quantificar as trocas entre a superfície e a atmosfera através de medidas de variações no vento vertical e nas variáveis escalares (p. ex. temperatura, umidade ou concentração de CO₂) em uma torre que se estende acima da vegetação. A principal limitação de aplicação deste método em terrenos complexos é devido ao fato de que as medidas em uma torre única só podem estimar as variações das variáveis na direção vertical e variações temporais. Para aplicar o método, é necessário assumir que o produto médio das flutuações temporais do vento e dos escalares medidos em um ponto é equivalente à média espacial dos vórtices turbulentos que são ‘carregados’ com o escoamento médio do vento. Assume-se também que o campo de escoamento é horizontalmente homogêneo de forma que divergências horizontais de fluxo e advecção são negligenciadas. Porém, sobre superfícies naturais são raros os terrenos que podem ser considerados homogêneos e fatores como variações espaciais na topografia ou cobertura vegetal podem

afetar significativamente o escoamento e aumentar as incertezas na interpretação das medições de fluxos.

Estes fatores representam um grande obstáculo para o monitoramento do balanço regional de carbono nos biomas terrestres, pois afetam diretamente as estimativas de emissões pela respiração noturna da vegetação, Aubinet (2008). Um estudo experimental realizado por Araujo et al. (2008) mediu as concentrações de CO₂ ao longo de perfis verticais e ao longo de uma toposequência em uma área de floresta tropical ao norte de Manaus, AM. O resultado mais crítico mostrou a influência da topografia típica da região (sequências de platôs e vales): grande parte do CO₂ emitido pela respiração do solo e da vegetação escoou do platô durante a noite e acumula nos vales, onde fica ‘aprisionado’ pela forte estabilidade da camada de ar abaixo do dossel até o meio da manhã seguinte. Uma avaliação quantitativa da perda por drenagem noturna, porém, só é possível realizando medições detalhadas dos gradientes horizontais de vento e concentração de CO₂, Tota et al (2008).

Os estudos mostram claramente que dificilmente será possível compreender o escoamento e reduzir as incertezas de medidas de fluxo em terrenos complexos sem uma metodologia que reconheça a complexidade da variabilidade espacial do campo de variáveis ambientais. Com esta motivação, o presente trabalho apresenta uma ferramenta de *software* desenvolvida para testar um protótipo de sensores ambientais (geossensores) em uma área experimental de floresta tropical na Amazônia. A idéia advém de um projeto financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo em parceria com a *Microsoft Research*. Será montando uma rede de sensores de alta resolução espacial para capturar dados de interesse geral de floresta tropical na Amazônia, esses dados serão armazenados, processados e posteriormente divulgados para a comunidade acadêmica geral, além de serem utilizados como dados reais de monitoramento dos fluxos emitidos pela floresta tropical na Amazônia.

2. Metodologia de Trabalho

Os dados coletados pela rede de geossensores são centralizados, processados, armazenados e distribuídos para os usuários interessados, de forma que também possam ser usados como fonte de dados para outras aplicações. Para atingir este objetivo fora desenvolvido um sistema de gerenciamento destes dados capaz de suportar as necessidades especificadas. Este sistema possui as seguintes funções e/ou subsistemas:

- Entrada de dados obtidos dos sensores, armazenamento no banco de dados e criação dos metadados sobre a entrada.
- Relatórios simples, visualização simples dos dados coletados;
- Web services essenciais para acesso ao banco de dados independente dos relatórios por consultas personalizadas de usuários remotos.

Aproveitando os conhecimentos já adquiridos em visitas técnicas ao Departamento de Física e Astronomia da Universidade *Johns Hopkins* e a colaboração desta equipe, e visando reduzir o esforço de implementação, fora implementado um modelo de dados baseado no já utilizado no projeto *Life Under Your Feet*, no qual a equipe do projeto usa a tecnologia Microsoft (*SQL Server 2008* como banco de dados, *Server 2008* como servidor de aplicação e *C#* como linguagem de desenvolvimento). Além dessas ferramentas, também está em uso a ferramenta *Silverlight* como *front-end* para aplicações de visualização de dados. A figura 2.1 mostra um exemplo do modelo de banco de dados utilizado na construção da ferramenta de visualização; o modelo utilizado é constituído de dados reais coletados de plataformas de monitoramento espalhadas por diversas regiões do Brasil. A premissa é utilizar bancos já existentes para modelagem e fazer um assim um modelo adaptável para vários tipos de dados

e/ou medidas, uma vez que não existe uma unicidade na maneira com os dados são coletados por sensores diferentes.

hora_pcd	satelite	estacao	procod	id_pcd	local	latitude	longitude	hora_sensor	id_sensor	nome_sensor	valor_bruto	valor_calibrado	unidade
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	193	UmidInt#4	1261656000	0	0.000000	%
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	194	UmidRel#1	1261634400	16	22	%
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	194	UmidRel#1	1261645200	16	22	%
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	194	UmidRel#1	1261656000	1D	29	%
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	205	VelVento10m#4	1261634400	26	3.800000	m/s
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	205	VelVento10m#4	1261645200	13	1.900000	m/s
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	205	VelVento10m#4	1261656000	48	7.200000	m/s
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	216	VelVentoMax#1	1261656000	A1	16.100000	m/s
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	225	RadSolAcum#7	1261623600	0	0.000000	MJ/m2
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	225	RadSolAcum#7	1261634400	0	0.000000	MJ/m2
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	225	RadSolAcum#7	1261645200	0	0.000000	MJ/m2
1261656737	130	45	1	31929	Honoropolis	-19 30 0	50 0 0	225	RadSolAcum#7	1261656000	23	3.500000	MJ/m2
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	18	Bateria#8	1261656000	81	12.900000	Volt
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	29	CodifExo	1261638000	0	0.000000	cm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	29	CodifExo	1261641600	0	0.000000	cm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	29	CodifExo	1261645200	0	0.000000	cm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	29	CodifExo	1261648800	0	0.000000	cm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	29	CodifExo	1261652400	0	0.000000	cm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	29	CodifExo	1261656000	0	0.000000	cm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	36	ComPSol#3	1261656000	8C	140	Logico
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	71	LgcTsol#1	1261656000	0	0	Logico
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	76	NivRegua#1	1261656000	7E	126	cm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	99	Pluvio#5	1261638000	E8	232	mm
1261656738	130	45	1	32748	Rio do Sul Novo	-27 11 0	49 37 0	99	Pluvio#5	1261641600	E8	232	mm

Figura 2.1 Exemplo de banco de dados utilizado na construção da ferramenta de mineração.

A partir dos modelos de banco de dados e de dados reais, iniciou-se o processo de desenvolvimento de serviços web e de protótipos de mineração e visualização de dados. Os protótipos ainda são modestos visto que ainda faltam definições a serem feitas acerca de como os dados devem ser visualizados ou que tipo de informações a comunidade necessita sobre esses dados ou ainda que tipos de tratamento esses dados possam vir a receber, isso acontece devido as variações de interesse e tipos de dados coletados. Uma vez que os dados foram coletados e armazenado inicia-se a etapa de processamento dos dados, que consiste em uma avaliação e tratamento dos dados por meios de técnicas de mineração de dados já conhecidas na literatura como Mapas Auto-Organizáveis de Kohonen, Redes Neurais de múltiplas camadas, Algoritmos de agrupamento hierárquico, Carl G (1997). Como resultado após a fase de tratamento e mineração tem-se um banco de dados sólido com menos ruídos como falta de dados ou dados errados (*science database*), além de se ter conhecimentos prévios de grupos ou padrões e também previsões sobre os dados coletados, na figura 2.2 é apresentada a página índice da ferramenta de visualização desenvolvida.

Como ponto de partida focou-se primeiro na visualização de dados, pois esse é o primeiro serviço web e é o mínimo que se espera para que algum membro da comunidade acadêmica possa ler os dados e então utilizar-los para algum fim de interesse acadêmico. Inicialmente fora criado serviços no servidor web utilizando a tecnologia *Silverlight* para visualização. Na página índice, os serviços colocados a disposição do pesquisador buscam dados de temperatura e umidade relativa coletados de sensores na Mata Atlântica, esse é mais um dos modelos de banco de dados usados na construção da ferramenta. Por meios de botões o usuário pode selecionar quais informações ele deseja e em seguida visualizar as informações em um gráfico com aparência estilizada e intuitiva, além disso, é permitido o controle da pesquisa pelo usuário, ou seja, o usuário pode selecionar um tipo de dado específico ou um período específico e obter assim a informação necessária com precisão, isso é importante no sentido de evitar um gasto desnecessário de processamento por parte do servidor e também agilizar o tempo que a informação é disposta na tela para consumo.

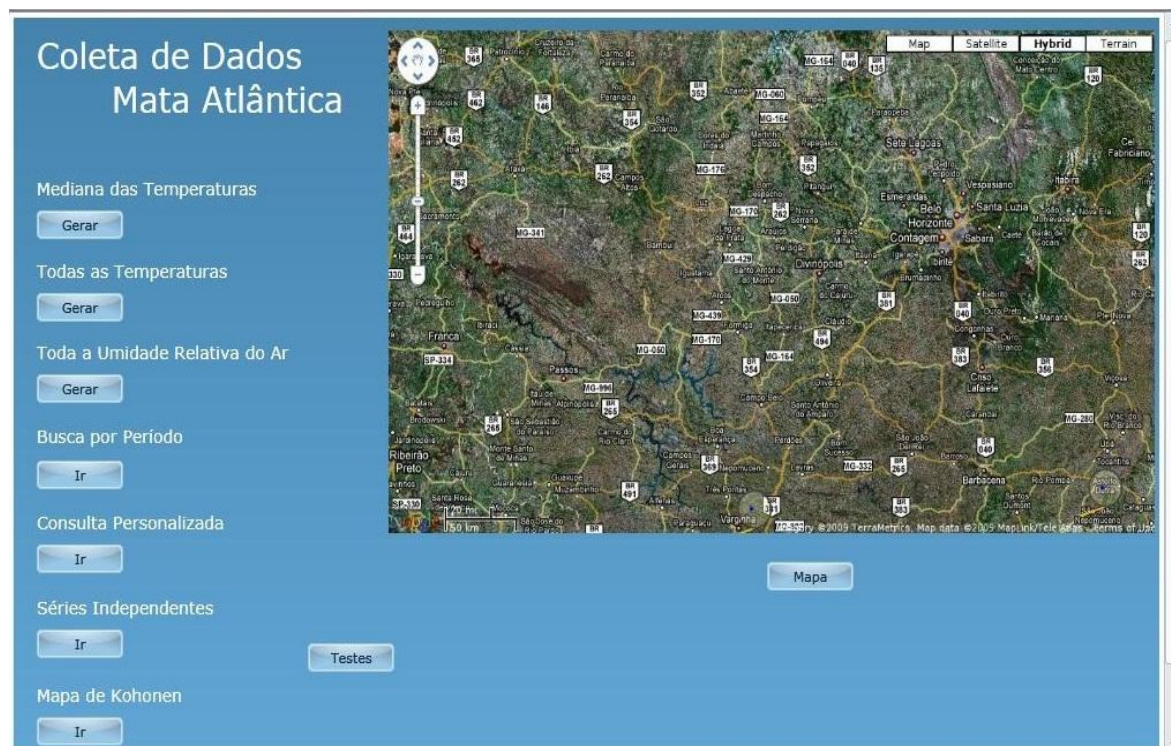


Figura 2.2 Página índice da ferramenta de visualização e mineração.

A página índice é amigável no sentido de ser fácil o acesso aos dados e as pesquisas, no canto superior direito da página é exibido um mapa no qual é possível ver de qual região do Brasil os dados foram retirados. Através dos botões é possível acessar outras páginas que contêm mais conteúdos para visualização. O mapa do canto superior direito é também uma ferramenta da *Microsoft* que foi acoplada a ferramenta de visualização, essa ferramenta é denominada de *Bing Maps* e está disponível na internet para que usuários interessados possam utilizá-la.

Como dito anteriormente outros tipos de consultas personalizadas também pode ser feitas, para isso foi criado uma aplicação em especial que tem por finalidade facilitar o uso da linguagem SQL (*Structured Query Language*), que é a linguagem amplamente usada para inserir, buscar, selecionar, apagar, copiar e manipular dados e tabelas em um banco de dados. Muitos usuários acadêmicos, principalmente os que não possuem afinidade com a linguagem SQL e com ferramentas computacionais como os sistemas gerenciadores de banco de dados (SGBD), têm dificuldades em pesquisar ou manipular dados utilizando SQL, portanto surgiu a idéia de se fazer uma aplicação que possa ajudar na construção de sentenças de pesquisa para que usuários não familiarizados com a linguagem possam fazer suas consultas e também aprender mais de como usar SQL. A figura 2.3 mostra a aplicação especial criada para dar suporte para quem não tem afinidade com a linguagem SQL. Nesta página o usuário pode montar consultas utilizando botões discriminados pelo tipo de comando SQL que geram, desse modo por meio de 'clics' o pesquisador pode formar sentenças capazes de extrair dados e gerar gráficos. Com o aprimoramento no conhecimento da linguagem será possível ignorar os botões e os comandos pré-compilados que eles representam e fazer a mão consultas próprias, para isso existe um quadro que serve tanto para mostrar as consultas feitas acionando os botões como para que usuários mais avançados escrevam em SQL as sentenças que desejarem e obtenham assim as informações desejadas do banco de dados.

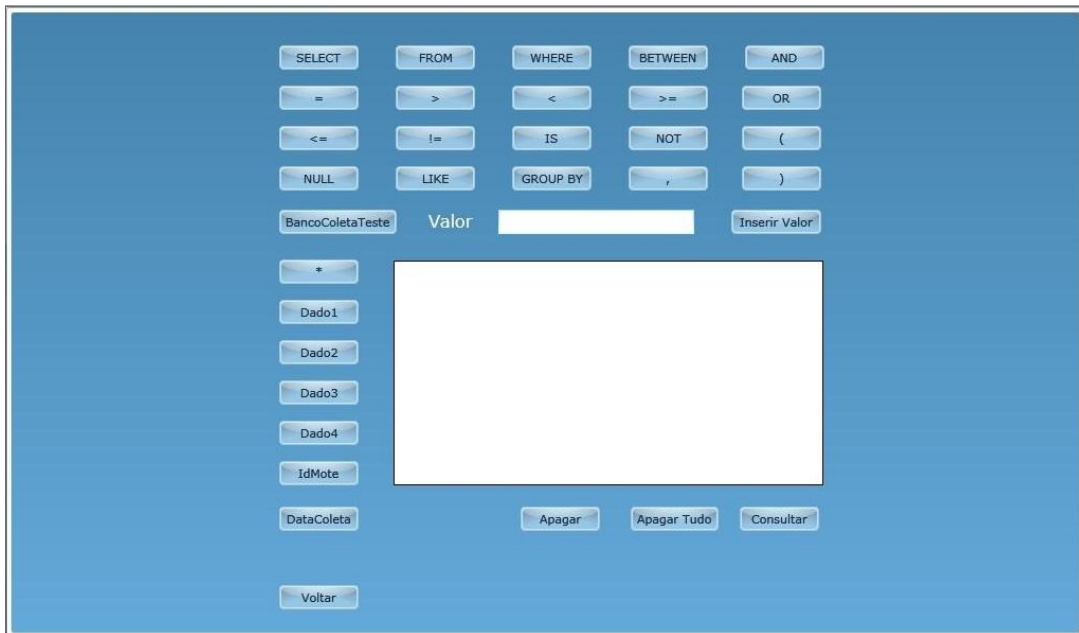


Figura 2.3 Aplicação para auxílio no desenvolvimento de sentenças SQL.

3. Resultados

Como resultado se obtém uma ferramenta capaz de armazenar, processar e distribuir dados de maneira rápida e eficaz. Na figura 3.1 é possível ver gráficos que mostram medidas de temperaturas para uma determinada região da Mata Atlântica e na figura 3.2 é apresentado gráficos de curvas de temperatura e umidade para a mesma região permitindo assim uma prévia avaliação visual das distribuições temporais de temperatura e umidade.

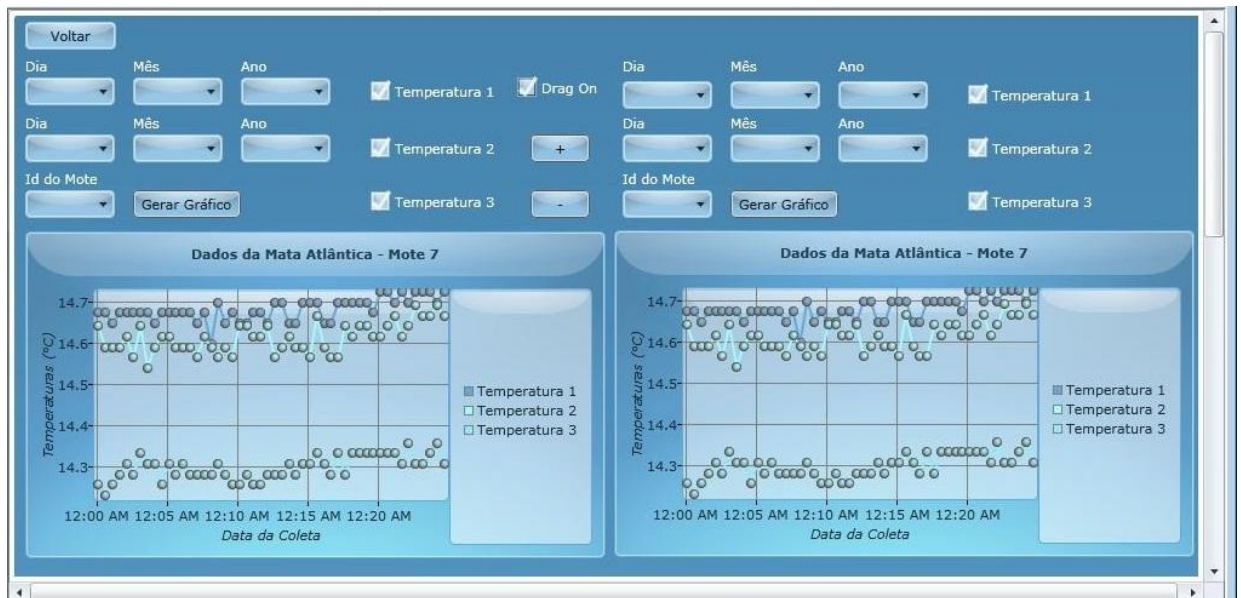


Figura 3.1 Gráficos de temperatura para Mata Atlântica.



Figura 3.2 Gráficos de curva de temperatura e umidade.

4. Conclusões

Ainda não se tem uma medida ou estimativa da influência do tipo de pesquisa realizada com o tempo de processamento dos dados para visualização, contudo é natural que dependendo da consulta ao banco, isto é, se a consulta busca informação de um grande volume de dados ou de dados dispostos e mais de uma tabela, o tempo de processamento tende a ser maior que de costume, mas o fato de se poder personalizar a consulta ajuda bastante no sentido de agilizar essa busca e diminuir o processamento. Com relação às técnicas de mineração, enriquecimento e redução de dados, foi feito um estudo utilizando livros e artigos científicos da área de ciência da computação para se ter um levantamento de quais técnicas podem ser usadas para extrair informações sobre os dados de sensores, entretanto talvez não seja possível disponibilizar todas as técnicas de mineração nas ferramenta de visualização devido ao peso computacional de algumas, por exemplo, técnicas como: *Dynamic Time Warping* (DTW), algoritmos de agrupamento hierárquico ou ainda métodos de agrupamento como *Nearest Neighbour* (NN), podem consumir muito recurso computacional dependendo da quantidade de dados utilizados, Lin et al, (2008), Han (2006).

5. Referências

- Araujo, A.C., Kruijt, B., Nobre, A. D., Dolman, A. J., Waterloo, M. J., Moors, E. J., Souza, J. S., 2008. Nocturnal Accumulation of CO₂ underneath a tropical forest canopy along a topographical gradient. *Ecological Applications*, 18(6): 1406-1419, doi:10.1890/06-0982.1.
- Aubinet, M., 2008. Eddy covariance CO₂ flux measurements in nocturnal conditions: An analysis of the problem. *Ecological Applications*, 18(6): 1368-1378, doi:10.1890/06-1336.1.
- Baldocchi, D., 2008. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*, 56: 1-26.

Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, 9(4): 479-492.

Carl G. Looney. *Pattern Recognition Using Neural Networks*. Oxford University Press, 1st edition, 1997.

Dabbert, W.F., Lenschow, D. H., Horst, T. W., Zimmenman, P. R., Oncley, S. P., Delany, A. C., 1993. Atmosphere-surface exchange measurements. *Science*, 260: 1472-1481.

Han, Jiawei. *Data Mining Concepts and Techniques*. Second edition. University of Illinois at Urbana-Champaign: Elsevier, 2006. 726. 13:978-1-55860-901-3.

Lin, Jessica. Keogh, Eamonn & Gunopulos, Dimitrios. Iterative Incremental Clustering of Time Series. Computer Science & Engineering Department, University Of California, Riverside. International Conference on Knowledge Discovery and Data Mining. July - 2008 Edmonton, Alberta, Canada. pp 102-111.

Tota, J., Fitzjarrald, D. R., Staebler, R. M., Sakai, R. K., Moraes, O. L. L., Acevedo, O. C., Wofsy, S. C., Manzi, A. O., 2008. Amazon rain forest subcanopy flow and the carbon budget: Santarém LBA-ECO site. *Journal of Geophysical Research*, 113: G00B02, doi:10.1029/2007JG000597.