



USO DE SISTEMA DE PRÉ-PROCESSADORES PARA OBTENÇÃO DE SÉRIES PLUVIOMÉTRICAS DE QUALIDADE

RAÍZA S. PRECINOTO^{1,5}, GERALDO R. L. FILHO², TAMÍRES P. CORREIA¹,
EDNALDO O. SANTOS³, GUSTAVO B. LYRA³, SÉRGIO M. S. DA CRUZ⁴

¹ Discente em Engenharia Florestal, Bolsista de Iniciação Científica PROIC, Instituto de Florestas, UFRRJ, Seropédica – RJ
⁵ raizasalomao@gmail.com

² Discente em Sistemas de Informação, Bolsista FAPERJ, Instituto de Ciências Exatas, UFRRJ, Seropédica - RJ

³ Meteorologista, Prof. Adjunto, Depto de Ciências Ambientais, Instituto de Florestas, UFRRJ, Seropédica - RJ

⁴ Químico, Prof. Adjunto, Programa de Pós-graduação em Modelagem Matemática e Computacional, UFRRJ, Seropédica - RJ

Apresentado no XVIII Congresso Brasileiro de Agrometeorologia – 02 a 06 de Setembro de 2013 – Centro de Convenções e Eventos Benedito Silva Nunes, Universidade Federal do Pará, Belém, PA

RESUMO: Séries históricas de precipitação são importantes para diversos estudos ambientais aplicados. Porém, diversos fatores afetam a qualidade dessas séries de dados, necessitando que sejam assimilados, tratados, além de passar por um controle de qualidade e preenchimento de falhas. Neste trabalho foi utilizado o sistema baseado em Pré-processadores para tratar dados de precipitação de 23 estações pluviométricas da Região Metropolitana do Rio de Janeiro. Ao todo, foram removidos 103 valores acima da faixa limite (500 mm), com média de cinco *outliers* por estação. Além disso, foram sinalizados 207 valores como suspeitos, com uma média de nove alertas por estação analisada. Finalmente, foi preenchido, pelo método de regressão linear, um total de 505 dados pluviométricos, representando uma média de 22 preenchimentos por estação. O sistema de Pré-processadores mostrou ser capaz de facilitar estudos de dados climáticos, graças à rapidez do processamento e completude das análises.

PALAVRAS-CHAVE: Gestão de Dados Meteorológicos, Qualidade de Dados.

USE OF PRE-PROCESSORS SYSTEMS TO GET RAINFALL SERIES WITH QUALITY

ABSTRACT: Historical series of rainfall are important for many environmental studies. However, several factors affect the quality of the series and that is why such data should be treated with the aid of quality control and gap filling techniques. In this study we used a system based on a set of pre-processors to handle data from 23 rainfall stations from the metropolitan region of Rio de Janeiro. Were removed 103 values above the physical limit range (500 mm) with five outliers per station. In addition, 207 values were tagged as suspicious, with an average of nine alerts per analyzed station. A total of 505 rainfall data





were filled by the linear regression method, such amount represent an average of 22 fills per station. The system of pre-processors used in this work is already available to help researcher to perform climate studies.

KEYWORDS: Meteorological Management of Data, Data Quality.

INTRODUÇÃO

Os estudos meteorológicos são fundamentais para se compreender o meio-ambiente de forma integral, e servem de base para diversas aplicações. Para tal, é preciso incluir as informações disponibilizadas pelos órgãos responsáveis pelo setor, como a Agência Nacional de Águas (ANA) e o Instituto Nacional de Meteorologia (INMET) e que os dados sejam confiáveis. Por isso, os mesmos devem ser assimilados e tratados, buscando séries homogêneas e de qualidade.

O tratamento dos dados compreende a obtenção de séries meteorológicas contínuas, sem falhas e sem dados espúrios, ou seja, consistentes e homogêneos. A maior parte das séries de dados climáticos de longo prazo contêm variações devido à fatores não climáticos como o deslocamento do local da estação, urbanização, etc. Esses fatores não climáticos muitas vezes criam heterogeneidades de várias magnitudes e em diferentes posições no tempo. Usar séries de dados que não foram ajustados para essas variações não climáticas pode afetar seriamente a correta avaliação de tendências e variabilidades climáticas, além de interferir com a identificação de qualquer sinal real de mudança climática. Por conseguinte, torna-se importante desenvolver e melhorar as técnicas para identificá-las adequadamente e ajustá-las para variações não climáticas (Vincent, 1998).

Em vista do exposto acima, o objetivo deste trabalho é assimilar e tratar dados provenientes de estações pluviométricas de diversas cidades da Região Metropolitana do Rio de Janeiro usando sistemas de pré-processadores criados para realizar o controle de qualidade e preencher as falhas de séries históricas.

MATERIAIS E MÉTODOS

Foram obtidos dados de chuva diária de 23 estações pluviométricas que abrangem a região metropolitana do estado do Rio de Janeiro, oriundos do banco de dados da ANA, através da plataforma *Hidroweb* (<http://hidroweb.ana.gov.br>). Foram selecionadas as estações cujas séries tivessem um mínimo de 20 anos de dados, e a partir do ano de 1960. A Figura 1 mostra a distribuição das 23 estações em questão no Estado do Rio de Janeiro.

As séries de dados são obtidas sob a forma de arquivo texto estruturado que foram incorporadas ao sistema de Pré-processadores, previamente desenvolvido pelo grupo de pesquisa, denominado *Meteoro* (Filho *et al*, 2013).

Neste sistema, é realizada a carga e transformação de dados e seu posterior tratamento através de módulos que não só avaliam a qualidade dos dados, como também registram os descritores



de proveniência (Freire *et al.*, 2008) das transformações aplicadas aos dados meteorológicos, com a indicação de dados suspeitos ou improváveis, e posteriormente o preenchimento de falhas das séries, através do método de regressão linear, de acordo com o raio de proximidade entre as estações meteorológicas.

O controle de qualidade dos dados é realizado com base na identificação dos limites físicos e climáticos e na comparação desses com os totais de chuva de estações na mesma região geográfica. Os limites climáticos foram obtidos nas normais do INMET para a região de estudo e na análise exploratória das séries das demais estações avaliadas com o uso do gráfico do tipo *box-plot*, que permitiu identificar os *outliers* das séries, os quais foram posteriormente avaliados em relação às estações na mesma mesorregião. As comparações foram realizadas por modelos de regressão linear, de modo que, se o valor se afastou até 100% da tendência, o mesmo foi considerado inacurado e excluído da série.

Para o preenchimento de falhas foram utilizadas estações dentro de um raio de 75 km, contanto que tivessem um mínimo de 80% dos dados contidos no mesmo intervalo da estação a serem preenchidas as falhas, preenchendo-se todas as possíveis. O preenchimento de falhas do Pré-processador é realizado tanto para dados em branco como para dados acima da faixa do limite físico (500 mm).

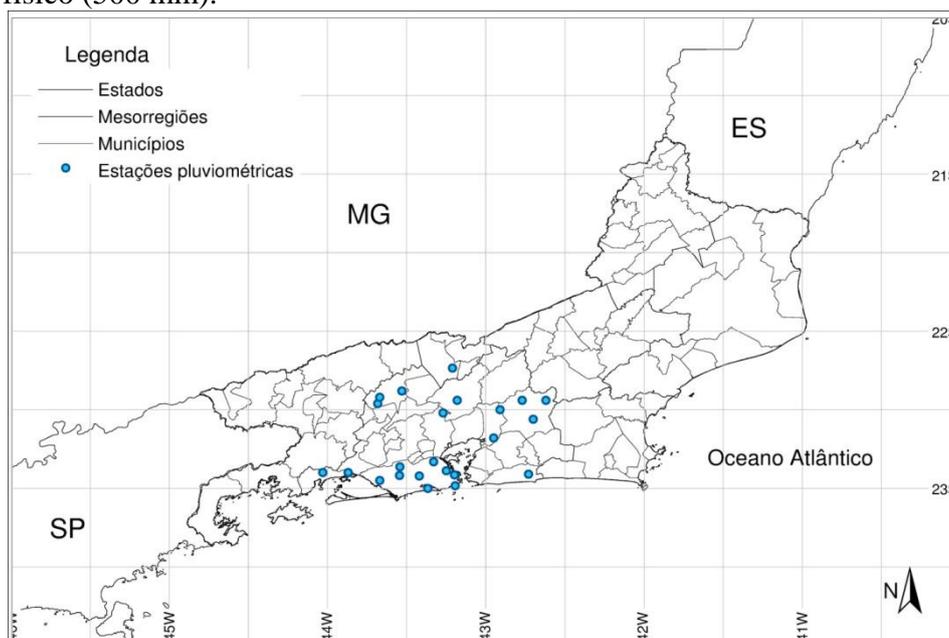


Figura 1 – Mapa da localização das 23 estações, contidas na Região Metropolitana do Rio de Janeiro.

RESULTADOS E DISCUSSÃO

Após a carga das séries das 23 estações pluviométricas no sistema, observou-se um percentual médio de meses com falhas ou dados em branco, em relação à quantidade de meses efetivos, de 36,4%, sendo que este parâmetro teve uma amplitude de 0% (Eletrobrás) a 214,5% (Piraí). Para realizar esta análise sem o pré-processador, seriam necessários aproximadamente dois meses de trabalho, mesmo usando programas de planilhas eletrônicas como Microsoft Excel (semi-automatizado). As análises foram automatizadas através do sistema *Meteoro*, um



XVIII Congresso Brasileiro de Agrometeorologia – XVIII CBA
2013 e VII Reunião Latino Americana de Agrometeorologia
Belém - PA, Brasil, 02 a 06 de Setembro 2013
Cenários de Mudanças Climáticas e a Sustentabilidade
Socioambiental e do Agronegócio na Amazônia



exemplo desse tipo de processamento pode ser observado na Figura 2. A Tabela 1 exibe os resultados obtidos do sistema de Pré-processamento de dados.



The screenshot shows a web interface for the 'Meteoro' system. On the left is a navigation menu with options: 'Cadastrar estação', 'Carga de dados', 'Qualidade de dados', 'Visualizar dados', 'Preenchimento de falhas', 'Resumo dos dados', and 'Sobre'. The main area displays a table of data under the heading 'Proveniência'. The table has columns for 'Data', 'Máxima', 'Dia Máxima', 'numDiasChuva', 'total', 'totalStatus', 'created', and 'Valor previsto'. The data rows represent monthly precipitation data from 1970-01-01 to 1971-03-01.

Proveniência							
Data	Máxima	Dia Máxima	numDiasChuva	total	totalStatus	created	Valor previsto
1970-01-01	0	0	0	-9999.99	0	2013-06-27 11:10:31	123
1970-02-01	0	0	0	-9999.99	0	2013-06-27 11:10:33	33
1970-03-01	0	0	0	-9999.99	0	2013-06-27 11:10:33	60
1970-04-01	0	0	0	-9999.99	0	2013-06-27 11:10:34	56
1970-05-01	0	0	0	-9999.99	0	2013-06-27 11:10:35	36
1970-06-01	0	0	0	-9999.99	0	2013-06-27 11:10:35	47
1970-07-01	30	15	6	58.00	1	2013-06-21 00:57:45	-
1970-08-01	25	30	9	99.90	1	2013-06-21 00:57:46	-
1970-09-01	9	11	10	40.40	1	2013-06-21 00:57:48	-
1970-10-01	20	9	14	96.80	1	2013-06-21 00:57:50	-
1970-11-01	61	22	15	162.90	1	2013-06-21 00:57:52	-
1970-12-01	32	21	11	112.60	1	2013-06-21 00:57:54	-
1971-01-01	121	14	11	227.10	1	2013-06-21 00:57:56	-
1971-02-01	216	26	6	-9999.99	0	2013-06-27 11:10:36	77
1971-03-01	38	1	14	102.20	1	2013-06-21 00:57:59	-

Figura 2 – Tela do sistema *Meteoro* contendo a visualização de dados após o tratamento nos pré-processadores.

Tabela 1 – Resumo dos Dados de Precipitação tratados pelo Pré-Processador.





XVIII Congresso Brasileiro de Agrometeorologia – XVIII CBA
2013 e VII Reunião Latino Americana de Agrometeorologia
Belém - PA, Brasil, 02 a 06 de Setembro 2013
Cenários de Mudanças Climáticas e a Sustentabilidade
Socioambiental e do Agronegócio na Amazônia



	ID	Nome	Município	M Efetivos	F Dia	F Mes	Red.	Alert - Out	Rem - Out	Alert - Box	Alert - Reg	Pre - Reg
1	2243141	BANGU	Rio de Janeiro	383	12196	135	215	2	1	14	0	19
2	2243149	ALTO DA BOA VISTA	Rio de Janeiro	513	10025	84	513	20	9	6	0	0
3	2243151	VASSOURAS	Vassouras	596	13071	332	672	4	1	17	0	0
4	2243155	TINGUA	Duque de Caxias	223	16558	338	0	8	5	5	0	56
5	2243157	NUCLEO COLONIAL S. BENTO	Duque de Caxias	169	7306	142	0	2	0	6	0	24
6	2243165	PIRAI	Pirai	220	20824	472	0	3	1	6	0	65
7	2243186	ECOLOGIA AGRIC. (KM 47)	Itaguaí	319	22900	353	377	3	0	8	0	60
8	2243210	FREI CANECA (ETD)	Rio de Janeiro	239	2126	122	300	0	0	5	0	53
9	2243232	ELEVATÓRIA DE COPACABANA	Rio de Janeiro	220	281	10	252	2	1	5	0	11
10	2243250	COROA GRANDE	Itaguaí	256	963	33	176	6	3	11	0	27
11	2244148	IBICUI	Mangaratiba	261	2437	76	188	5	1	5	0	0
12	2343007	VIA 11 - 9DR	Rio de Janeiro	270	1137	18	300	1	0	10	0	11
13	2243097	MENDANHA - 6DR	Rio de Janeiro	278	990	18	300	3	1	13	0	18
14	2243092	BENFICA - CCPL	Rio de Janeiro	300	741	21	936	0	2	11	0	8
15	2242010	MANUEL RIBEIRO	Rio de Janeiro	527	1419	48	451	2	1	12	0	21
16	2242011	ESTAÇÃO DE BOMB. DE IMUNANA	Magé	530	1050	28	448	2	1	8	0	12
17	2242012	REPRESA DO PARAISO	Magé	531	488	16	1443	36	38	7	0	47
18	2242013	FAZENDA DO CARMO	Cachoeiras de Macacu	529	1329	45	455	17	6	11	0	0
19	2242014	JAPUIBA	Cachoeiras de Macacu	535	1153	41	456	10	1	7	0	10
20	2242016	FAZENDA SAO JOAQUIM	Cachoeiras de Macacu	530	1484	46	1448	41	28	5	0	42
21	2243083	ELETROBRAS	Rio de Janeiro	295	513	0	312	0	1	10	0	4
22	2243086	IRAJA - 3DR	Rio de Janeiro	330	1159	18	360	2	0	11	0	11
23	2243089	CAMPO GRANDE - 7DR	Rio de Janeiro	340	440	11	1116	0	2	14	0	6





“M Efetivos”: Meses de dados fornecidos pelo Hidroweb; “F Dia”: Número de dias em branco ou faltando no arquivo dentro do intervalo da série histórica; “F Mês”: Número de meses em branco ou faltando; Red.: Número de linhas/datas (meses) que estão representados duas ou mais vezes em um mesmo arquivo; “Alert Out”: Dados mensais acima da faixa de alerta (400mm) e abaixo da faixa limite (500mm); “Rem Out”: Dados mensais acima da faixa limite (500 mm); “Alert Box”: Dados espúrios detectados pelo BoxPlots; “Alert Reg”: Dados espúrios detectados pela regressão linear; “Pre Reg”: Dados preenchidos pela regressão linear.

Através do processamento dos dados, controle de qualidade e preenchimento de falhas foram removidos 103 valores acima da faixa limite (500 mm), com média de cinco *outliers* por estação. Além disso, foram sinalizados 207 valores como suspeitos (entre a faixa de alerta, que é 400 mm, e a faixa limite), com uma média de nove alertas por estação analisada.

Por conseguinte, foi preenchido pelo método de regressão linear um total de 505 dados pluviométricos, representando uma média de 22 preenchimentos por estação.

CONSIDERAÇÕES FINAIS

Quanto maior o nível de automação de uma análise, menos sujeita a falhas humanas a mesma estará. Contudo, para isso é necessário um sistema de processamento eficiente, contemplando as exigências do tratamento de dados meteorológicos, usando as recomendações estabelecidas pela Organização Mundial de Meteorologia (OMM).

O sistema de pré-processadores utilizado neste trabalho já é capaz de automatizar e facilitar estudos climáticos, através da rapidez das análises, sendo possível observar o número de meses fornecidos pelo *Hidroweb*, a quantidade de dias e meses em branco ou faltando em cada série histórica, número de datas repetidas no mesmo arquivo, entre outros parâmetros. Além disso, são sinalizados os dados acima da faixa limite de 500 mm, e os suspeitos, que são aqueles acima da faixa de alerta, 400 mm, e menores que a faixa limite, sendo possível substituir estes valores, além de preencher os ausentes.

Contudo, ressalta-se que o Sistema aqui apresentado está em constante aperfeiçoamento, e serão implementadas novas rotinas para a inclusão de outros parâmetros meteorológicos importantes, principalmente temperatura do ar.

REFERÊNCIAS BIBLIOGRÁFICAS

- Filho, G. R. L., Precinoto, R. S., Correia, T. P. *et al.* **Assimilação, Controle de Qualidade e Análise de Dados Meteorológicos Apoiados por Proveniência.** XXXIII Congresso da Sociedade Brasileira de Computação, 2013.
- Freire, J., Koop, D., Santos, E. & Silva, C. T. **Provenance for Computational Tasks: A Survey.** Computing in Science and Engineering, v. 10, n. 3, p. 11-21. 2008.
- Vincent, L. A. **A Technique for the Identification of Inhomogeneities in Canadian Temperature Series.** Journal of Climate, American Meteorological Society, Volume 11, pages 1094-1104, 1998.

AGRADECIMENTOS

Agradecemos à FAPERJ pelo financiamento APQ1 e uma bolsa de iniciação científica e ao programa PROIC/UFRRJ por outras duas bolsas de iniciação científica.

