Utilização de Analítica Visual para complementação da análise estatística sobre dados coletados pelo Sistema Integrado de Monitoramento Ambiental - SIMA

Alisson Fernando Coelho do Carmo[1], **Milton Hirokazu Shimabukuro**[1], **Enner Herenio de Alcântara**[1] **José Luiz Stech**[2], **Vilma Mayumi Tachibana**[1]

[1] Programa de Pós-Graduação em Ciências Cartográficas (PPGCC)
Faculdade de Ciências e Tecnologias (FCT)
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)
Presidente Prudente – SP – Brasil

[2]Divisão de Sensoriamento Remoto (DSR)
Instituto Nacional de Pesquisas Espaciais (INPE)
São José dos Campos – SP – Brasil

alisondocarmo@gmail.com, {miltonhs, enner, vilma}@fct.unesp.br
stech@dsr.inpe.br

Abstract. *The development of sensor devices increased the potential for environmental monitoring enabling the automatic and periodic data collection. Interactive data exploration and visual analysis techniques can be used to support the analytical process of the data set. This paper presents some possibilities of using Visual Analytics techniques to complement the use of other methods of exploratory data analysis applied to environmental data. Some techniques were used, as Paralell Coordinates and Horizon Chart to support the interactive exploration of environmental data collected by the Integrated System for Environmental Monitoring and complement the results of descriptive statistical analysis and Principal Components Analysis.*

Resumo. *A evolução dos dispositivos sensores aumentou as possibilidades para o monitoramento ambiental permitindo a aquisição constante de dados. Técnicas de exploração e análise visual e interativa de dados podem ser aplicadas sobre os dados para auxiliar o processo analítico. Este trabalho apresenta algumas possibilidades de utilização de técnicas de Analítica Visual para complementar outros métodos de análise exploratória de dados aplicadas sobre dados ambientais. Foram utilizadas técnicas como Paralell Coordinates e Horizon Chart, para auxiliar a exploração interativa dos dados ambientais coletados pelo Sistema Integrado de Monitoramento Ambiental e complementar os resultados de análises estatísticas descritivas e por Componentes Principais.*

1. Introdução

A intervenção causada pela ação do homem no meio ambiente tem grande influência sobre o comportamento normal deste contexto. A necessidade do registro e acompanhamento de observações das características ambientais é cada vez maior para que seja possível interpretar as mudanças ocorridas e controlá-las ativamente. As características de ambientes

aquáticos podem oferecer um grande conjunto de variáveis representativas que permitem a avaliação do ambiente.

Atualmente, em razão do desenvolvimento da área de sensoriamento remoto, as imagens produzidas por satélites orbitais e fotografias geradas por outros dispositivos imageadores aéreos, têm se consolidado como recursos importantes para a obtenção de dados referentes às características da superfície. Os sensores remotos oferecem uma alternativa para a análise, que agregada às metodologias de coletas clássicas de levantamentos e amostragem locais (*in situ*) representam as principais abordagens de obtenção de dados.

A utilização de sensoriamento remoto e a coleta local de dados apresentam aspectos que podem interferir em seu potencial de análise. [STECH et al. 2011] afirmam que para o total entendimento dos processos físicos, químicos e biológicos que agem sobre ambientes aquáticos é necessária a manipulação de séries temporais tão grandes quanto possíveis e com vários atributos meteorológicos e limnológicos. Para permitir a identificação de determinados fenômenos que acontecem rapidamente, é necessário uma alta resolução temporal de dados, ou seja, coletas efetuadas constantemente. Já para analisar fenômenos mais lentos, é necessário um longo registro histórico que permita acompanhar sua ação gradativa.

Considerando estas premissas, a coleta de dados por amostragem local com uma grande periodicidade se torna altamente custosa, podendo inviabilizar sua execução. A utilização de tecnologias de sensoriamento remoto pode se tornar limitada para monitoramento de recursos e qualidade da água devido a resolução temporal, espacial e espectral de alguns sensores, como citado por [Ritchie et al. 2003], que concluem que a integração entre diferentes tecnologias, como dados de sensoriamento remoto, GPS (*Global Positioning System*) e SIG (Sistema de Informação Geográfica) é imprescindível.

Em razão do desenvolvimento tecnológico de dispositivos sensores, processamento e comunicação, uma abordagem que está sendo crescentemente utilizada é a coleta automática e periódicas de dados. O Sistema Integrado de Monitoramento Ambiental (SIMA) é formado por um conjunto de tecnologias aplicadas à coleta de dados e monitoramento da hidrosfera [INPE 2014]. O sistema SIMA é composto de uma rede de plataformas que possuem sensores que coletam dados do ar e da água. As plataformas SIMA realizam a leitura dos sinais dos sensores em um período de uma hora. Após a leitura, os dados coletados são transmitidos via enlace de satélite para servidores intermediários em estações terrestres, que são responsáveis por receber os dados e realizar a verificação da existência de erros na transmissão dos sinais. Com os dados identificados como consistentes, estes são finalmente transmitidos ao servidor no centro de armazenamento, os quais passam pelo processo de decodificação, processamento e armazenamento, tornando-se disponíveis em um portal da internet. [Alcântara et al. 2013] apresentam algumas estatísticas básicas sobre as plataformas e discutem sobre os principais problemas encontrados, relacionados à degradação dos sensores e comunicação com o satélite.

Embora a existência de grande quantidade de dados e variáveis seja de suma importância para a análise de dados, a interpretação destes dados pode requerer processamentos mais robustos. Uma das formas de potencializar o processo de análise é a utilização de técnicas de *Visual Analytics* (Analítica Visual) com recursos visuais que ofereçam interatividade para a exploração do conjunto de dados. [Keim et al. 2008] sa-

lientam a importância da colaboração entre homem e máquina no processo analítico e sintetizam as tarefas envolvidas em um ciclo de execução chamado *sense-making loop*. [Ward et al. 2010] ressaltam os benefício que podem ser conseguidos utilizando recursos que permita a interação do usuário com os dados. Tais técnicas podem ser utilizadas para complementar outras abordagens de análise, pois facilitam a identificação de comportamentos dos valores dos dados e aproveitam a capacidade de percepção visual humana.

Neste sentido, este trabalho tem o objetivo de apresentar algumas técnicas de Analítica Visual, como Coordenadas Paralelas (*Paralell Coordinates*) e *Horizon Charts*, que podem ser utilizadas em conjunto com outros métodos de análise de dados. Foram aplicadas técnicas de análise estatística descritiva do conjunto de dados e, visando investigar possíveis relações entre as diferentes variáveis coletadas para produzir um conjunto reduzido de variáveis sem relacionamento direto, foi utilizada a técnica de estatística multivariada de Análise de Componentes Principais (*Principal Component Analysis* - PCA). De forma geral, segundo [Johnson 1998], PCA utiliza algumas combinações lineares entre variáveis originais para a criação de um novo conjunto de variáveis capazes de explicar a estrutura de variância e covariância do conjunto, com o objetivo de reduzir a dimensão dos dados e facilitar a interpretação. Esta técnica é comum na literatura em diversos contextos, como avaliação da qualidade da água [Guedes et al. 2012], classificação de elementos químicos [da Silva Lyra et al. 2010], entre outros. Assim como proposto por [Jeong et al. 2009b], as técnicas de Analítica Visual podem facilitar o entendimento das composições dos Componentes Principais, uma vez que o processo de obtenção destes fatores pode ser obscuro para aqueles que não o dominam.

O processo de análise foi aplicado à um conjunto de dados coletados por uma plataforma SIMA, referentes ao reservatório de Itaipu. No final do processo, foi possível identificar a relação direta entre algumas variáveis, bem como a sumarização em apenas 3 fatores capazes de representar grande parte da variância do conjunto total de variáveis. Algumas das características ressaltadas pela análise estatística puderam ser notadas utilizando representações visuais e interativas dos dados, mostrando-se útil para complementar a análise.

As demais Seções descrevem os principais aspectos envolvidos no desenvolvimento deste trabalho. A exploração e seleção inicial dos dados de análise é descrita na Seção 2. Os resultados dos processamentos, técnicas de Analítica Visual e aplicação da técnica PCA são apresentados na Seção 3 para então, subsidiar as considerações finais apresentadas na Seção 4.

2. Exploração e seleção dos dados

Neste trabalho foram utilizados os dados ambientais coletados pelo projeto SIMA. Os dados foram adquiridos por meio do portal do projeto SIMA (<*www.dsr.inpe.br/hidrosfera/sima/login.php*>), no qual permite a obtenção do conjunto de dados desejado em formato de planilha eletrônica. Uma vez adquiridos os dados de todas as plataformas SIMA, estes foram inseridos em um Sistema Gerenciador de Banco de Dados (SGBD) para facilitar a manipulação e filtragem dos dados a serem processados. O SGBD utilizado foi o PostgreSQL, escolha motivada por ser um sistema *open source* que possui integração com a extensão espacial PostGIS, também *open source*.

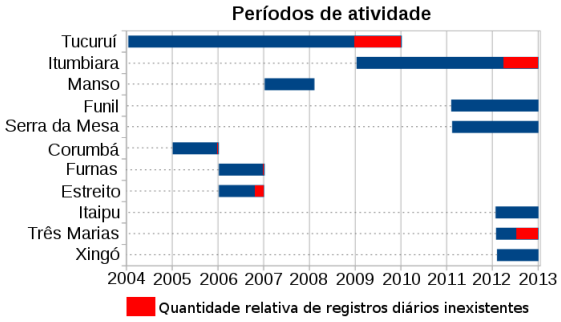

Figura 1. Intervalos de tempo ativos das plataformas e quantidade relativa de dias sem coleta de dados

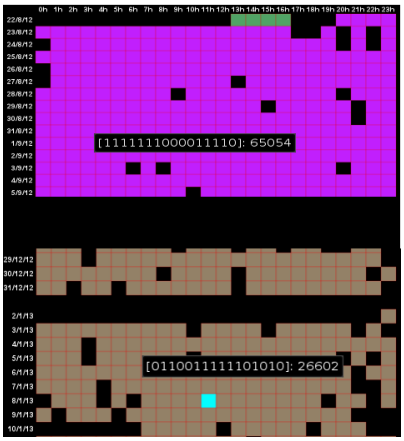

Figura 2. Identificação de erros em conjuntos de sensores.[do Carmo et al. 2013]

O primeiro fator observado, com o início da exploração dos dados, foi o tempo de atividade de cada plataforma, que demonstra o período em que a plataforma SIMA esteve ativa, coletando e transmitindo dados, como pode ser visto na Figura 1, bem como a existência de dias completos sem a realização de nenhuma coleta de dados.

Conforme discutido por [Alcântara et al. 2013], existem diversas causas para a existência dos erros nos dados, como degradação dos sensores, falha na transmissão dos valores coletados e erro na conversão dos sinais elétricos em componente digital. [do Carmo et al. 2013] sugerem a utilização de uma visualização baseada em *Calendar View* para exibir uma máscara binária, em escala de cores, capaz de identificar a existência de erros e ilustram o caso de um intervalo de dados da plataforma SIMA do reservatório de Três Marias, conforme Figura 2, na qual é possível notar a existência de dois padrões de cores que representam dois diferentes conjunto de sensores com erros.

Os erros nos dados podem ocorrer constantemente em diferentes intervalos da série temporal das plataformas SIMA. Como o foco deste trabalho está na investigação entre as possíveis relações entre as variáveis capturadas pelas plataformas SIMA, as lacunas de dados podem interferir negativamente durante a análise e cálculo dos componentes principais, pois influenciam diretamente na variância dos dados. Observando o gráfico apresentado na Figura 1 é possível identificar a plataforma SIMA do reservatório de Itaipu com inexistência de dias completos sem dados. Após análise exploratória da base de dados da plataforma SIMA do reservatório de Itaipu e utilizando a visualização apresentada na Figura 2, foi possível selecionar um período em que todas as variáveis foram coletadas, compreendido entre 20 de julho de 2012 à 30 de agosto de 2012, contendo 887 registros. Uma representação geral do conjunto de dados selecionados, utilizando a técnica *point-based*, é apresentada na Figura 3, na qual permite observar a variabilidade dos dados e possíveis padrões de alterações por meio da observação da cor que representa cada registro -azul claro para valores baixos e azul escuro para valores mais altos. Nesta representação os dados são dispostos ordenadamente de acrodo com o tempo de coleta e os valores de cada atributos são convertidos para o espaço de cores.

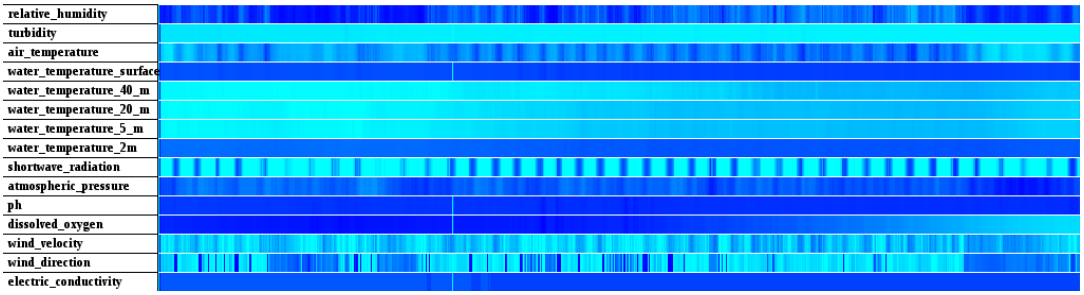

Figura 3. Representação visual geral dos dados selecionados

3. Análise Estatística

Inicialmente, para verificar o comportamento geral das diferentes variáveis, algumas métricas de estatísticas descritivas foram calculadas sobre o conjunto de dados. O conjunto de dados descrito pelas métricas estatísticas é apresentado na captura de tela exibida na Figura 4.

Variable	N	Mean	StDev	Variance	Minimum	Median	Maximum	Range
condutividade	887	0,043337	0,002369	0,000006	0,040000	0,045000	0,050000	0,010000
O2_dissolvido	887	7,2191	2,7411	7,5137	1,0800	8,8100	10,3200	9,2400
pH	887	7,3497	0,1646	0,0271	7,0000	7,3600	8,1800	1,1800
agua_temp_0m	887	21,231	1,057	1,118	19,220	21,660	23,520	4,300
agua_temp_3.5m	887	21,031	1,056	1,116	19,240	21,440	22,470	3,230
agua_temp_5m	887	20,992	1,065	1,134	19,110	21,310	22,340	3,230
agua_temp_10m	887	20,864	1,100	1,211	19,110	21,180	22,210	3,100
agua_temp_30m	887	20,221	1,163	1,354	18,720	19,880	22,080	3,360
turbidez	887	8,4992	2,3959	5,7405	5,8700	7,8300	17,6100	11,7400
vento_dir	887	107,89	103,07	10623,59	0,00	57,66	358,59	358,59
vento_vel	887	3,4463	2,1852	4,7753	0,0000	3,1400	9,8000	9,8000
pressao_atm	887	990,24	3,05	9,29	983,53	989,65	998,82	15,29
radiacao_oc	887	165,50	242,03	58577,95	0,00	5,88	852,94	852,94
ar_temp	887	21,074	3,824	14,623	12,010	21,170	30,450	18,440
humidade_relativa	887	70,860	16,329	266,644	25,100	70,590	99,220	74,120

Figura 4. Métricas da estatística descritiva sobre o conjunto de dados

As métricas de estatísticas descritivas demonstram uma grande similaridade entre os dados de temperatura da água, referentes aos atributos: agua_temp_0m, agua_temp_3.5m, agua_temp_5m, agua_temp_10m, agua_temp_30m. O comportamento homogêneo dos das da temperatura da água pode ser observado também considerando a visualização dos valores da série temporal, Figura 5, e a matriz de gráficos de dispersão, Figura 6, relacionados a estes dados.

Além dos métodos gráficos convencionais de representação, exibidos nas Figuras 5 e 6, outras técnicas de Visualização de Informação podem facilitar a observação de relação entre a variabilidade dos dados, como a técnica baseada em *Horizon Charts*. A técnica *Horizon Chart* busca aumentar a eficiência na utilização do espaço em representações de séries temporais [Heer et al. 2009], além de possibilitar a apresentação simultânea de vários atributos, aspecto relevante na manipulação de séries temporais multivariadas. Os dados sobre a temperatura da água são apresentados na Figura 7 utilizando a técnica *Horizon Chart*.

Na representação visual apresentada na Figura 7, os dados estão sendo exibidos de forma relativa à temperatura 20.86º, média dos valores de temperatura da água do conjunto selecionado. As áreas em azul e verde representam a variação positiva e negativa, respectivamente, do valor em relação à temperatura média. Já a intensidade da cor representa a amplitude da variação. A visualização *Horizon Charts* na Figura 7 comple-

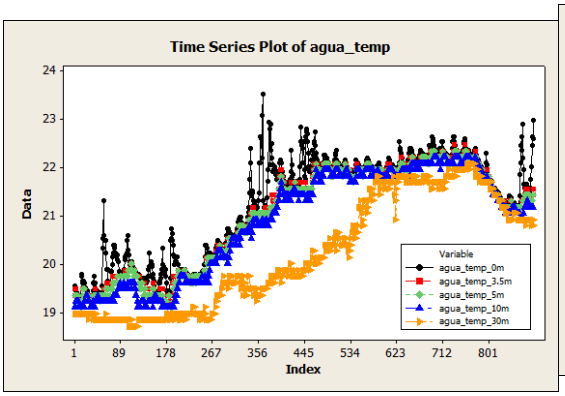

Figura 5. Série temporal dos dados de temperatura da água

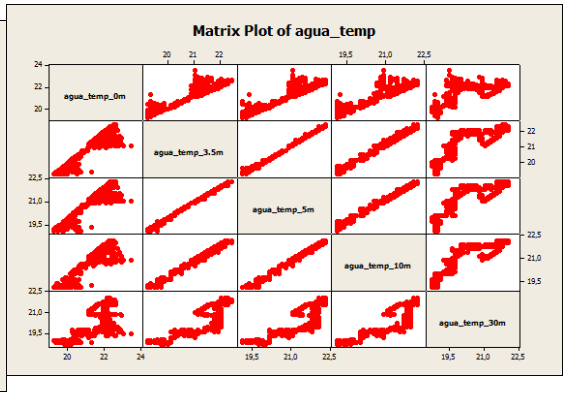

Figura 6. Matriz de gráficos de dispersão dos dados de temperatura da água

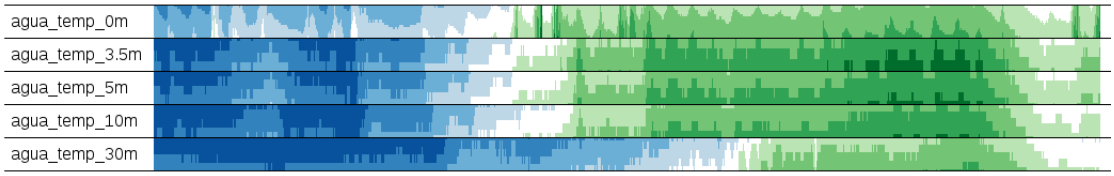

Figura 7. Visualização *Horizon Chart* sobre os dados de temperatura da água

menta as informações de existência de relação entre os dados de temperatura da água em diferentes níveis de profundidade, indicados também nos gráficos das Figuras 5 e 6.

Para evitar o destaque da evidente relação entre os valores da temperatura da água em diferentes níveis de profundidade e facilitar a identificação de possíveis relações entre outras variáveis, estes atributos foram sumarizados em uma única variável. A técnica PCA foi aplicada sobre os dados de temperatura da água, sumarizando em um único fator capaz de explicar mais de 94% da variação dos dados de temperatura, como pode ser observado na Figura 8 que exibe os autovalores e autovetores relevantes. Considerando as contribuições homogêneas do primeiro Componente Principal com peso distribuído entre todas as cinco variáveis, foi criada uma nova variável para representar a temperatura da água obtida com PCA.

Autovalor	4,7302	0,2052	0,0571	0,0058	0,0017
Explicação	0,946	0,041	0,011	0,001	0,000
Acumulado	0,946	0,987	0,998	1,000	1,000

Componentes Principais

Variáveis	PC1	PC2	PC3	PC4
agua_temp_0m	0,442	-0,452	-0,769	-0,094
agua_temp_3.5m	0,457	-0,159	0,267	0,535
agua_temp_5m	0,457	-0,133	0,331	0,306
agua_temp_10m	0,457	-0,070	0,394	-0,781
agua_temp_30m	0,422	0,865	-0,270	0,033

Figura 8. Componentes Principais dos dados da temperatura da água

Após o processamento inicial e formatação dos dados, o conjunto de dados foi submetido à técnica PCA. Para isto, foi realizado o cálculo da matriz de correlação para ser utilizada no cálculo de autovalores e autovetores. A matriz de correlação, que pode ser vista na Figura 9, foi utilizada em razão da não padronização das unidades métricas das variáveis.

A matriz de correlação demonstra comportamentos que podem ser úteis para o processo de análise e exploração dos dados. A informação de relação entre as variações

condutividade	1,00000										
O2_dissolvido	-0,56247	1,00000									
pH	0,69572	-0,48641	1,00000								
pca_agua_temp	0,90233	-0,64459	0,68550	1,00000							
turbidez	-0,83436	0,64616	-0,70744	-0,87945	1,00000						
vento_dir	-0,08491	-0,09208	0,09111	-0,13875	0,05325	1,00000					
vento_vel	0,17378	-0,28885	-0,06792	0,26052	-0,20911	0,05571	1,00000				
pressao_atm	0,34151	-0,61517	0,29791	0,32453	-0,45484	0,25167	0,21971	1,00000			
radiacao_oc	0,19308	-0,06200	0,21790	0,14822	-0,12543	0,07301	0,34284	0,09310	1,00000		
ar_temp	0,29465	0,18717	0,20438	0,36824	-0,22195	-0,37793	0,11972	-0,46290	0,37883	1,00000	
humidade_relativa	-0,45497	0,19650	-0,21609	-0,55314	0,38728	0,46629	-0,26849	0,16845	-0,33711	-0,77440	1,00000

Figura 9. Matriz de correlação. As maiores correlações são destacadas

dos dados pode ser útil para o refinamento de técnicas de visualização que consideram a proximidade do relacionamento entre as variáveis, como a técnica *Paralell Coordinates* (Coordenadas Paralelas). A técnica de Coordenadas Paralelas foi introduzida por [Inselberg and Dimsdale 1990] como um recurso capaz de representar dados multidimensionais sem a necessidade de redução do espaço R^n, utilizando projeções para o espaço bi ou tridimensional. Um dos importantes fatos que estão relacionados com a facilidade de interpretação da visualização em Coordenadas Paralelas é a ordenação das variáveis de acordo com suas relações, informação que poderia ser extraída da matriz de correlação. A Figura 10 apresenta a visualização inicial do conjunto de dados em Coordenadas Paralelas. Esta visualização permite a reordenação dos eixos de forma interativa, bem como a seleção de determinados valores para serem destacados em todos os eixos. Tal representação ressalta alguma relações indicadas pela matriz de correlação, como a relação inversa entre condutividade e oxigênio dissolvido, relação direta entre temperatura da água e condutividade, relação inversa entre turbidez e temperatura da água, entre outras.

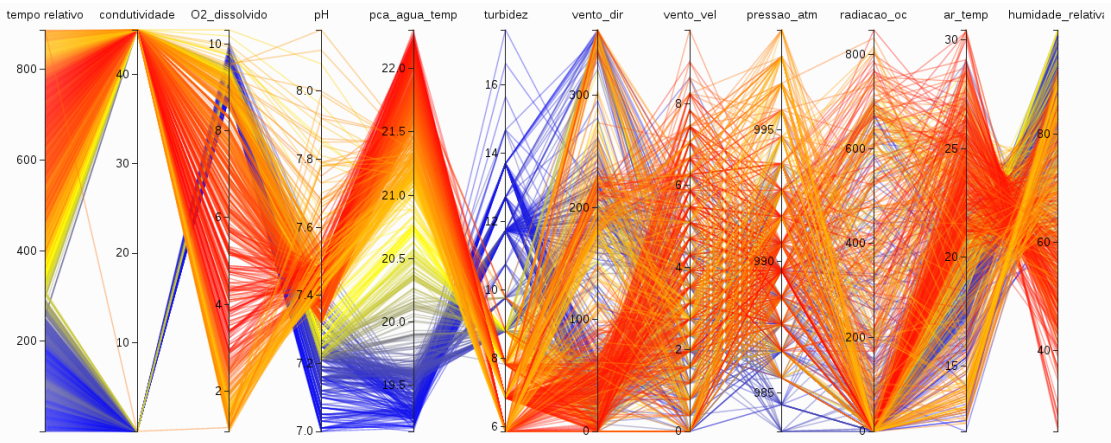

Figura 10. Representação em Coordenadas Paralelas dos dados selecionados

Após a obtenção da matriz de correlação, cálculo dos autovalores e respectivos autovetores, os Componentes Principais foram obtidos de acordo com sua respectiva proporção de explicação da variância dos dados. A Figura 11 exibe a síntese abordando os autovalores e Componentes Principais (*Principal Component* - PC) obtidos.

Existem diversas abordagens que buscam definir a maneira ótima de escolher a quantidade de PCs que representam o conjunto de dados de forma relevante, mas segundo [Johnson 1998], não existe uma resposta definitiva à esta circunstância. Poderiam ser escolhidos apenas três PCs, que seriam capazes de explicar 75.2% da variabilidade dos dados. A escolha dos três PCs mais relevantes permite segmentar a contribuição das

#Autovalores

Autovalor	4,5117	2,4278	1,3313	0,9841	0,5853	0,3527	0,2929	0,2141	0,1468	0,0948	0,0584
Explicação	0,410	0,221	0,121	0,089	0,053	0,032	0,027	0,019	0,013	0,009	0,005
Acumulado	0,410	0,631	0,752	0,841	0,895	0,927	0,953	0,973	0,986	0,995	1,000

#Componentes Principais

Variável	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
condutividade	0,428	0,016	-0,139	0,065	-0,093	-0,264	-0,167	0,593	-0,368	-0,032	0,447
O2_dissolvido	-0,334	-0,299	-0,040	0,261	-0,059	-0,671	-0,340	-0,118	-0,160	0,278	-0,209
pH	0,354	0,097	-0,222	0,451	0,080	-0,088	0,513	-0,454	-0,334	0,132	0,016
pca_agua_temp	0,448	-0,018	-0,120	-0,057	-0,185	-0,056	-0,062	0,170	-0,017	-0,242	-0,808
turbidez	-0,429	-0,092	0,133	0,012	0,088	0,231	0,120	0,129	-0,735	-0,366	-0,150
vento_dir	-0,052	0,386	0,278	0,544	-0,550	0,281	-0,297	-0,012	-0,015	0,059	-0,008
vento_vel	0,154	-0,016	0,665	-0,387	-0,385	-0,320	0,295	-0,156	-0,131	0,039	0,060
pressao_atm	0,201	0,464	0,163	-0,169	0,430	-0,126	-0,517	-0,383	-0,199	-0,187	0,019
radiacao_oc	0,150	-0,168	0,587	0,427	0,531	-0,002	0,091	0,293	0,177	0,043	-0,114
ar_temp	0,174	-0,551	0,034	0,192	-0,147	0,028	-0,187	-0,324	0,138	-0,620	0,250
humidade_relativa	-0,276	0,443	-0,075	0,164	-0,008	-0,468	0,299	0,146	0,281	-0,535	0,033

Figura 11. Componentes Principais escolhidos e variáveis mais representativas

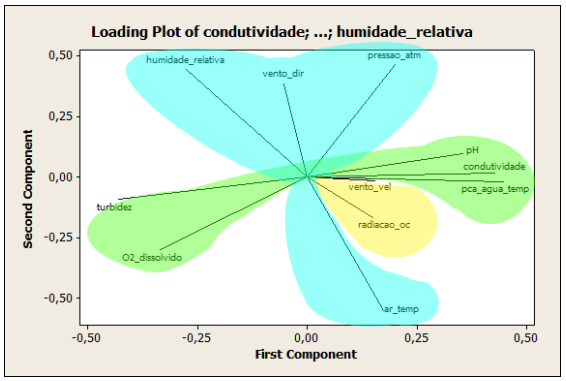

Figura 12. Relação de contribuição de cada variável nos PCs

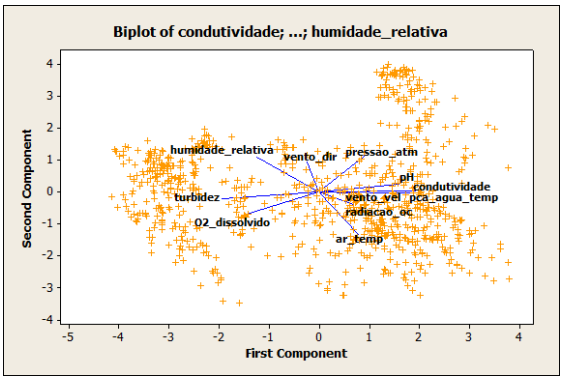

Figura 13. Gráfico integrado entre a carga dos fatores e o conjunto de dados

variáveis em fatores específicos, de forma que a contribuição ou peso de uma variável seja relevante em apenas um dos três PCs. Esta característica pode ser observada por meio do gráfico apresentado na Figura 12 no qual são destacados os agrupamentos relacionados aos respectivos PCs. Para ilustrar a relação com os dados, a Figura 13 integra a visão das cargas dos fatores (*Loadings*) com o conjunto de amostras dos dados (*Score*).

4. Conclusão

Este trabalho apresentou uma abordagem para redução do número de variáveis estudando suas relações utilizando a técnica de Análise de Componentes Principais. O resultado da aplicação desta técnica em um conjunto de dados da plataforma SIMA fundeada no reservatório de Itaipu permitiu a descrição da variância dos dados utilizando apenas três Componentes Principais, os quais representam mais de 75% da variância dos dados. A quantidade de fatores e porcentagem de explicação se apresenta em consonância com o trabalho de [Guedes et al. 2012] que também utilizou três PCs que explicam 74.3% da variância total no estudo da qualidade da água.

As representações visuais utilizadas demonstraram possibilidades para potencializar o processo de análise, sobretudo em relação a facilidade da percepção visual dos comportamentos dos dados e capacidade de interação. Este fato reforça a potencialidade da integração de técnicas de Anlítica Visual para ampliar o entendimento de outras

abordagens analíticas, assim como apontado por [Jeong et al. 2009a], que apresentam um estudo comparativo entre os resultados obtidos com a aplicação de análises estatísticas e sua integração com técnicas de Analítica Visual dispostas em um ambiente de Múltiplas Visões Coordenadas na plataforma interativa chamada IPCA. Baseados na acurácia, velocidade e qualidade das respostas dos usuários/analistas que foram submetidos aos testes, [Jeong et al. 2009a] concluem que a integração com as representações visuais interativas ajudam os usuários à entenderem melhor as relações entre os dados e as características estatísticas extraídas.

A técnica de *Horizon Chart* conseguiu exibir o comportamento relacionado em diferentes profundidades de temperatura complementando a informação apontada pelas estatísticas descritivas. O relacionamento entre o comportamento de todas as variáveis puderam ser visualizados utilizando a técnica de Coordenadas Paralelas, na qual todos os atributos dos conjuntos de dados são exibidos simultaneamente e pode ser potencializada com a reordenação dos eixos de acordo com o valor de correlação entre cada variável.

Além da sintetização das variáveis, os PCs gerados no presente trabalho permitiram o agrupamento de variáveis específicas a cada fator. Como pode ser visto na Figura 12, as variáveis que representam mais significativamente o primeiro PC estão relacionadas com os atributos da água, podendo ser nomeado como fatores de características da água ou então fatores limnológicos. Já o segundo componente principal é substancialmente formado por meio dos atributos relacionados ao ar, permitindo a atribuição de fator de características do ar, ou meteorológicos.

O terceiro componente principal merece atenção especial, pois é constituído basicamente por duas variáveis que se referem à velocidade do vento e a radiação de ondas curtas. Observando a correlação entre estas variáveis, é possível obter o valor de 0.34 que não representa um valor expressivo, porém é o maior índice de relação comparando estas variáveis com o restante do conjunto.

Por fim, espera-se que este trabalho possa contribuir com os estudos das relações entre as variáveis coletadas pelas plataformas SIMA, que podem ser complementadas e apoiadas com técnicas de representação visual de dados e exploração interativa presentes no contexto da Analítica Visual.

Agradecimentos

Os autores agradecem o Programa de Pós-Graduação em Ciências Cartográficas (PPGCC) da Faculdade de Ciências e Tecnologia/UNESP (FCT/UNESP)- Campus de Presidente Prudente - por permitir o desenvolvimento desta investigação; a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio dedicado ao projeto; e ao Instituto Nacional de Pesquisas Espaciais (INPE) pela cessão dos dados SIMA.

Referências

[Alcântara et al. 2013] Alcântara, E., Curtarelli, M., Ogashawara, I., Stech, J., and Souza, A. (2013). A system for environmental monitoring of hydroelectric reservoirs in brazil. *Ambiente e Água - An Interdisciplinary Journal of Applied Science*, 8(1).

[da Silva Lyra et al. 2010] da Silva Lyra, W., da Silva, E. C., de Araújo, M. C. U., FragosoI, W. D., and VerasI, G. (2010). Classificação periódica: um exemplo didático para ensinar análise de componentes principais. *Química Nova*, 33:1594 – 1597.

[do Carmo et al. 2013] do Carmo, A. F. C., Shimabukuro, M. H., and de Alcântara, E. H. (2013). Exploração visual interativa de dados coletados pelo sistema integrado de monitoramento ambiental - sima. In *XIV Brazilian Symposium on Geoinformatics*, GEOINFO 2013, pages 127 – 132.

[Guedes et al. 2012] Guedes, H. A. S., da Silva, D. D., Elesbon, A. A. A., Ribeiro, C. B. M., de Matos, A. T., and Soares, J. H. P. (2012). Aplicação da análise estatística multivariada no estudo da qualidade da água do rio pomba, mg. *Revista Brasileira de Engenharia AgrÃcola e Ambiental*, 16:558 – 563.

[Heer et al. 2009] Heer, J., Kong, N., and Agrawala, M. (2009). Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1303–1312, New York, NY, USA. ACM.

[INPE 2014] INPE, H. (2014). Sima: Sistema integrado de monitoramento ambiental. `http://www.dsr.inpe.br/hidrosfera/sima/`. Acessado em Março de 2014.

[Inselberg and Dimsdale 1990] Inselberg, A. and Dimsdale, B. (1990). Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st Conference on Visualization '90*, VIS '90, pages 361–378, Los Alamitos, CA, USA. IEEE Computer Society Press.

[Jeong et al. 2009a] Jeong, D. H., Ziemkiewicz, C., Fisher, B., Ribarsky, W., and Chang, R. (2009a). ipca: An interactive system for pca-based visual analytics. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'09, pages 767–774, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.

[Jeong et al. 2009b] Jeong, D. H., Ziemkiewicz, C., Ribarsky, W., and Chang, R. (2009b). Understanding principal component analysis using a visual analytics tool. *2009 US-Korea Conference on Science, Technology and Entrepreneurship*.

[Johnson 1998] Johnson, R. (1998). *Applied multivariate statistical analysis*. Prentice Hall, Upper Saddle River, NJ [u.a.], 4. ed edition.

[Keim et al. 2008] Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Jörn, and Melançon, G. (2008). Information visualization. chapter Visual Analytics: Definition, Process, and Challenges, pages 154–175. Springer-Verlag, Berlin, Heidelberg.

[Ritchie et al. 2003] Ritchie, J. C., Zimba, P. V., and Everitt, J. H. (2003). Remote sensing techniques to assess water quality. *Photogrammetric Engineering and Remote Sensing*, 69(6):695–704.

[STECH et al. 2011] STECH, J. L., ALCANTARA, E. H., LORENZZETTI, J. A., and LIMA, I. B. T. (2011). Uso de tecnologia espacial para coleta automática de dados limnologicos e meteorologicos: Aplicações nos reservatórios hidreletricos de manso e corumbá. In *Novas tecnologias para o monitoramento e estudo de reservatórios hidrelétricos e grandes lagos*, chapter 4, pages 119–162. Parêntese.

[Ward et al. 2010] Ward, M., Grinstein, G., and Keim, D. (2010). *Interactive Data Visualization: Foundations, Techiniques, and Applications*. A K Peters, Ltd.