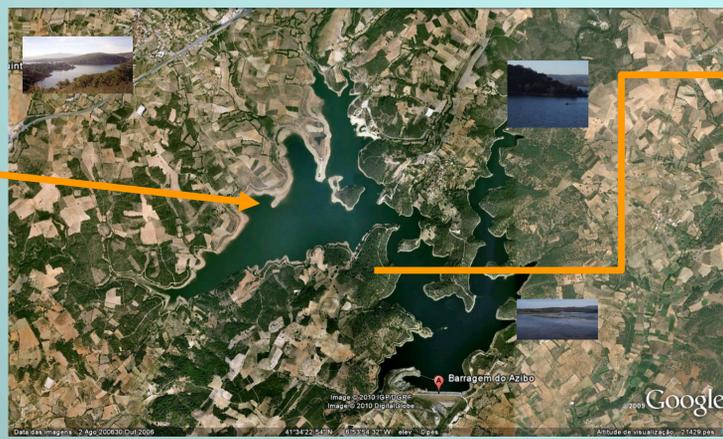


DATA MINING: UMA POTENCIAL FERRAMENTA PARA A ANÁLISE DE TENDÊNCIAS EM DADOS LIMNOLÓGICOS?

Ana M. Gerales & Pedro Bastos

CIMO, Escola Superior Agrária de Bragança, Campus de Santa Apolónia 5301-885 Bragança, Portugal (geraldes@ipb.pt; bastos@ipb.pt)



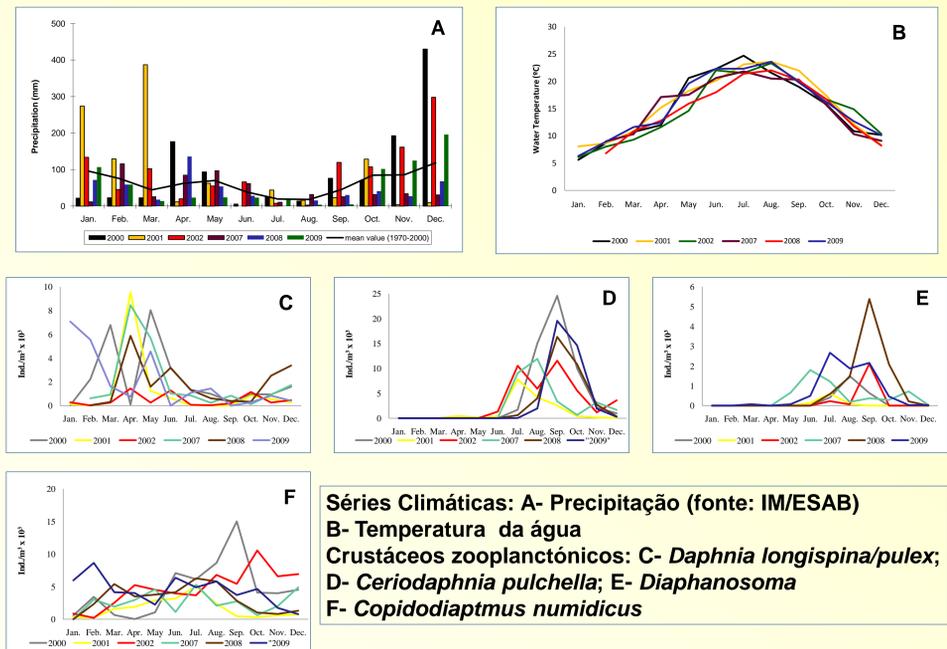
CARACTERÍSTICAS DA ALBUFEIRA

Capacidade total (10³m³): 54470
 Área da sub-bacia (km²): 89
 Área da albufeira (ha): 4,10
 Profundidade máxima (m): 30
 Profundidade média(m): 13,2
 Ano de entrada em funcionamento: 1982
 Principais utilizações: Recreio, Abastecimento urbano
 Altitude média (m): 500
 Variação do nível da água (m): 1,5 - 2,0
 Tempo de residência da água (anos): 2,22
 Ocupação do solo: Matos, matas autóctones, lameiros, culturas, infra-estruturas de recreio
 Estado trófico: Meso-eutrófica (TSI- Carlson, 1977)

Introdução

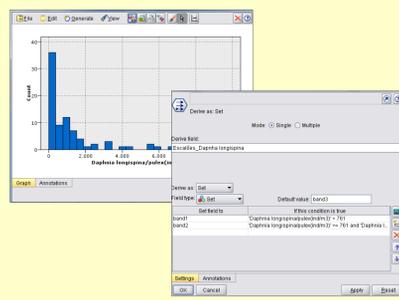
Hoje em dia, a quantidade de dados armazenados excede a capacidade de análise dos mesmos se não se recorrer ao uso de técnicas de análise automatizadas. Como consequência deste aumento efectivo de informação, o processamento através de métodos tradicionais tornou-se mais difícil e complexo. A descoberta de conhecimento em bases de dados é um campo que está em evolução de forma a fornecer soluções de análise automatizadas. O conhecimento obtido pode ser expresso na elaboração de modelos predictivos. No presente trabalho pretende-se avaliar, ainda que de forma preliminar, a potencial utilidade destas ferramentas para a detecção de tendências e padrões em dados de natureza limnológica. Os dados utilizados nesta primeira abordagem foram obtidos na Albufeira do Azibo (Latitude: 41°34'32, 25"N; Longitude: 6° 53'51, 63"W) ao longo de 84 campanhas de amostragem que decorreram mensalmente ao longo dos anos de 2000, 2001, 2002, 2007, 2008 e 2009. Nesta primeira fase investigam-se padrões de resposta das populações das espécies mais abundantes de crustáceos zooplancónicos à variação da precipitação e da temperatura da água.

DADOS UTILIZADOS NO PRESENTE ESTUDO



Clementine v 11.1 foi a ferramenta de “data mining” utilizada neste estudo. Na ferramenta existem uma variedade de modelos e técnicas. No entanto, este estudo assentou unicamente na aplicação de redes neuronais artificiais (RNA). Esta decisão aplicou-se devido ao facto das RNA serem uma das mais conhecidas e usadas técnicas em “data mining” (Berson et al. 2000). As RNA surgem como uma boa escolha na resolução de problemas de classificação, segmentação e previsão, quando os resultados do modelo são mais importantes do que compreender o modelo em si e a forma como este funciona (Santos & Azevedo 2005).

Para melhor compreensão dos resultados foram criados escalões sobre os dados em estudo. Este processo assentou sobre a aplicação de histogramas.



Conclusões do Estudo

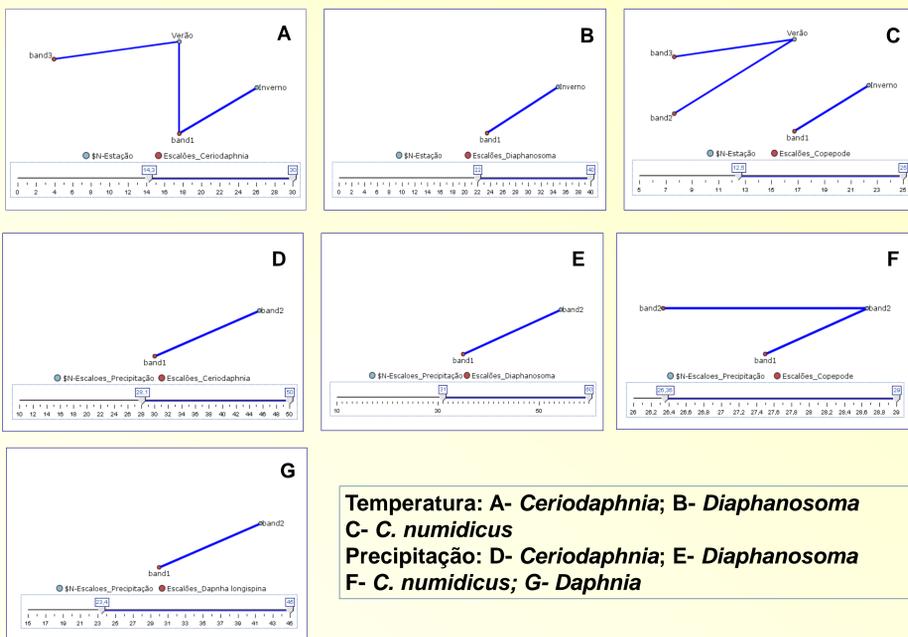
- **Influência da temperatura**
 - A aplicação da ferramenta de “Data Mining” indicia que a população de *Ceriodaphnia* apresenta densidades mais elevadas no período em que as temperaturas são mais altas.
 - No que diz respeito à população de *Diaphanosoma* apenas indicia que as densidades mais baixas ocorrem no Inverno. De facto, sendo esta espécie típica de águas quentes está ausente desta albufeira nos meses mais frios, atingido a densidade máxima a partir de Junho.
 - Relativamente a *C. numidicus* é possível inferir um pico máximo e outro intermédio durante o Verão, sendo a densidade mais baixa atingida no Inverno.
 - Não foi possível inferir qualquer tendência para *Daphnia*.

- **Influência da precipitação**

- Quando a precipitação atinge valores elevados as densidades de todas as espécies estudadas tendem a atingir baixas densidades com excepção da população de *C. numidicus*. Nesta população é possível inferir dois picos, um que corresponde a uma densidade baixa e outro que corresponde a uma densidade mais elevada. No entanto, a densidade máxima ocorre nos períodos em que a precipitação é reduzida.

Apesar das limitações criadas pela pequena dimensão da série de dados e da consequente impossibilidade de validação dos resultados obtidos todas as tendências indicadas pela aplicação destas ferramentas informáticas são corroboradas por estudos anteriores realizados nesta albufeira (Gerales & Boavida 2004a, 2004b). Assim, no futuro com séries de dados maiores, estas ferramentas poderão ser utilizadas para criar modelos robustos que permitam prever quais serão as respostas que terão os diferentes componentes abióticos e bióticos face a determinados factores de perturbação, permitindo assim a elaboração de medidas de gestão que permitam manter a integridade ecológica deste sistema.

RESULTADOS DA APLICAÇÃO DA FERRAMENTA DE DATA MINING



Bibliografia

- Berson A., Smith S. & Thearling K. 2000. Building Data Mining Applications for CRM. McGraw-Hill, USA.
 Carlson R. E. 1977. A trophic state index for lakes. *Limnol. Oceanogr.*, 22 (2): 361-369.
 Gerales A.M. & Boavida M. J. 2004a. What factors affect the pelagic cladocerans of the Azibo meso-eutrophic reservoir? *Ann Limnol- Int J Limnol* 40:101-111
 Gerales, A M e M J Boavida, 2004b. Limnological variations of a reservoir during two successive years: One wet, another dry. *Lakes and Reservoirs: Research & Management*. 9:143-152.
 Santos M. F. & Azevedo C. 2005. Data Mining - Descoberta de conhecimento de bases de dados. FCA - Editora de Informática.