

Programa Nacional do Meio Ambiente II – PNMA II

Componente: Desenvolvimento Institucional

Subcomponente: Monitoramento da Qualidade da Água

Projeto: Monitoramento da Qualidade da Água como Instrumento de Controle Ambiental e Gestão de Recursos Hídricos no Estado de Pernambuco

**Análise Estatística de Dados
do Monitoramento da Qualidade das Águas
do Rio Ipojuca e do Reservatório Tapacurá**

Relatório Final Consolidado

Maria Fernanda Pimentel

Recife, julho de 2003

Apresentação

Este documento reúne todos os relatórios de análise estatística dos dados de monitoramento da qualidade das águas do Rio Ipojuca e do Reservatório Tapacurá, elaborados como parte integrante do Projeto **Monitoramento da Qualidade da Água como Instrumento de Controle Ambiental e Gestão de Recursos Hídricos no Estado de Pernambuco**. Os resultados são apresentados de acordo com a ordem cronológica de sua obtenção, estabelecida no Plano de Trabalho acordado previamente (Produto 1, página 4).

Na primeira etapa da análise, procedeu-se à formatação dos bancos de dados disponibilizados, de forma a torná-los compatíveis com o exigido pelos programas computacionais empregados na análise estatística. Depois de verificados os valores registrados, e corrigidos ou excluídos os valores anômalos, chegou-se às planilhas que serviram de base para todos os procedimentos estatísticos. O conjunto de dados relativos ao Rio Ipojuca ficou com 842 amostras e 25 variáveis; o conjunto de dados do Reservatório Tapacurá, com apenas 35 amostras e 21 variáveis (Produto 2, página 8).

Em seguida foi realizada a análise estatística descritiva dos dados das duas áreas, na qual foram calculadas medianas, amplitudes, quartis e coeficientes de correlação. Foram construídos também, como parte das análises uni- e bivariadas, gráficos de linha, gráficos de caixa, gráficos matriciais e gráficos tridimensionais. Conforme a necessidade, esses gráficos foram estratificados por varredura e/ou por variáveis (Produto 3, página 16).

Na terceira etapa (Produto 4, página 47), foram realizadas análises de componentes principais dos conjuntos de dados completos e de alguns subconjuntos. Nos gráficos dos escores obtidos nessas análises foi possível identificar padrões de associação relacionados com a estação climática e com as estações de coleta das amostras (no caso do Rio Ipojuca). Os gráficos dos pesos confirmaram padrões de associação já conhecidos, como a alta correlação existente entre o teor de cloreto e a condutividade.

Finalmente, constatou-se que análises de componentes principais realizadas com um número reduzido de amostras, para simular uma amostragem bimestral, apresentam padrões muito semelhantes aos das análises completas (Produto 5, página 67). Isto é uma indicação de que uma política de amostragem reduzida pode vir a ser implementada. Uma decisão nesse sentido, porém, deve levar em conta, além de questões de logística, outros parâmetros de interesse ambiental que não puderam ser incluídos nas análises constantes deste documento.

Como previsto no plano de trabalho, todas as atividades contaram com a participação de dois técnicos (Helena Alves Moreira e Luciano Costa Almeida), que ao longo do projeto receberam treinamento nos seguintes procedimentos, no ambiente do programa computacional *Statistica 5.5*:

- Reconhecimento dos elementos da interface gráfica
- Como importar planilhas do formato *Excel*
- Como acrescentar, excluir e modificar casos e variáveis
- Como recodificar os valores numéricos de uma variável
- Como inserir valores de texto numa variável
- Como definir uma variável através de uma função
- Como exportar e salvar tabelas e gráficos
- Construção de subconjuntos de dados obedecendo a critérios de seleção pré-estabelecidos
- Reordenamento hierárquico de casos de acordo com os valores de variáveis de interesse
- Cálculo de várias estatísticas descritivas
- Construção de gráficos seqüenciais para variáveis e casos (gráficos de linha)
- Construção de gráficos de caixa, gráficos matriciais e gráficos tridimensionais
- Estratificação de conjuntos de dados por variável e por varredura
- Rotulação de gráficos de acordo com os valores de variáveis de interesse
- Modificação dos elementos tipográficos e da aparência de um gráfico qualquer
- Análise e interpretação de componentes principais
- Análise e interpretação de agrupamentos hierárquicos

Recife, 30 de julho de 2003.

Maria Fernanda Pimentel

Programa Nacional do Meio Ambiente II – PNMA II

Componente: Desenvolvimento Institucional

Subcomponente: Monitoramento da Qualidade da Água

Projeto: Monitoramento da Qualidade da Água como Instrumento de Controle Ambiental e Gestão de Recursos Hídricos no Estado de Pernambuco

**Análise Estatística de Dados
do Monitoramento da Qualidade das Águas
do Rio Ipojuca e do Reservatório Tapacurá**

Produto 1 – Plano de Trabalho

Maria Fernanda Pimentel

Recife, Novembro de 2002

Introdução

O presente trabalho tem como objetivo a análise estatística de dados históricos do monitoramento da qualidade das águas da Bacia do Rio Ipojuca e do Reservatório de Tapacurá, ambos situados no Estado de Pernambuco, visando fornecer subsídios à CPRH/SRH e à SECTMA para propor eventuais alterações no atual sistema de monitoramento. Toda a análise deverá ser acompanhada por técnicos das instituições envolvidas no projeto, que assim serão treinados na metodologia analítica empregada e deverão tornar-se capazes de utilizá-la para abordar problemas análogos, com dados de outra proveniência.

A análise estatística de um grande volume de dados é praticamente inviável sem o auxílio de programas computacionais especializados. O primeiro passo do trabalho, portanto, consistirá em adaptar os dados ao formato exigido pelos programas que serão usados na análise, uma cópia dos quais estará à disposição dos participantes, durante o treinamento.

Uma vez que os dados estejam devidamente formatados, proceder-se-á à análise propriamente dita, que inicialmente consistirá na descrição, tanto quantitativa quanto gráfica, das distribuições uni- e bivariadas do conjunto de dados. Nesse ponto, será possível fazer uma análise preliminar dos dados das duas áreas, identificando possíveis valores anômalos, bem como subconjuntos de amostras e variáveis que serão destacados para uma análise mais detalhada, por meio de técnicas estatísticas multivariadas.

A análise multivariada exploratória empregará principalmente a análise de componentes principais, uma técnica de redução da dimensionalidade que permite projetar o máximo de informação no menor número possível de dimensões não correlacionadas. Analisando-se os gráficos dos escores e dos pesos resultantes dessa análise, pode-se identificar facilmente padrões de associação e similaridade tanto entre amostras quanto entre variáveis. Esses padrões servirão de base para estabelecer relações espaciais e/ou temporais entre os parâmetros de qualidade de águas monitoradas, bem como entre as estações de monitoramento.

Paralelamente à análise de componentes principais, realizar-se-á uma análise de agrupamentos hierárquica, cujo resultado é uma figura bidimensional conhecida como dendrograma, construída a partir da matriz das distâncias entre os pontos no espaço multidimensional. O dendrograma tem a característica de utilizar toda a informação, e não apenas parte dela, como ocorre na análise de componentes principais. A inspeção visual de um dendrograma também permite identificar padrões de associação.

As conclusões e recomendações do trabalho deverão basear-se nas duas técnicas de análise multivariada, que são ferramentas complementares.

Atividades

As atividades a serem desenvolvidas são detalhadas abaixo:

1. Formatação dos bancos de dados do monitoramento da qualidade das águas da Bacia do Rio Ipojuca e do Reservatório de Tapacurá, de forma a torná-los compatíveis com o exigido pelos programas computacionais empregados na análise estatística.
2. Importação do resultado do item 1 para os programas de análise estatística de dados e subsequente verificação, para identificar e eventualmente corrigir possíveis valores anômalos.
3. Análise estatística descritiva dos dados das duas áreas, utilizando, pelo menos, as seguintes ferramentas: cálculo de várias estatísticas descritivas, construção de gráficos de linha, gráficos de caixa, gráficos matriciais, gráficos tridimensionais, estratificação por varredura e/ou por variáveis.
4. Identificação de subconjuntos de variáveis e amostras para a realização da análise estatística multivariada.
5. Análise estatística multivariada, através das técnicas de projeções em componentes principais e análise de agrupamentos hierárquica, visando identificar padrões de associação ou semelhança entre amostras ou variáveis.
6. Reunião com técnicos da CPRH, da SECTMA e da SRH, para divulgação dos resultados do trabalho.
7. Participação em oficina de trabalho relativa à identificação de zonas homogêneas de monitoramento.
8. Preparação e apresentação do relatório técnico final, contemplando as discussões e conclusões do trabalho.

O desenvolvimento de todas essas atividades será acompanhado por técnicos designados pela SECTMA, pela CPRH e pela SRH. Esses profissionais receberão um treinamento que consistirá em discussões teóricas sobre as técnicas estatísticas e gráficas empregadas, acompanhadas de exercícios práticos com os programas computacionais e os dados utilizados nas análises.

Cronograma

O Quadro 1 apresenta o cronograma a ser seguido no desenvolvimento das atividades descritas anteriormente, bem como os produtos resultantes.

Quadro 1 – Cronograma de atividades/produtos.

Atividades/Produtos	Prazos			
	03/12/02	02/01/03	03/02/02	03/03/03
Atividades 1 e 2 Produto 2 (Relatório Técnico 1)	■			
Atividade 3 Produto 3 (Relatório Técnico 2)		■		
Atividades 4, 5, 6 e 7 Produto 4 (Relatório Técnico 3)			■	
Atividade 8 Produto 5 (Relatório Técnico Final)				■

Programa Nacional do Meio Ambiente II – PNMA II

Componente: Desenvolvimento Institucional

Subcomponente: Monitoramento da Qualidade da Água

Projeto: Monitoramento da Qualidade da Água como Instrumento de Controle Ambiental e Gestão de Recursos Hídricos no Estado de Pernambuco

**Análise Estatística de Dados
do Monitoramento da Qualidade das Águas
do Rio Ipojuca e do Reservatório Tapacurá**

Produto 2 – Relatório 1

Maria Fernanda Pimentel

Recife, dezembro de 2002

Introdução

O presente relatório se refere ao desenvolvimento das atividades 1 e 2 previstas no plano de trabalho, a saber:

1. Formatação dos bancos de dados do monitoramento da qualidade das águas da Bacia do Rio Ipojuca e do Reservatório de Tapacurá, de forma a torná-los compatíveis com o exigido pelos programas computacionais empregados na análise estatística.
2. Importação do resultado do item 1 para os programas de análise estatística de dados e subsequente verificação, para identificar e eventualmente corrigir possíveis valores anômalos.

No que segue passamos a descrever o desenvolvimento das duas atividades previstas. As Tabelas e Figuras mencionadas no texto encontram-se agrupadas no final do documento

Atividade 1

Os dados fornecidos pela SRH e pela CPRH foram digitados em duas planilhas do programa computacional *MS-Excel*, no formato convencional da análise multivariada, onde cada linha corresponde a uma amostra e cada coluna corresponde a uma variável. Disso resultaram dois arquivos, um referente à Bacia do Rio Ipojuca e outro correspondente ao reservatório de Tapacurá. O conjunto de dados do Rio Ipojuca é muito mais extenso, já que alguns de seus valores remontam a 1986, ao passo que os registros mais antigos referentes a Tapacurá são de maio de 1999. Algumas das variáveis medidas são comuns aos dois arquivos, mas outras são específicas de um deles, como mostra a Tabela 1.

Atividade 2

As duas planilhas foram convertidas do formato Excel para o do programa computacional *Statistica*, onde deverão ser realizadas todas as análises subsequentes. Os dois conjuntos de dados resultantes dessa operação foram então submetidos a uma análise gráfica preliminar, com o objetivo de identificar e retificar incorreções. Vários problemas foram detectados e corrigidos, entre os quais podemos mencionar:

- Linhas ou colunas totalmente vazias
- Dados de texto lidos como se fossem valores numéricos
- Blocos de dados associados a amostras incorretas
- Ponto decimal deslocado ou com notação tipográfica inadequada

- Valores ausentes tomados como nulos
- Valores numéricos em discordância com os registros originais

Depois dessa etapa de depuração, ficamos com dois conjuntos de dados, no formato do programa *Statistica*, prontos para análise. O conjunto relativo a Tapacurá contém 35 casos (amostras) e 21 variáveis. O de Ipojuca é constituído por 842 casos e 25 variáveis. Tais conjuntos, porém, apresentam um número considerável de dados ausentes, sobretudo no caso de Ipojuca. Isto aparentemente é uma conseqüência de mudanças nos procedimentos de coleta ou de registro de dados ocorridas ao longo dos anos. Por outro lado, certamente representará uma limitação quanto ao tipo de análise que poderá ser realizado com determinadas variáveis.

A Tabela 2 mostra quantos valores numéricos válidos estão disponíveis para cada variável nos dois conjuntos de dados. As únicas variáveis medidas para todas as amostras foram a salinidade e a condutividade elétrica, e isso ocorreu apenas para os dados de Tapacurá.

Uma outra restrição que deve ser ressaltada é a falta de coincidência entre as lacunas. Muitas vezes os dados de uma variável foram obtidos para uma amostra onde os dados de uma segunda variável estão faltando, enquanto essa segunda variável foi registrada para outra amostra para a qual os dados da primeira variável estão ausentes. Dependendo do subconjunto escolhido para a análise multivariada, esta pode vir a ser uma restrição severa.

Finalmente, a análise preliminar permitiu a identificação de algumas amostras anômalas, em algumas variáveis. Tais amostras, embora representadas por valores legítimos, apresentam um comportamento visivelmente diferente das demais. Esse aspecto também terá de ser levado em conta durante a realização das análises, já que amostras anômalas podem distorcer gravemente determinados testes estatísticos. Alguns exemplos de anomalias são mostrados nas Figuras 1 - 4.

Treinamento

Como previsto no plano de trabalho, todas as atividades contaram com a participação dos técnicos Helena Alves Moreira (CPRH) e Luciano Costa Almeida (SRH), que receberam treinamento nos seguintes procedimentos, no ambiente do programa *Statistica 5.5*:

- Reconhecimento dos elementos da interface gráfica
- Como importar planilhas do formato *Excel*
- Como acrescentar, excluir e modificar casos e variáveis

- Como recodificar os valores numéricos de uma variável
- Como inserir valores de texto numa variável
- Como definir uma variável através de uma função
- Como exportar e salvar tabelas e gráficos

Recife, 2 de dezembro de 2002.

Maria Fernanda Pimentel

Tabela 1 – Variáveis acompanhadas na Bacia do Rio Ipojuca e no Reservatório de Tapacurá.

Variável	Unidade	Tapacurá	Ipojuca
Condutividade elétrica	µS/cm	•	•
Salinidade	%	•	
Oxigênio dissolvido	mg L ⁻¹	•	•
O. D. saturado	%	•	•
Nitrato	mg L ⁻¹	•	•
Nitrito	mg L ⁻¹	•	
Clorofila-a total	µg/L	•	
Turbidez	NTU	•	•
Sólidos em suspensão	mg L ⁻¹	•	
Sólidos dissolvidos	mg L ⁻¹	•	
Sólidos totais a 105°C	mg L ⁻¹		•
Microfitoplancton total	org mL ⁻¹	•	
Microfitoplancton total	células /ml	•	
Microfitoplancton – cianobactéria	cianobactérias mL ⁻¹	•	
Coliformes fecais	nmp/100 mL	•	•
Cloreto	mg L ⁻¹	•	•
Alcalinidade	mg L ⁻¹	•	
pH	---	•	•
Temperatura	°C	•	•
Nitrogênio amoniacal	mg L ⁻¹ em N	•	•
Fósforo como PO ₄	mg L ⁻¹ em P	•	•
Fósforo particulado	mg L ⁻¹ em P	•	
D. B. O. (5 dias a 20 °C)	mg L ⁻¹		•
Cor	Escala Pt-Co		•
% de saturação	%		•
Cádmio	mg L ⁻¹		•
Chumbo	mg L ⁻¹		•
Cobalto	mg L ⁻¹		•
Cobre	mg L ⁻¹		•
Cromo	mg L ⁻¹		•
Ferro	mg L ⁻¹		•
Manganês	mg L ⁻¹		•
Níquel	mg L ⁻¹		•
Zinco	mg L ⁻¹		•
Índice de fenóis	mg L ⁻¹		•

Tabela 2 – Valores numéricos válidos nos dois conjuntos de dados.

Variável	Tapacurá	Ipojuca
Condutividade elétrica	35	249
Salinidade	35	
Oxigênio dissolvido	33	803
O. D. saturado	4	111
Nitrato	17	130
Nitrito	18	
Clorofila -a total	18	
Turbidez	18	196
Sólidos em suspensão	18	
Sólidos dissolvidos	27	
Sólidos totais a 105°C		138
Microfitoplancton total	20	
Microfitoplancton total	10	
Microfitoplancton – cianobactéria	10	
Coliformes fecais	28	663
Cloreto	17	244
Alcalinidade	15	
pH	32	323
Temperatura	34	322
Nitrogênio amoniacal	18	171
Fósforo como PO ₄	18	66
Fósforo particulado	17	
D. B. O. (5 dias a 20 °C)		797
Cor		204
% de saturação		110
Cádmio		17
Chumbo		17
Cobalto		1
Cobre		17
Cromo		17
Ferro		17
Manganês		17
Níquel		17
Zinco		17
Índice de fenóis		8

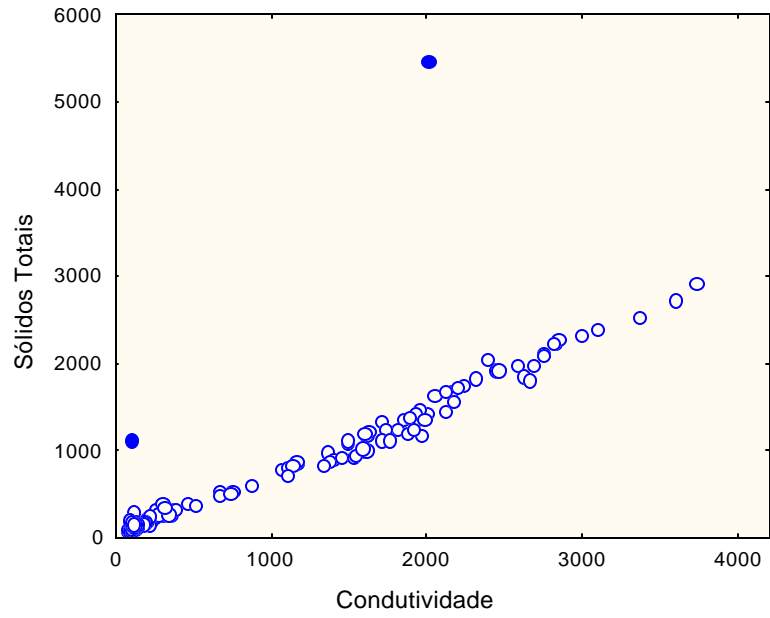


Figura 1 – Condutividade elétrica e sólidos totais na Bacia do Rio Ipojuca.

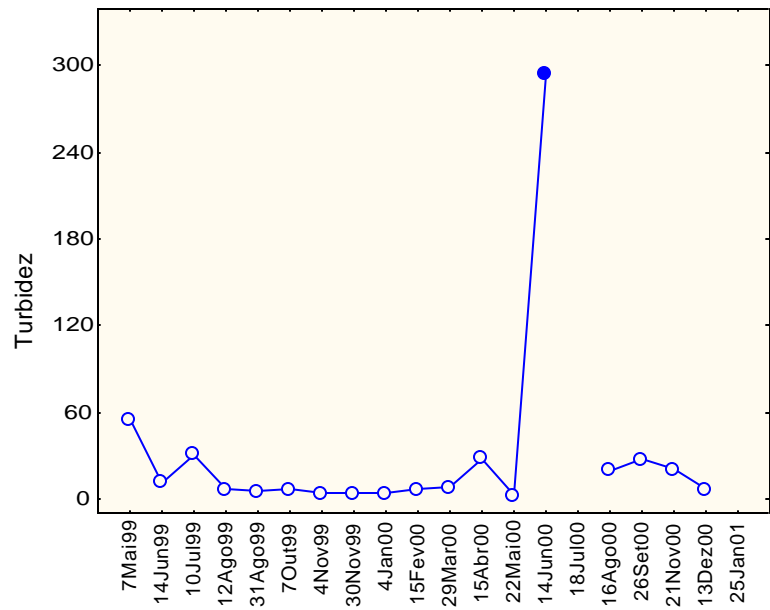


Figura 2 – Turbidez no Reservatório de Tapacurá.

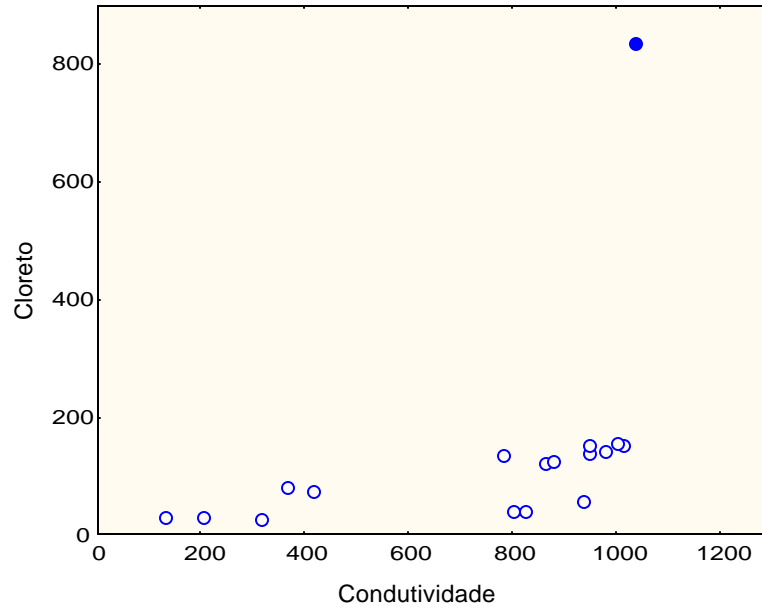


Figura 3 – Condutividade elétrica e cloreto no Reservatório de Tapacurá.

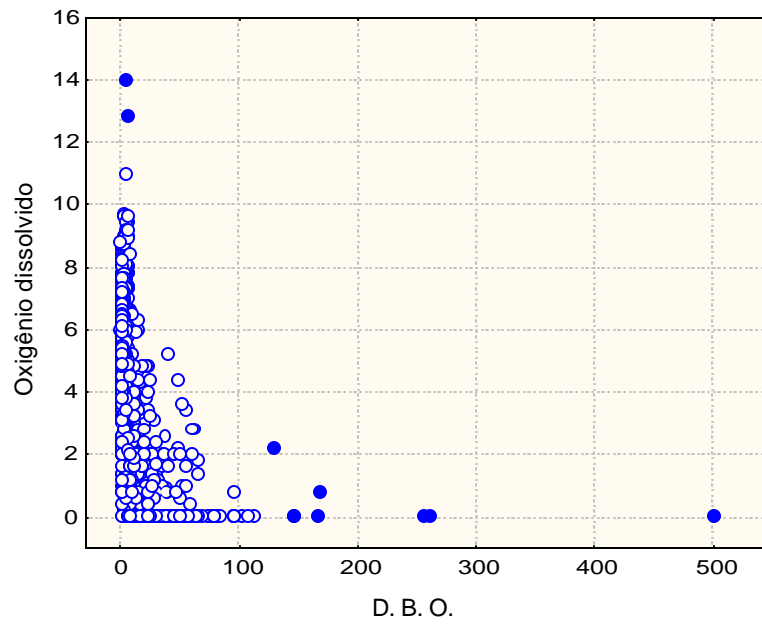


Figura 4 – Oxigênio dissolvido e D. B. O. na Bacia do Rio Ipojuca.

Programa Nacional do Meio Ambiente II – PNMA II

Componente: Desenvolvimento Institucional

Subcomponente: Monitoramento da Qualidade da Água

Projeto: Monitoramento da Qualidade da Água como Instrumento de Controle Ambiental e Gestão de Recursos Hídricos no Estado de Pernambuco

**Análise Estatística de Dados
do Monitoramento da Qualidade das Águas
do Rio Ipojuca e do Reservatório Tapacurá**

Produto 3 – Relatório 2

Maria Fernanda Pimentel

Recife, janeiro de 2003

Introdução

O presente relatório se refere ao desenvolvimento da atividade 3 prevista no plano de trabalho, a saber:

3. Análise estatística descritiva dos dados das duas áreas, utilizando, pelo menos, as seguintes ferramentas: cálculo de várias estatísticas descritivas, construção de gráficos de linha, gráficos de caixa, gráficos matriciais, gráficos tridimensionais, estratificação por varredura e/ou por variáveis.

No que segue passamos a descrever o desenvolvimento da atividade prevista, agrupando os resultados de acordo com a região analisada.

Análise dos dados do Rio Ipojuca

Os dados relativos ao Rio Ipojuca, disponibilizados pela contratante, constituem uma série histórica bastante extensa, contendo um total de 842 amostras, a primeira das quais colhida em setembro de 1986, e a mais recente obtida em dezembro de 2001. Considerando-se o conjunto como um todo, dezesseis pontos de coleta estão representados. Ao longo do tempo, porém, parece ter havido modificações na metodologia, fazendo com que apenas cinco estações de coleta (50, 65/64, 70, 85 e 95) contenham registros para todo o período (Tabela 1). Mesmo estas estações apresentam lacunas em diversas variáveis, o que aliás se verifica em todo o conjunto de dados. Nenhuma variável teve o seu valor registrado para todas as 842 amostras. As que mais se aproximaram disto foram o oxigênio dissolvido e a D. B. O. (803 e 797 registros, respectivamente). No outro extremo está o teor de cobalto, com uma única determinação, e esta, aliás, abaixo do limite de detecção. A Figura 1 mostra o número de valores válidos para todas as variáveis. Dela podemos concluir imediatamente que a maioria das análises estatísticas multivariadas terá de ser feita com um número bastante reduzido de valores.

As estações de coleta representam diferentes realidades ecológicas e sócio-econômicas ao longo do Rio, o que se reflete na grande variação existente entre elas, para a maioria dos parâmetros.

Os gráficos de caixa das Figuras 2 – 10 mostram o comportamento dos diversos parâmetros monitorados no Ipojuca durante o ano de 2001. Como a numeração das estações cresce da nascente ao estuário, podemos acompanhar nesses gráficos a variação nos valores dos parâmetros ao longo do Rio. De estação para estação mudam não só os valores centrais (escolhemos como representante da tendência central a mediana, isto é, o ponto que separa os 50% valores inferiores do conjunto de dados dos 50% valores superiores, por causa da evidente assimetria de alguns conjuntos de valores), como também a amplitude (isto é, a

diferença entre os valores mínimo e máximo) e a própria simetria (isto é, a forma da distribuição raramente é simétrica em torno do valor central). As retas em vermelho indicam os limites estabelecidos pela Resolução 20/86 do Conama para águas da Classe 2, por orientação da contratante.

Tomando como exemplo o gráfico de caixa da D. B. O. (Figura 2) vemos que o limite da Classe 2 (<5 mg L⁻¹) só não foi ultrapassado pelas duas amostras coletadas na estação 40. Nas demais, pelo menos um valor ultrapassou o limite da classe. Na estação 25, e sobretudo na estação 50, todos os valores excederam 5 mg L⁻¹. Os demais gráficos de caixa (Figuras 3-10) revelam a mesma heterogeneidade para os demais parâmetros. Nesses gráficos o ponto central é a mediana, as linhas definem os limites da amplitude, e a caixa define a distância interquartílica, isto é, a região abaixo da qual estão os 25% valores mais baixos e acima da qual estão os 25% valores mais altos do conjunto de dados para a variável e a estação representadas.

Estas considerações mostram que é conveniente realizarmos análises de séries temporais separadas de acordo com a estação, para que a diferença de uma estação para outra seja eliminada e possamos tentar identificar alguma tendência de evolução dos parâmetros com o tempo (Figuras 12 a 16).

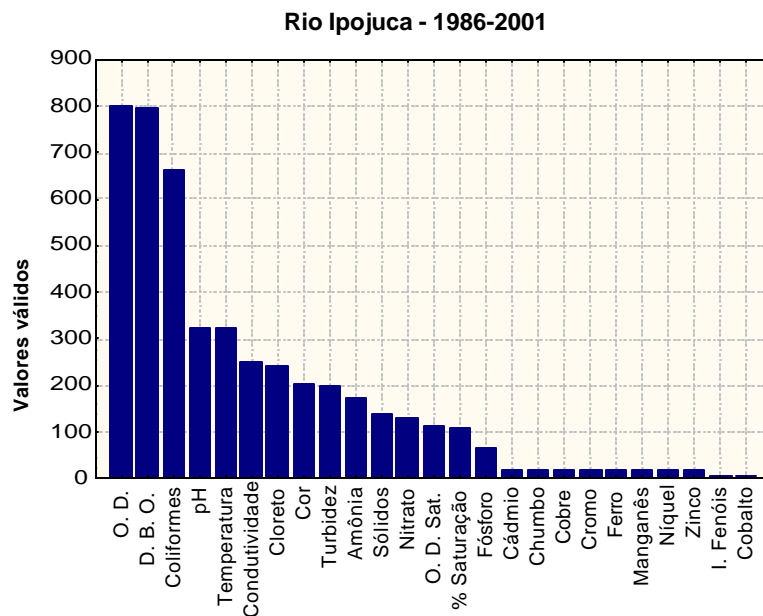


Figura 1 – Valores válidos para os parâmetros monitorados na Bacia do Rio Ipojuca, no período 1986-2001.

Tabela 1 – Amostras coletadas na Bacia do Rio Ipojuca, e seus períodos de coleta. *N* representa o número de amostras coletadas em cada ponto.

Ponto	<i>N</i>	1980					1990					2000					
		6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
IP 2-05	7						■	■	■								
IP 2-07	3							■	■								
IP 2-10	61	■	■	■	■	■	■	■	■	■	■	■	■	■			
IP 2-15	8						■	■	■					■			
IP 2-25	34						■	■	■	■	■	■	■			■	
IP 2-32	27							■	■		■	■	■				
IP 2-40	31						■	■	■		■	■	■			■	
IP 2-45	23	■	■	■	■	■		■	■								
IP 2-50	98	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
IP 2-55	60						■	■	■		■	■	■	■	■	■	■
IP 2-64	70							■	■	■	■	■	■	■	■	■	■
IP 2-65	32	■	■	■	■	■	■	■	■								
IP 2-70	102	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
IP 2-85	108	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
IP 2-90	75						■	■	■		■	■	■	■	■	■	■
IP 2-95	79	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Na análise das séries temporais, para que a representatividade estatística seja a maior possível, vamos nos restringir às cinco estações que foram monitoradas continuamente, e aos três parâmetros com maior número de determinações – O. D., D. B. O e Coliforme fecal. Ao invés de usar os valores originais, preferimos realizar um pré-processamento antes da análise, para compensar o fato de que a D. B. O. e o Coliforme fecal variam por mais de uma ordem de grandeza, e também para exibir mais claramente possíveis afastamentos em relação aos limites da classe 2 da Resolução 20/86 do Conama, onde todas as estações de coleta da Bacia do Ipojuca foram enquadradas. Foram escolhidas as seguintes transformações:

$$O. D. = O. D. - 5 \text{ mg L}^{-1};$$

$$D. B. O. = \log \frac{D. B. O.}{5 \text{ mg L}^{-1}};$$

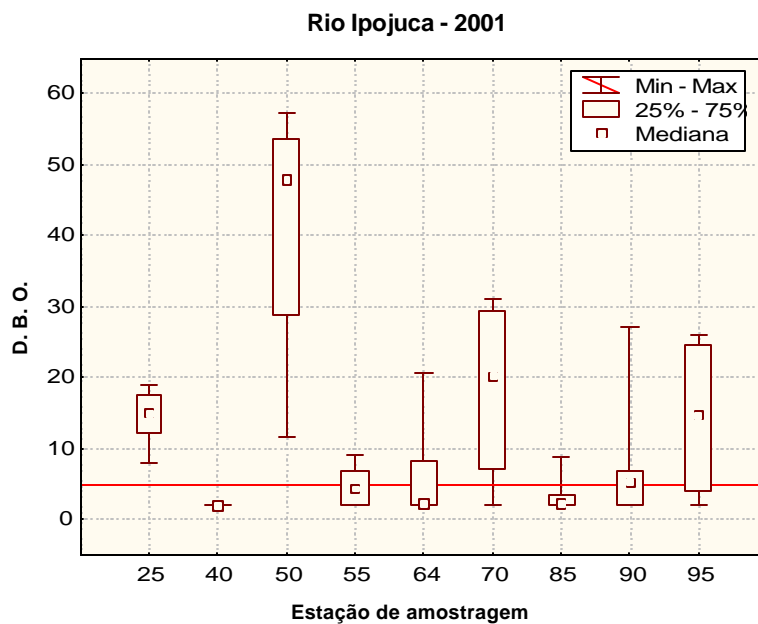


Figura 2 – Estatística descritiva da D. B. O. nas estações do Rio Ipojuca em 2001. A reta indica o valor limite estabelecido pela Resolução 20/86 do Conama para águas da Classe 2.

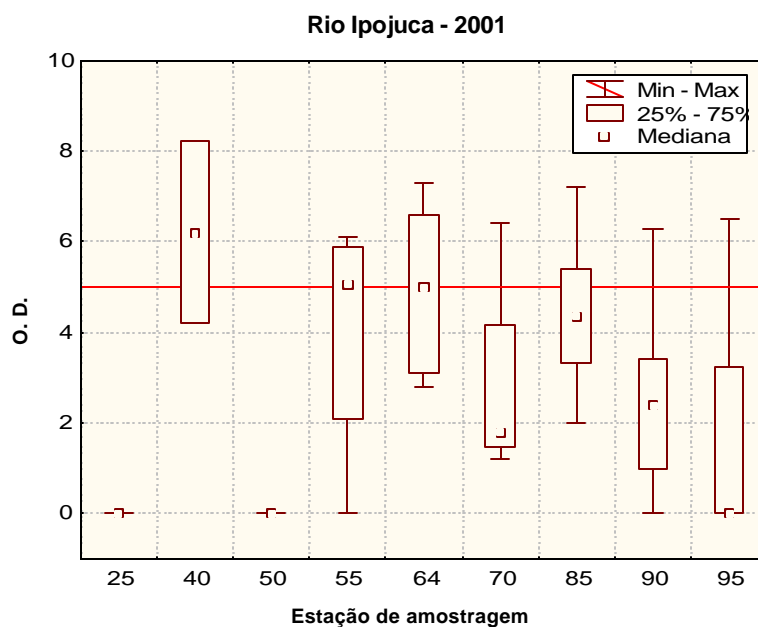


Figura 3 – Estatística descritiva do oxigênio dissolvido nas estações do Rio Ipojuca em 2001. A reta indica o valor limite estabelecido pela Resolução 20/86 do Conama para águas da Classe 2.

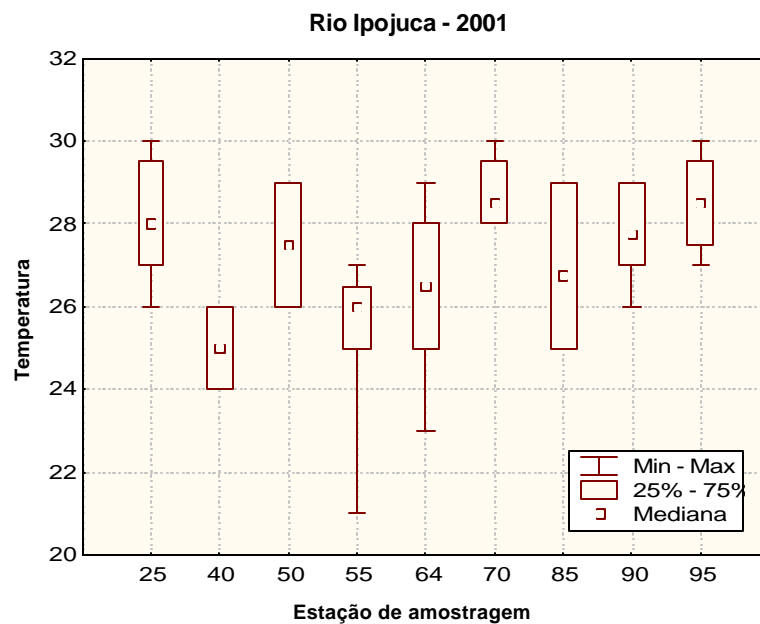


Figura 4 – Estatística descritiva da Temperatura nas estações do Rio Ipojuca em 2001.

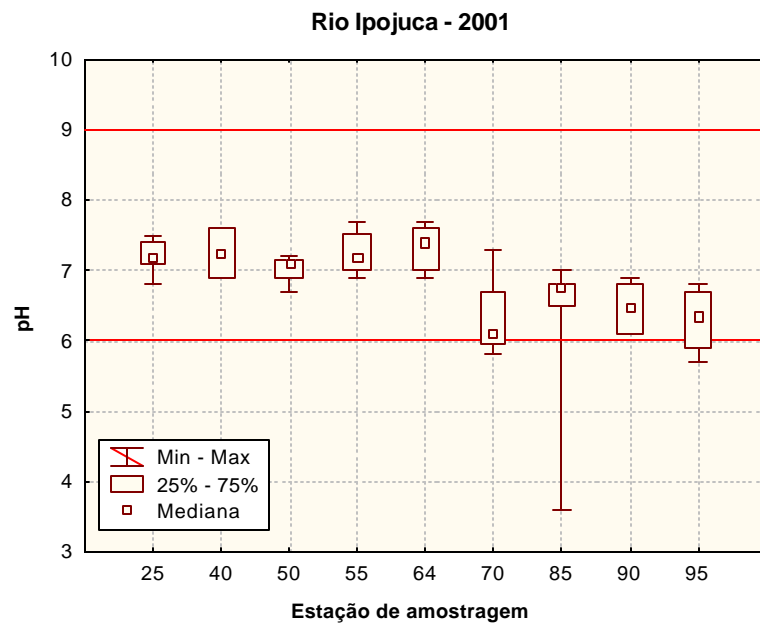


Figura 5 – Estatística descritiva do pH nas estações do Rio Ipojuca em 2001. As retas indicam os limites estabelecidos pela Resolução Conama 20/86 para águas da Classe 2.

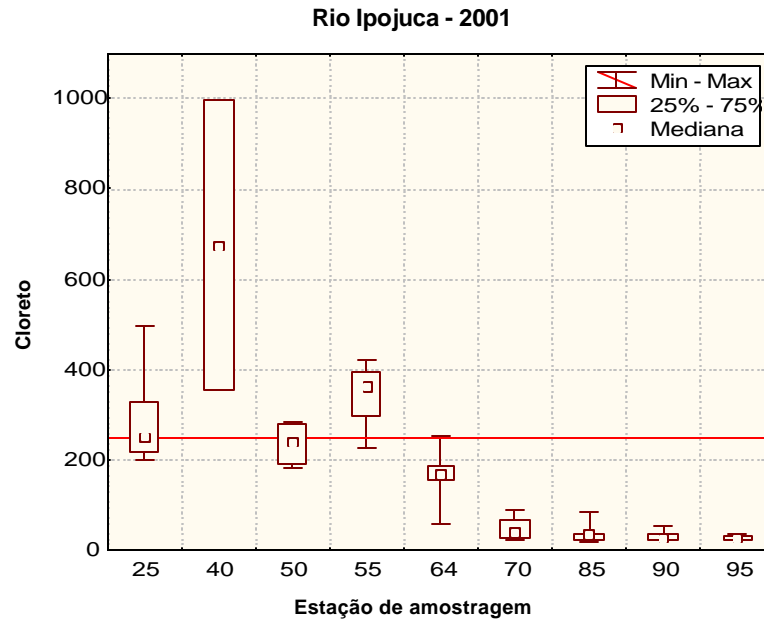


Figura 6 – Estatística descritiva do Cloro nas estações do Rio Ipojuca em 2001. A reta indica o valor limite estabelecido pela Resolução Conama 20/86 para águas da Classe 2.

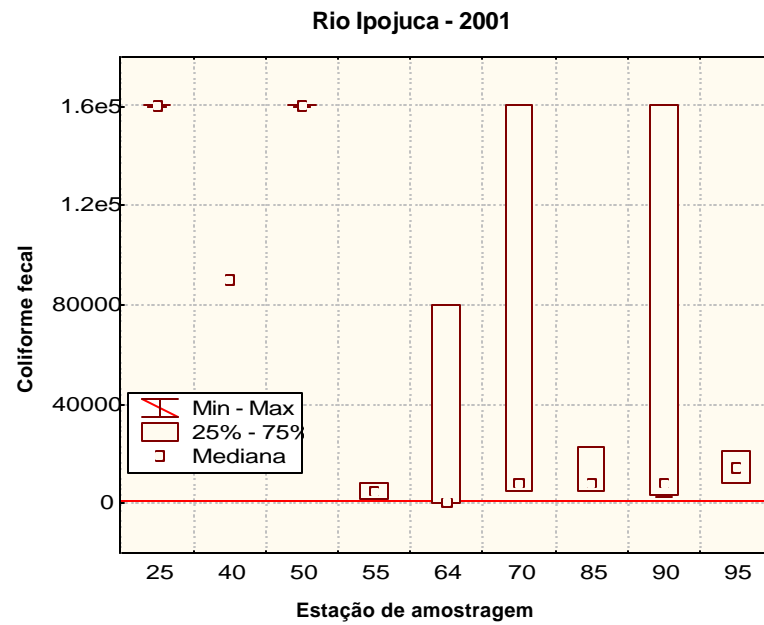


Figura 7 – Estatística descritiva do Coliforme fecal nas estações do Rio Ipojuca em 2001. A reta indica o valor limite estabelecido pela Resolução 20/86 do Conama para águas da Classe 2.

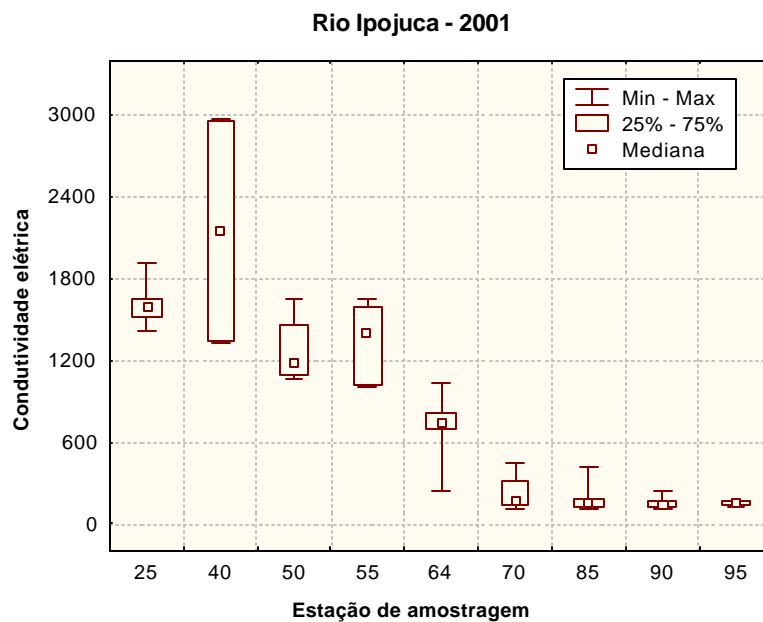


Figura 8 – Estatística descritiva da condutividade elétrica nas estações do Rio Ipojuca em 2001.

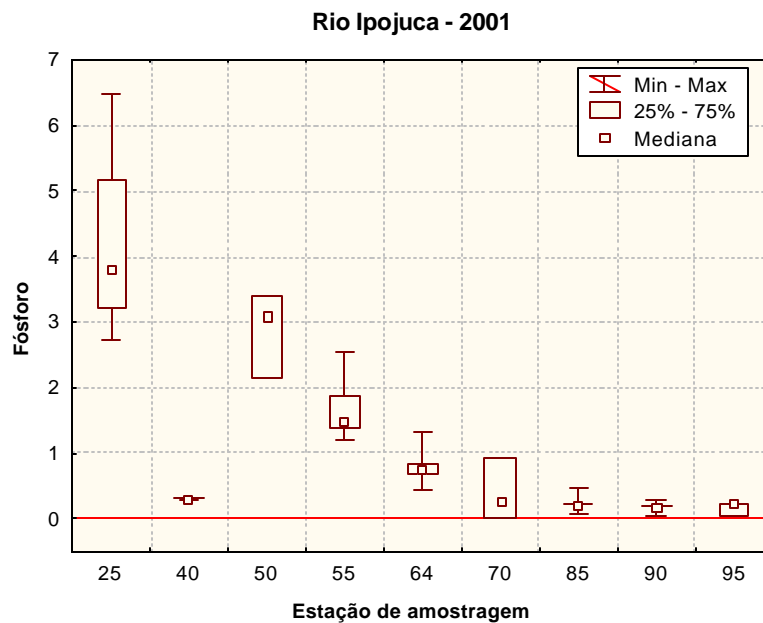


Figura 9 – Estatística descritiva do Fósforo nas estações do Rio Ipojuca em 2001. A reta indica o valor limite estabelecido pela Resolução 20/86 do Conama para águas da Classe 2.

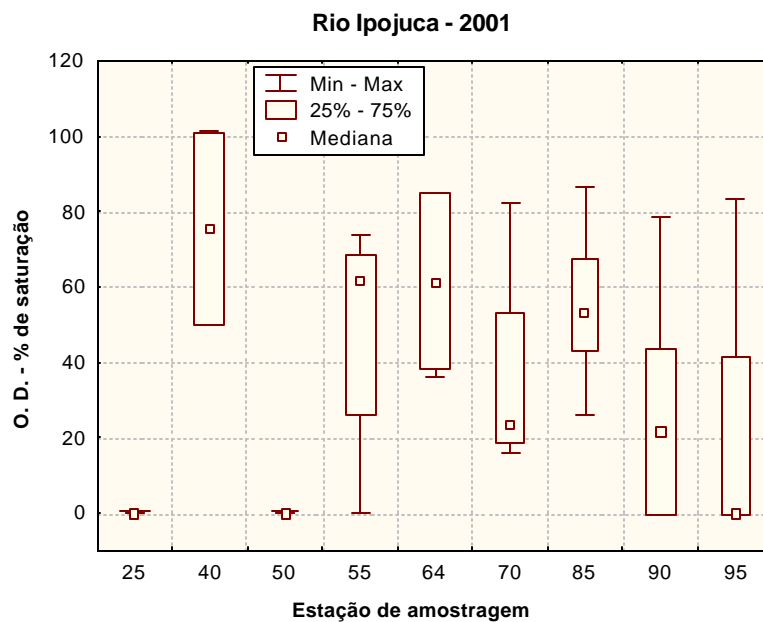


Figura 10 – Estatística descritiva da percentagem de saturação do oxigênio dissolvido nas estações do Rio Ipojuca em 2001.

$$Coliforme\ fecal = \log \frac{Coliforme\ fecal}{1.000\ NMP/100\ mL} .$$

Com estas transformações, o valor zero passa a representar o valor limite das três variáveis para a classe 2, e torna-se mais fácil observar possíveis padrões de comportamento em referência a ele. A Figura 11 ilustra o efeito desse pré-processamento sobre a D. B. O.

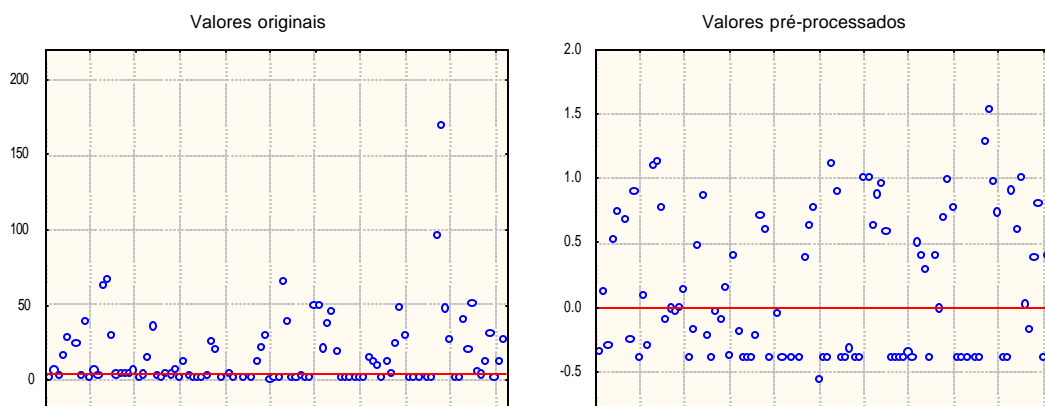


Figura 11 – Efeito do pré-processamento sobre a série dos valores da D. B. O. na Estação 70. As retas em vermelho sinalizam o limite da classe 2.

As Figuras 12-16, nas páginas seguintes, mostram como o O. D., a D. B. O. e o Coliforme fecal comportaram-se nas cinco estações selecionadas, ao longo dos anos 1986-2001. Os gráficos estão estratificados de acordo com o período de coleta – inverno (março a julho para as estações IP-25, IP-40, IP-50 e IP-55, e março a agosto para as demais) ou verão (os meses complementares, obviamente), já que o comportamento desses parâmetros pode ser influenciado pelo regime pluviométrico.

Da análise das séries temporais mostradas nas Figuras 12–16 podemos extrair as seguintes conclusões:

- No tocante às três variáveis analisadas, a maioria dos pontos situa-se fora dos padrões estabelecidos pela Resolução 20/86 do Conama para a Classe 2. A situação apresenta-se mais crítica nos meses de verão.
- Considerando-se o conjunto das cinco estações, a variável com maior número de pontos fora do padrão de classe é o Coliforme fecal. Essa variável apresenta também uma tendência de piora ao longo do tempo, que é mais clara nas Estações 70, 85 e 95.
- O comportamento do O. D. e da D. B. O. é particularmente crítico na Estação 50, situada no município de Caruaru.
- O efeito da alternância inverno/verão sobre o O. D. e a D. B. O. é mais visível nas Estações 85 e 95, localizadas mais próximas do estuário.
- Nos últimos anos, a Estação 65/64 apresenta uma melhora nos níveis de D. B. O., particularmente durante o verão. Essa mudança parece estar associada à substituição da Estação 65 pela Estação 64, ocorrida em 1992, já que a série temporal da D. B. O. apresenta uma nítida mudança de patamar nos valores do verão entre 1992 e 1993.

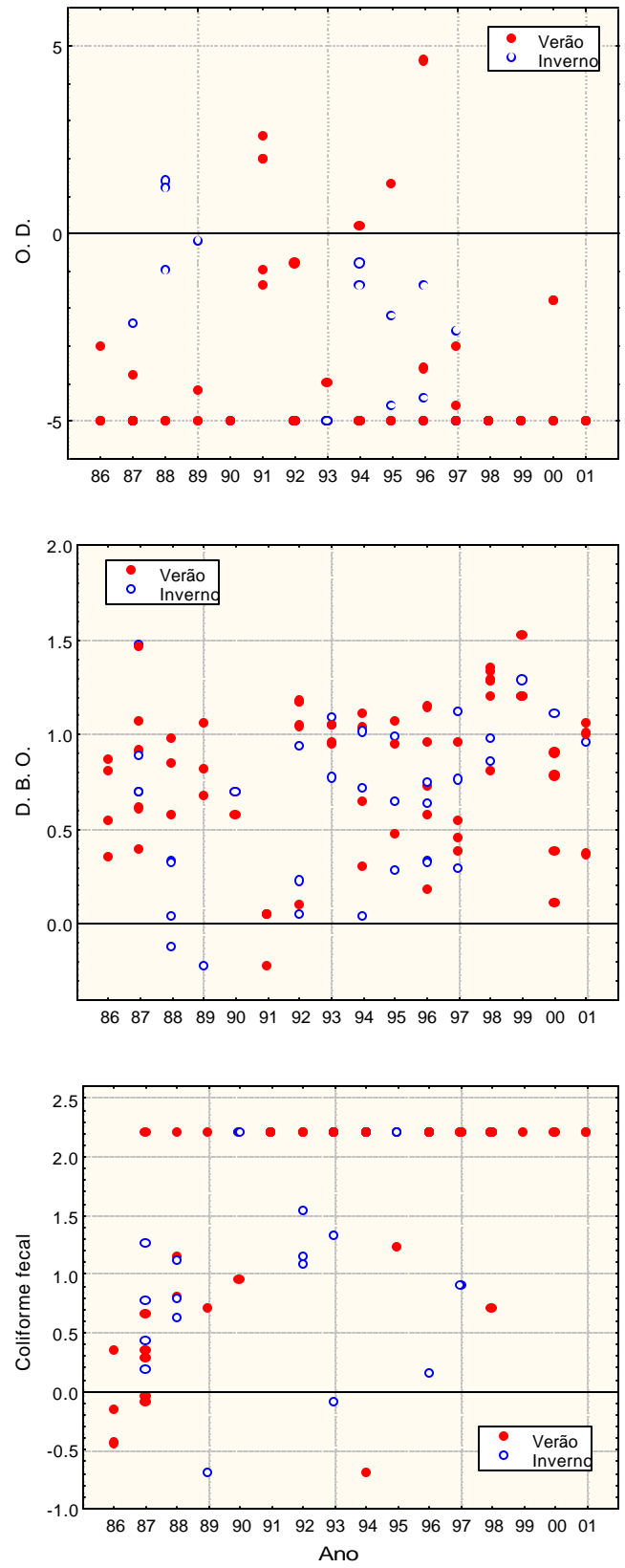


Figura 12 – Evolução temporal dos dados do Rio Ipojuca– Estação 50.

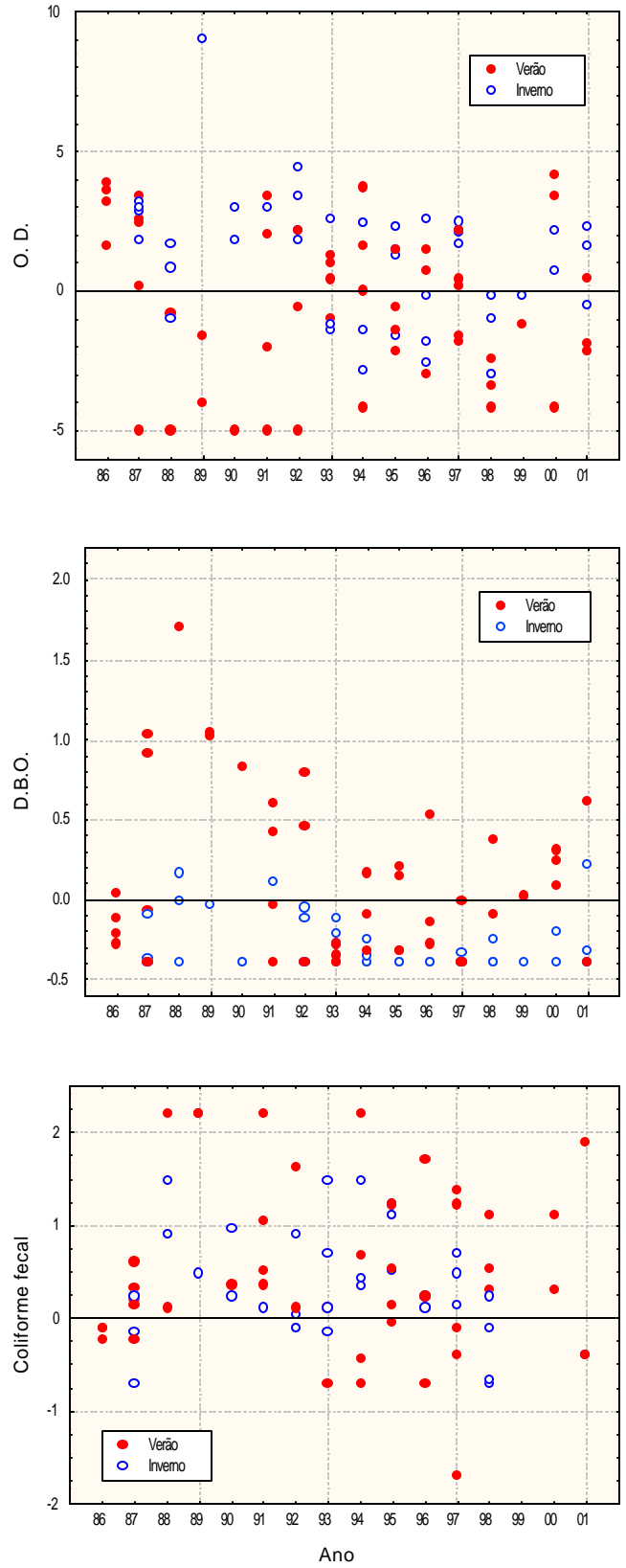


Figura 13 – Evolução temporal dos dados do Rio Ipojuca – Estação 65/64.

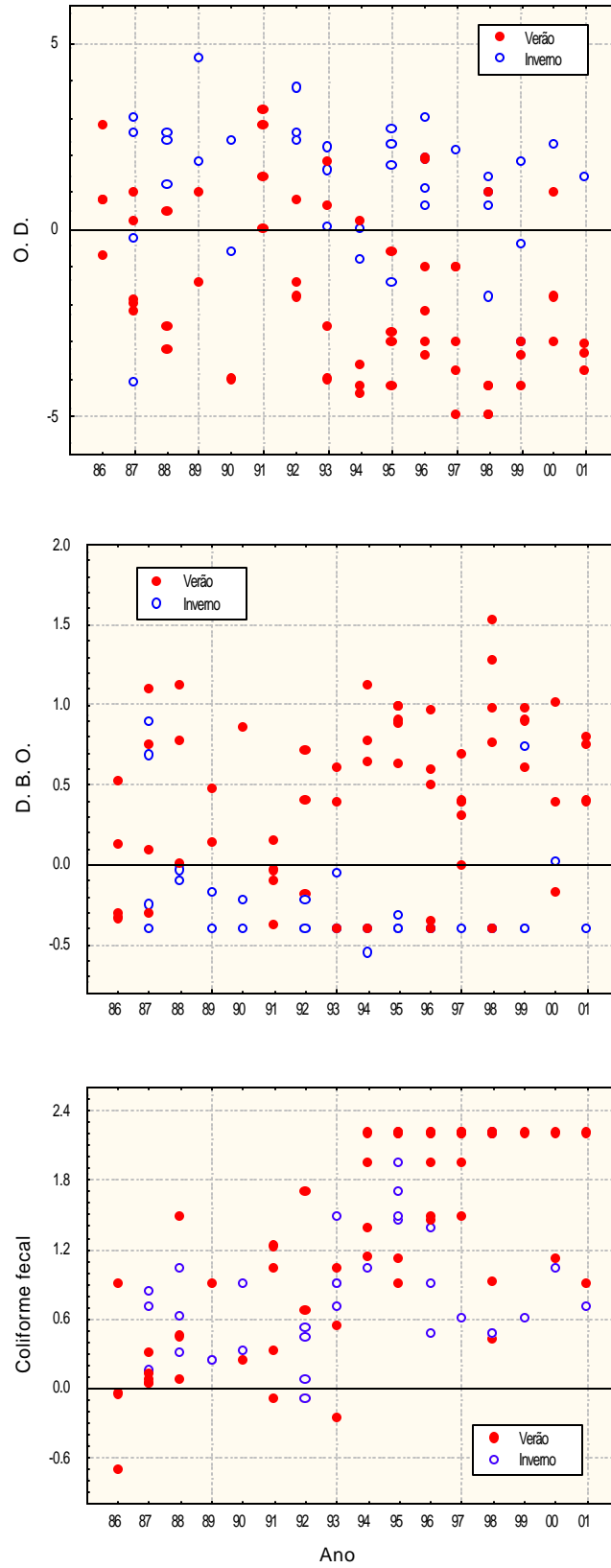


Figura 14 – Evolução temporal dos dados do Rio Ipojuca– Estação 70.

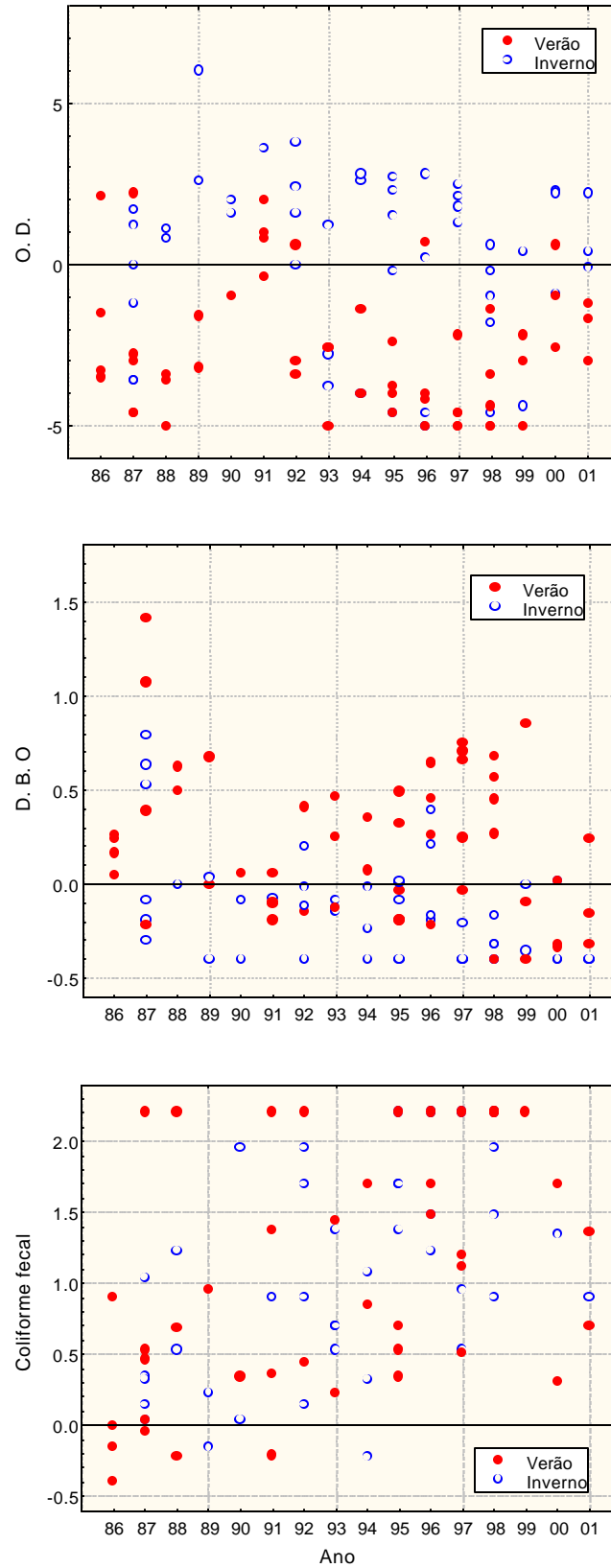


Figura 15 – Evolução temporal dos dados do Rio Ipojuca– Estação 85.

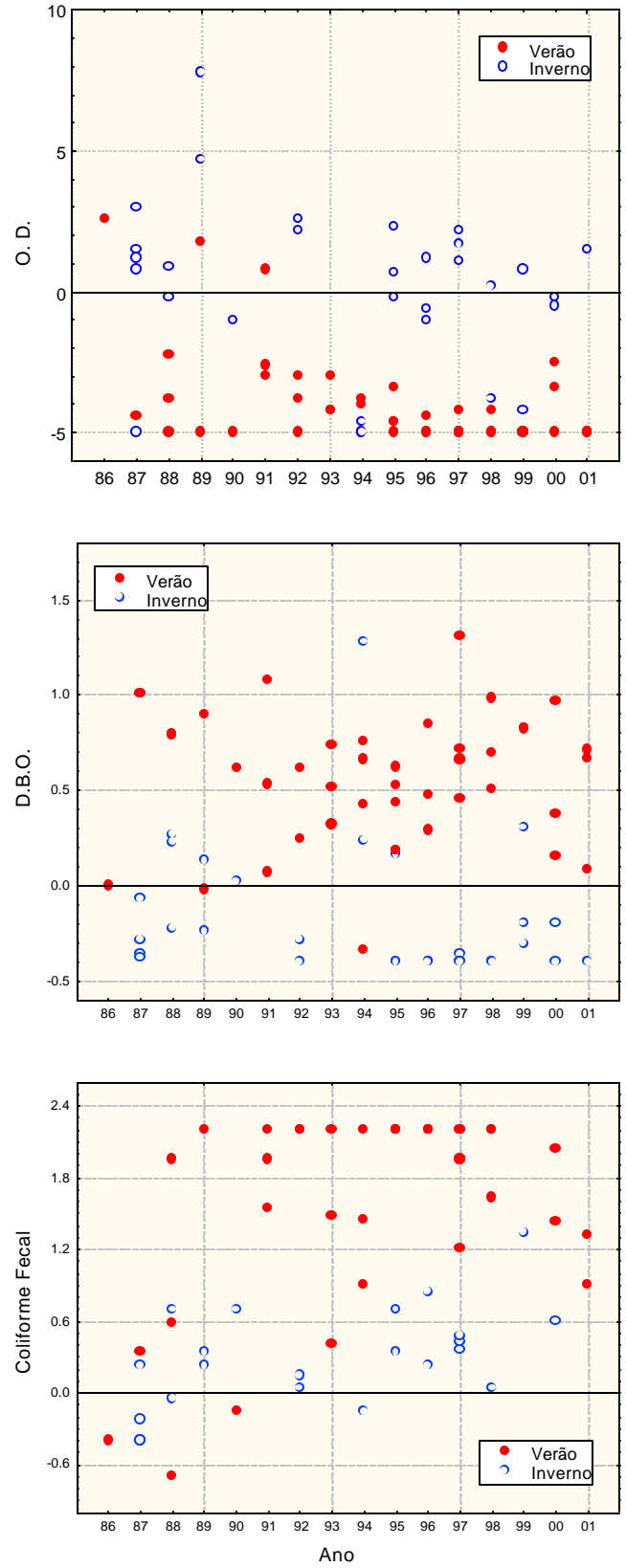


Figura 16 – Evolução temporal dos dados do Rio Ipojuca – Estação 95.

Numa análise subsequente, realizamos um estudo das correlações entre as cinco variáveis mais representadas no conjunto de dados. Várias correlações significativas foram identificadas, mas é importante salientar que todo cuidado é necessário na sua interpretação, já que esta é uma medida da associação *linear* entre duas variáveis, altamente sensível a pontos anômalos. Às vezes, um único ponto bastante afastado de um conjunto de outros pontos, possivelmente ocasionado por um mero erro de digitação, pode produzir correlações estatisticamente muito significativas, mas que de nada valem na prática. Considerando, por exemplo os valores de D. B. O. e O. D. para a Estação 50, calculamos um coeficiente de correlação de $-0,46$, que foi baseado em 91 pares de valores e mostrou-se estatisticamente significativo no nível de 95% de confiança. O diagrama matricial e os histogramas (isto é, diagramas de frequência) da Figura 17, porém, revelam que esse valor é altamente influenciado por pontos extremos nos dois eixos, e reflete na verdade duas distribuições muito assimétricas. Decidimos, conseqüentemente, associar a análise das correlações a um exame dos diagramas de dispersão das variáveis envolvidas.

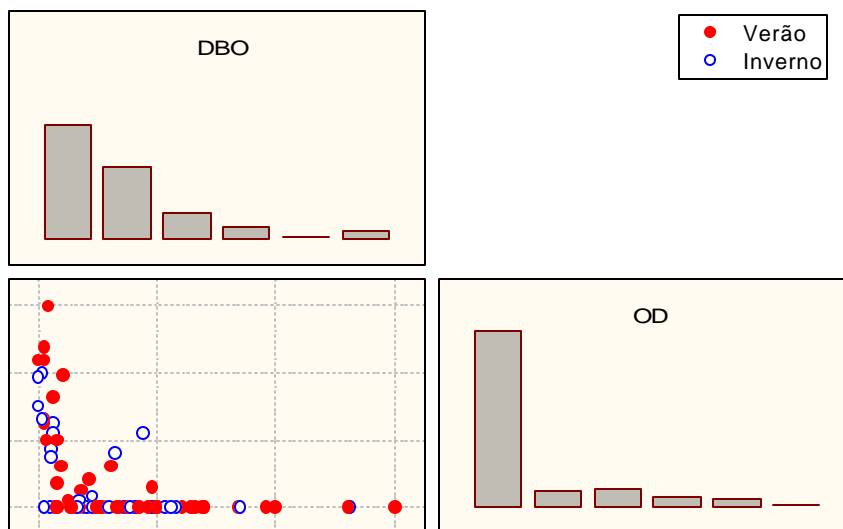


Figura 17 – Gráfico matricial de 91 pares de valores de O. D. e D. B. O. para a estação 50.

Os resultados numéricos e gráficos são mostrados nas Tabelas 2-6 e Figuras 18–22, nas páginas seguintes, onde os dados estão separados por estação de coleta, e estratificados pela distinção sazonal inverno/verão. O Coliforme fecal está representado pelo seu logaritmo, e as demais variáveis pelos valores originais.

- Para a estação 50, dois coeficientes de correlação, ambos positivos, mostraram-se significativos: Temperatura vs. Coliforme e Temperatura vs. pH. No entanto, levando em conta a aparência dos diagramas de dispersão da Figura 18, devemos encarar esses valores com desconfiança. No caso do pH, deve-se observar inclusive que a correlação positiva contrasta com as correlações negativas observadas nas estações 70, 85 e 95.
- A análise dos resultados para a estação 65/64 leva a conclusões semelhantes, embora neste caso deva-se ressaltar que na verdade estamos combinando dados de duas estações de coleta. De qualquer modo, a única correlação significativa observada, de 0,51 entre o oxigênio dissolvido e o pH, é consistente com os valores registrados nas demais estações.
- Nas estações 70 e 95 os padrões de associação são muito mais definidos, e mais fáceis de interpretar. Todos os coeficientes de correlação apresentaram valores significativos, apontando para as mesmas conclusões, que podemos associar à transição inverno-verão. Com a chegada do verão, como é óbvio, a temperatura média da água tende a subir. Ao mesmo tempo, como indicam as correlações, o número de coliformes fecais e a D. B. O. também tendem a elevar-se (correlações positivas com a temperatura), enquanto o oxigênio dissolvido e o pH tendem a reduzir-se (correlações negativas). Esta análise também permite explicar as demais correlações exibidas pelo pH: positiva com o oxigênio dissolvido, e negativa com o Coliforme e a D. B. O. O mesmo argumento pode ser aplicado às correlações negativas O. D. vs. D. B. O. e O. D. vs. Coliforme, bem como à correlação positiva entre D. B. O. e Coliforme. Com um gráfico tridimensional, podemos visualizar simultaneamente as correlações entre três pares de variáveis quaisquer, como ilustrado na Figura 23.
- Para a estação 85, que se localiza entre as estações 70 e 95, o quadro é essencialmente o mesmo que acabamos de descrever. As correlações pouco significativas apresentadas pelo pH podem ser atribuídas à presença de um ponto anômalo, com um pH excepcionalmente baixo, como se percebe claramente na parte inferior da Figura 21. Esse ponto corresponde a uma amostra colhida em outubro de 2001, que apresentou um pH de 3,6. Se excluirmos da análise esse único ponto, o coeficiente de correlação entre o pH e a D. B. O. torna-se significativo e com um valor de $-0,60$, o que é consistente com o observado nas estações vizinhas.
- A ausência de correlação significativa entre o Coliforme e a Temperatura na estação 70 pode ser uma consequência fortuita do número relativamente pequeno de amostras contendo valores para ambas as variáveis – apenas 32, de um total de

108. Para os valores usados no cálculo, de fato não parece haver ligação entre essas duas variáveis, como podemos ver pelo diagrama correspondente na Figura 20.

- É interessante contrastar o comportamento observado nas estações 70, 85 e 95, onde a influência climatológica é clara, com o observado nas outras duas estações, 50 e 65/64. Nestas duas, que se localizam mais a montante do Rio, praticamente não se nota a sazonalidade. Ressalvado o que já foi dito a respeito da restrição no volume de dados cujas correlações foram computadas, podemos especular que a diferença pode estar associada à atividade sucroalcooleira que é praticada sobretudo no trecho onde estão as estações 70, 85 e 95, e que é sabidamente de natureza sazonal. A variação de temperatura por si só não parece ser o principal fator, já que sua amplitude é praticamente a mesma em todas as cinco estações analisadas.

Tabela 2 – Coeficientes de correlação entre as cinco variáveis mais freqüentemente medidas na Bacia do Rio Ipojuca – Estação 50. Os valores em vermelho são estatisticamente significativos no nível de 95% de confiança. *N* representa o número de pares de valores válidos para as cinco variáveis.

<i>N</i> = 23	D. B. O.	O. D.	Coliforme	Temperatura
D. B. O.				
O. D.	-0.40			
Coliforme	-0.02	0.04		
Temperatura	-0.26	0.11	0.44	
pH	0.08	0.18	0.28	0.56

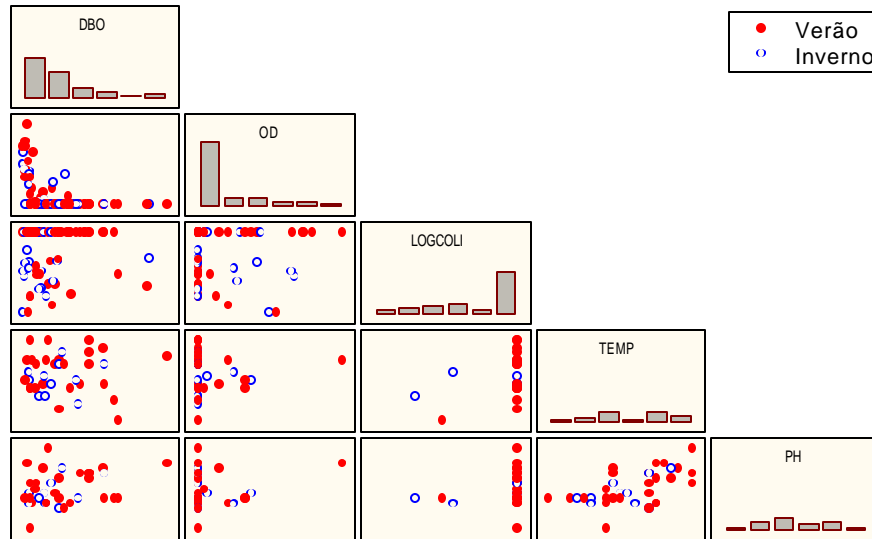


Figura 18 – Gráfico matricial das cinco variáveis mais freqüentes nos dados do Rio Ipojuca – Estação 50.

Tabela 3 – Coeficientes de correlação entre as cinco variáveis mais frequentemente medidas na Bacia do Rio Ipojuca – Estação 65/64. O valor em vermelho é estatisticamente significativo no nível de 95% de confiança. *N* representa o número de pares de valores válidos para as cinco variáveis.

<i>N</i> = 26	D. B. O.	O. D.	Coliforme	Temperatura
D. B. O.				
O. D.	-0.32			
Coliforme	0.11	-0.09		
Temperatura	0.16	0.06	0.20	
pH	0.04	0.51	0.15	0.25

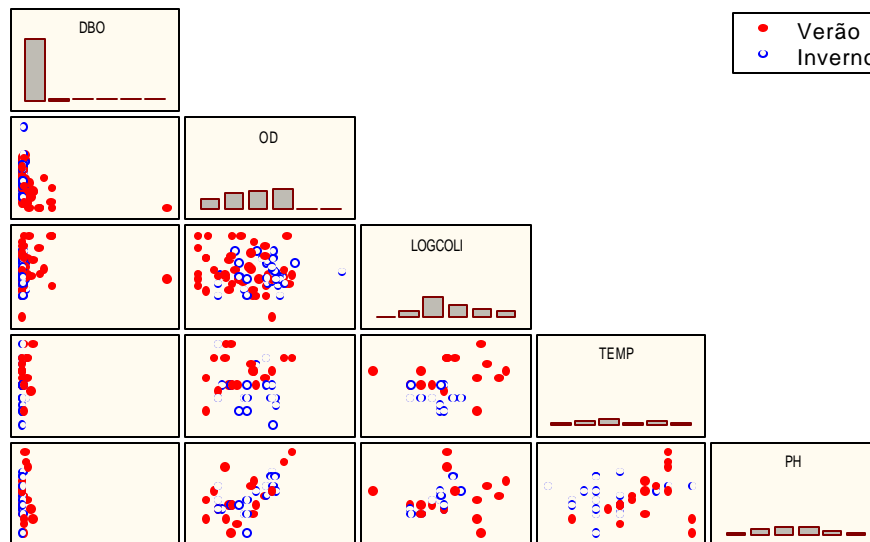


Figura 19 – Gráfico matricial das cinco variáveis mais frequentes nos dados do Rio Ipojuca – Estação 65/64.

Tabela 4 – Coeficientes de correlação entre as cinco variáveis mais frequentemente medidas na Bacia do Rio Ipojuca – Estação 70. Todos os valores são estatisticamente significativos no nível de 95% de confiança. *N* representa o número de pares de valores válidos para as cinco variáveis.

<i>N</i> = 29	D. B. O.	O. D.	Coliforme	Temperatura
D. B. O.				
O. D.	-0.60			
Coliforme	0.42	-0.52		
Temperatura	0.54	-0.85	0.50	
pH	-0.77	0.72	-0.54	-0.58

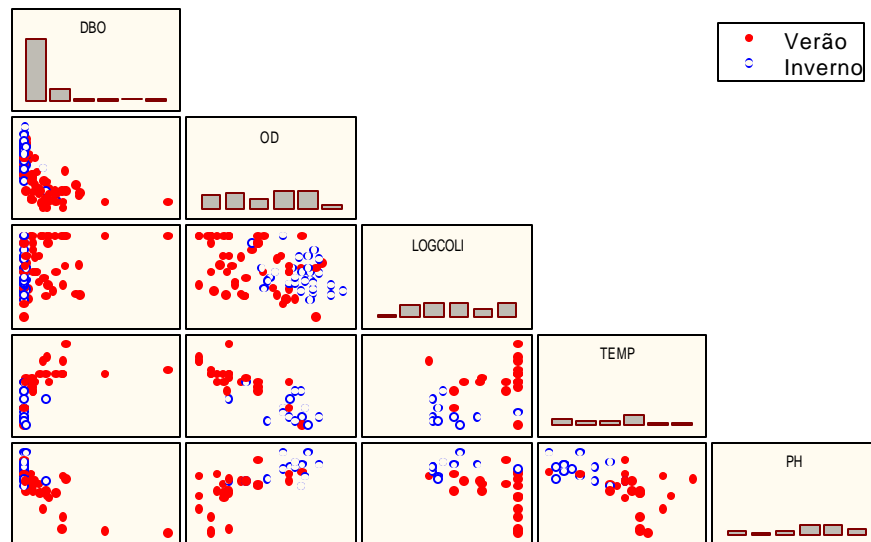


Figura 20 – Gráfico matricial das cinco variáveis mais frequentes nos dados do Rio Ipojuca – Estação 70.

Tabela 5 – Coeficientes de correlação entre as cinco variáveis mais freqüentemente medidas na Bacia do Rio Ipojuca – Estação 85. Os valores em vermelho são estatisticamente significativos no nível de 95% de confiança. *N* representa o número de pares de valores válidos para as cinco variáveis.

<i>N</i> = 32	D. B. O.	O. D.	Coliforme	Temperatura
D. B. O.				
O. D.	-0.75			
Coliforme	0.52	-0.36		
Temperatura	0.64	-0.85	0.30	
pH	-0.25	0.36	0.06	-0.39

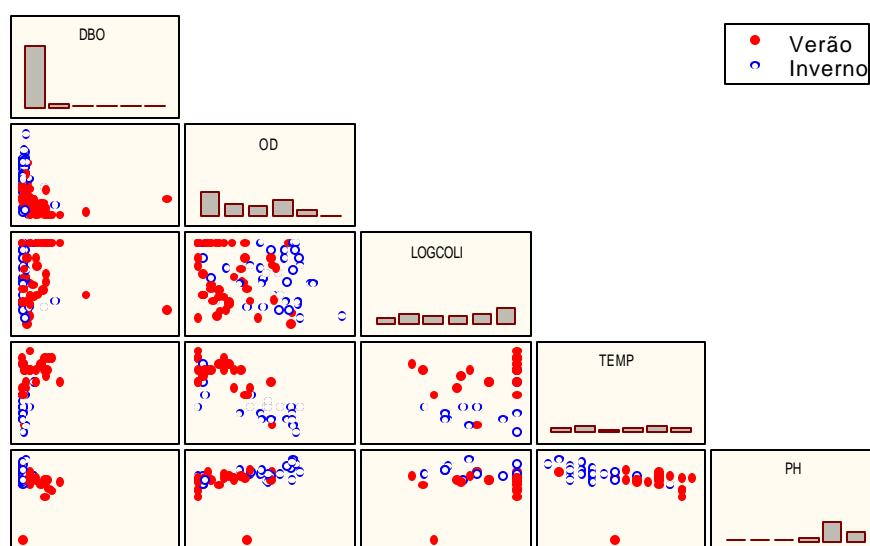


Figura 21 – Gráfico matricial das cinco variáveis mais freqüentes nos dados do Rio Ipojuca – Estação 85.

Tabela 6 – Coeficientes de correlação entre as cinco variáveis mais frequentemente medidas na Bacia do Rio Ipojuca – Estação 95. Todos os valores são estatisticamente significativos no nível de 95% de confiança. *N* representa o número de pares de valores válidos para as cinco variáveis.

<i>N</i> = 23	D. B. O.	O. D.	Coliforme	Temperatura
D. B. O.				
O. D.	-0.54			
Coliforme	0.61	-0.86		
Temperatura	0.72	-0.79	0.87	
pH	-0.83	0.67	-0.63	-0.65

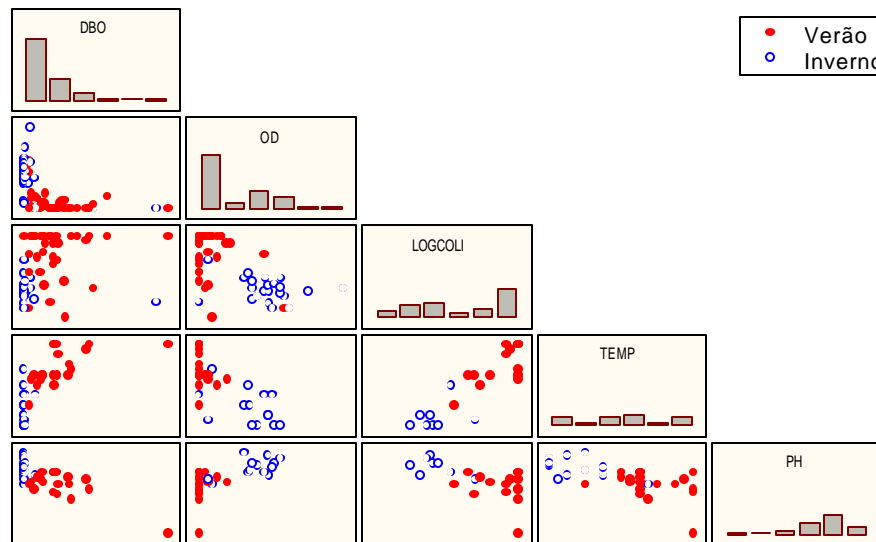


Figura 22 – Gráfico matricial das cinco variáveis mais frequentes nos dados do Rio Ipojuca – Estação 95.

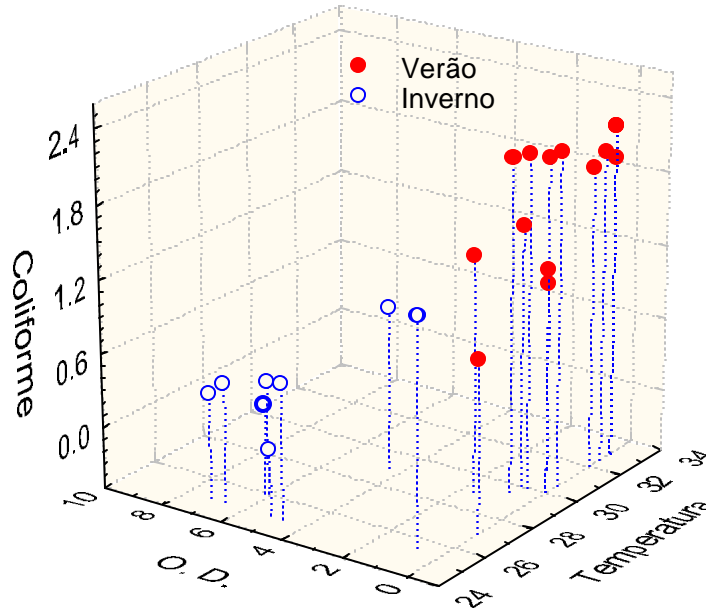


Figura 23 – Gráfico tridimensional para Temperatura, Coliforme e Oxigênio dissolvido. Dados do Rio Ipojuca – Estação 95.

Análise dos dados do Reservatório Tapacurá

O conjunto de dados do Reservatório Tapacurá apresenta duas importantes diferenças em relação ao conjunto da Bacia do Ipojuca. Em primeiro lugar, o número de amostras é muito menor: apenas 35, cobrindo o período 1999-2002. Em segundo lugar, não existem mais diferentes estações de coleta, o que deixa como possibilidade mais óbvia de estratificação apenas a baseada na distinção entre os períodos chuvoso (abril a julho) e seco (o resto do ano). Já com um número de casos bastante reduzido, o conjunto de Tapacurá ainda contém várias lacunas, como mostra a Figura 24. Esses dois fatores, como sabemos, restringem a abrangência das conclusões de qualquer análise estatística.

Mesmo com estas restrições, as séries temporais para algumas variáveis permitem-nos identificar algumas flutuações sazonais ocorridas no período em análise. As Figuras 25–27 mostram a evolução temporal de algumas variáveis nesses quatro anos. A estratificação permite identificar os dados de acordo com a estação pluviométrica, com os pontos cheios correspondendo ao período chuvoso. A água do reservatório Tapacurá é enquadrada na Classe 3.

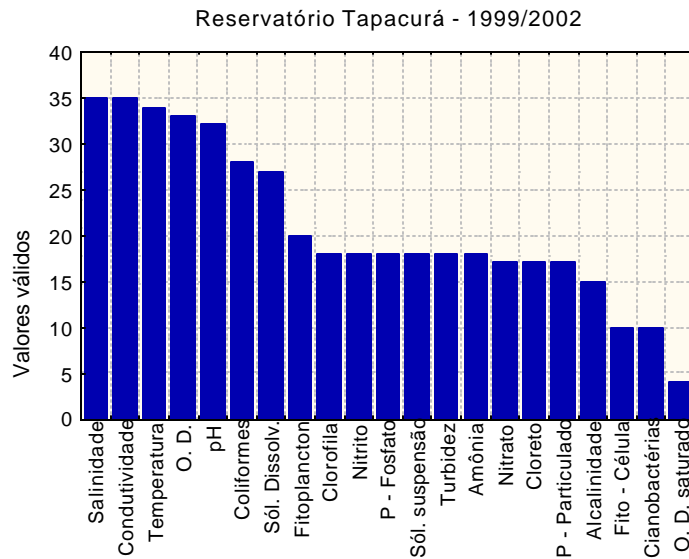


Figura 24 – Valores válidos para os parâmetros monitorados no Reservatório Tapacurá.

Na figura 25 percebe-se claramente o aumento da salinidade associado à seca de 1999, e a redução de patamar ocorrida nesse parâmetro a partir de 2000, como consequência do aumento do volume de água acumulada a partir de então. A evolução temporal da condutividade e dos sólidos dissolvidos apresenta o mesmo padrão, como seria de se esperar, e não é mostrada aqui.

A Figura 26 mostra o comportamento do oxigênio dissolvido. A maioria das amostras satisfaz o enquadramento na Classe 3. Das cinco exceções, quatro ocorreram no período chuvoso.

O comportamento do pH é apresentado na Figura 27, onde se pode observar que todas as amostras estão enquadradas no intervalo da Classe 3 (6 – 9). Parece haver também um ciclo constituído por uma elevação do pH ao longo da estação seca, seguida de um decréscimo após a chegada da estação chuvosa, bem como uma discreta tendência de elevação do pH médio ao longo de todo o período. O número de pontos, porém, não é suficiente para dar muita convicção estatística a essas observações.

Para as demais variáveis não nos foi possível detectar padrões temporais, possivelmente em consequência da relativa escassez de dados.

Na análise das correlações, tivemos de investigar um par de variáveis de cada vez, porque o conjunto de dados contém poucos trechos continuamente preenchidos para vários parâmetros. Passamos a comentar os resultados que mais nos chamaram a atenção.

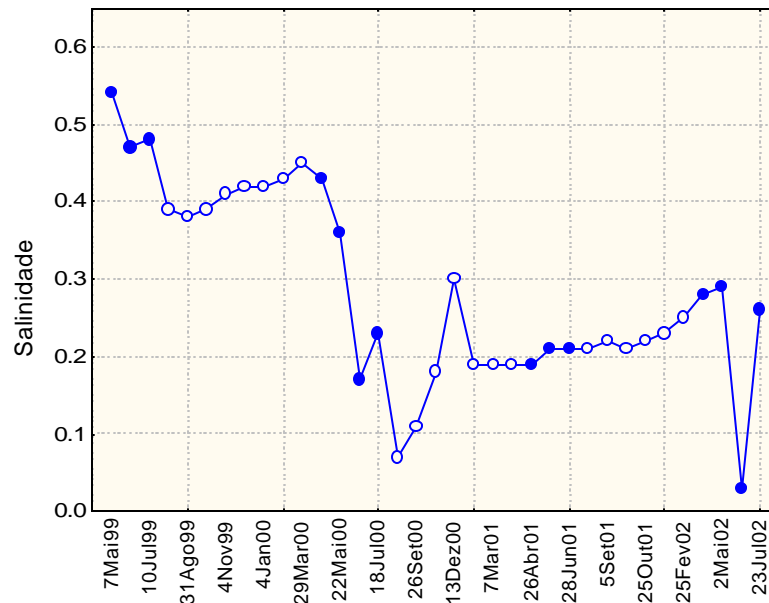


Figura 25 – Evolução temporal da salinidade no Reservatório Tapacurá, no período 1999-2002. Os pontos cheios correspondem ao período chuvoso.

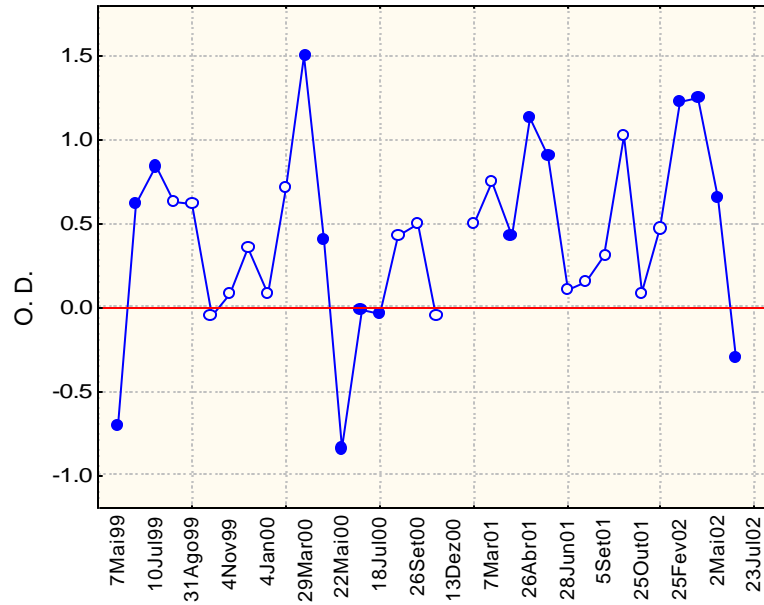


Figura 26 – Evolução temporal do oxigênio dissolvido no Reservatório Tapacurá, no período 1999-2002. Os pontos cheios correspondem ao período chuvoso. A reta indica o limite da Classe 3, codificado.

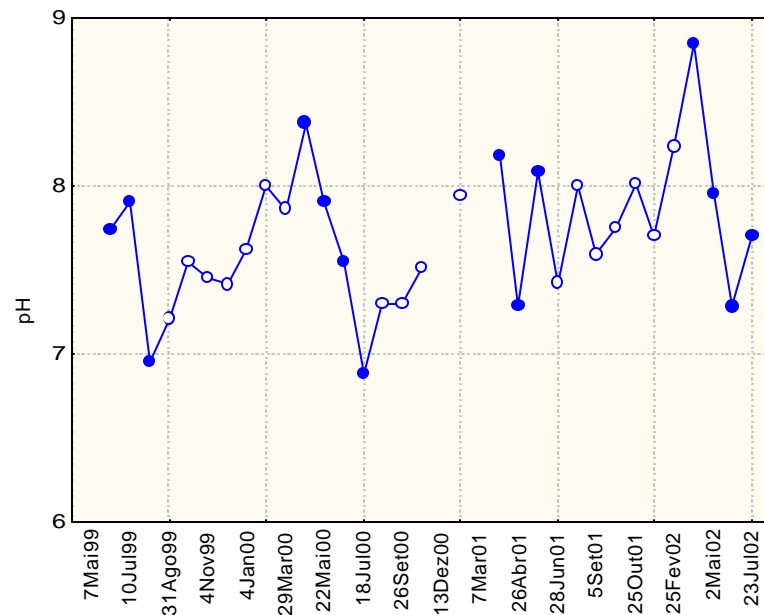


Figura 27 – Evolução temporal do pH no Reservatório Tapacurá, no período 1999-2002. Os pontos cheios correspondem ao período chuvoso.

A Figura 28 ilustra a correlação entre a temperatura da água e o oxigênio dissolvido. A correlação é positiva (0,40, para 32 amostras) e significativa no nível de 95% de confiança. Este é um resultado que nos parece curioso, uma vez que na análise dos dados do Rio Ipojuca o que observamos foram correlações *negativas* entre essas variáveis.

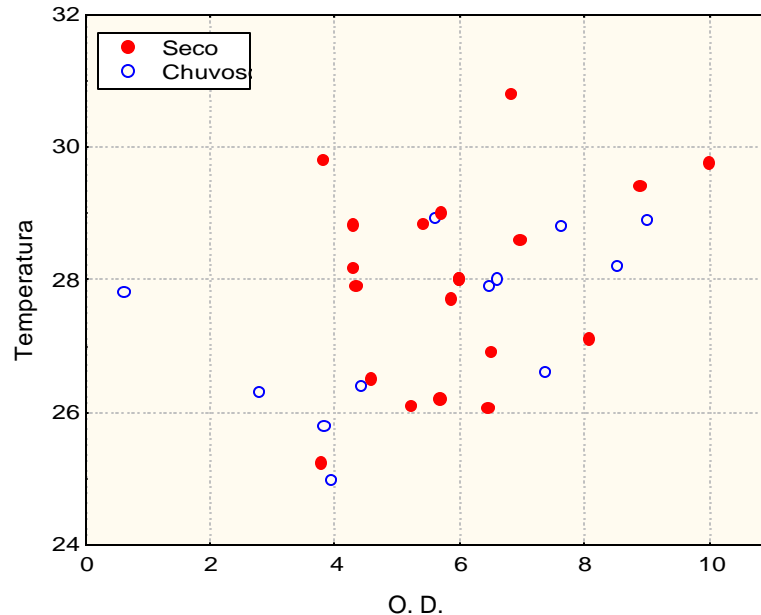


Figura 28 – Diagrama de dispersão para a temperatura e o oxigênio dissolvido, no Reservatório Tapacurá.

A correlação entre cloreto e salinidade é positiva, como seria de se esperar, mas chama a atenção pela presença de um ponto anômalo, que provavelmente se deve a um erro de registro, até porque se trata de uma amostra colhida na estação chuvosa (Figura 29). Esse único ponto, aliás, faz com que o valor do coeficiente de correlação deixe de ser significativo.

As correlações entre Nitrato, Nitrito, Amônia e Fósforo, de um lado, e Clorofila, Matéria orgânica e Coliformes, de outro, também foram investigadas. Apenas três coeficientes de correlação, nesse grupo, revelaram-se estatisticamente significativos. Um deles, entre o Nitrato e o Coliforme (0,97, 11 casos), é claramente o resultado da influência de dois pontos anômalos (Figura 30), pelo menos neste conjunto de dados. Os outros envolvem a Clorofila, a Amônia e o Fósforo particulado, e são ilustrados pelos diagramas das Figuras 31 e 32. Esses talvez justifiquem uma análise mais detalhada.

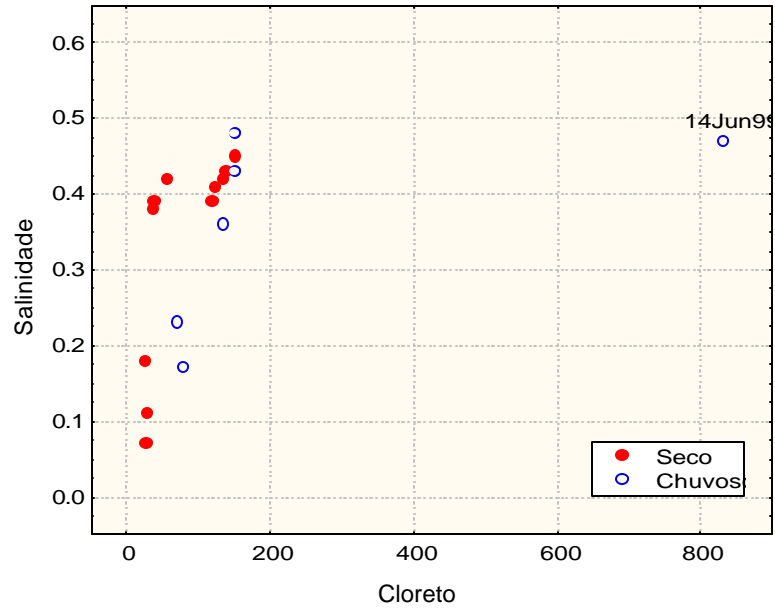


Figura 29 – Diagrama de dispersão para a Salinidade e o Cloreto no Reservatório Tapacurá.

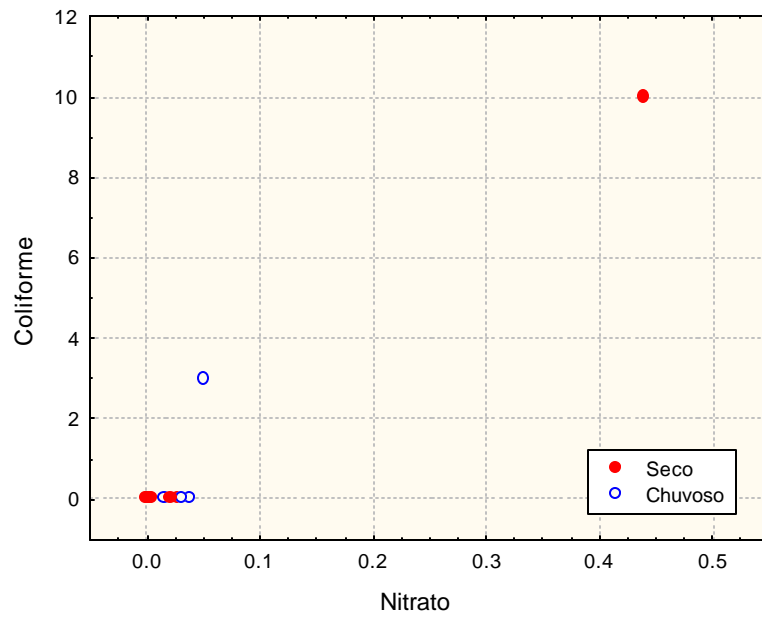


Figura 30 – Diagrama de dispersão para o Coliforme e o Nitrato, no Reservatório Tapacurá.

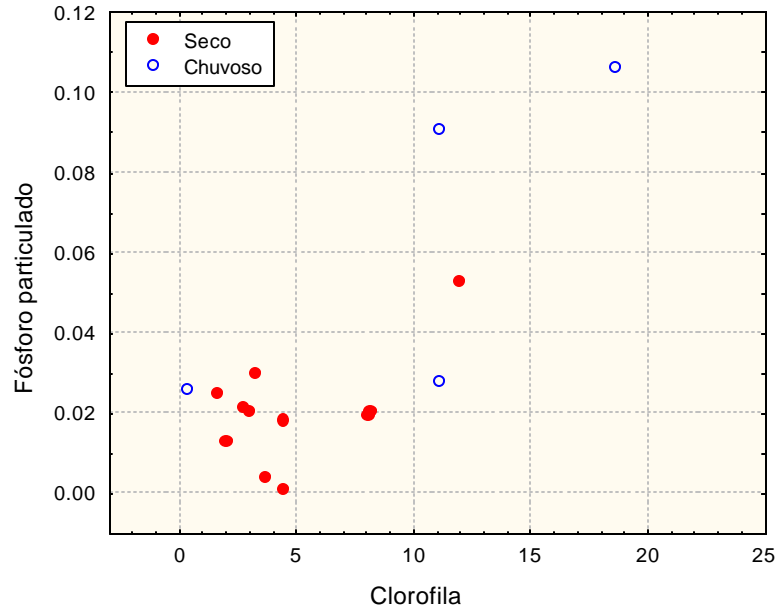


Figura 31 – Diagrama de dispersão para a Clorofila e o Fósforo particulado, no Reservatório Tapacurá.

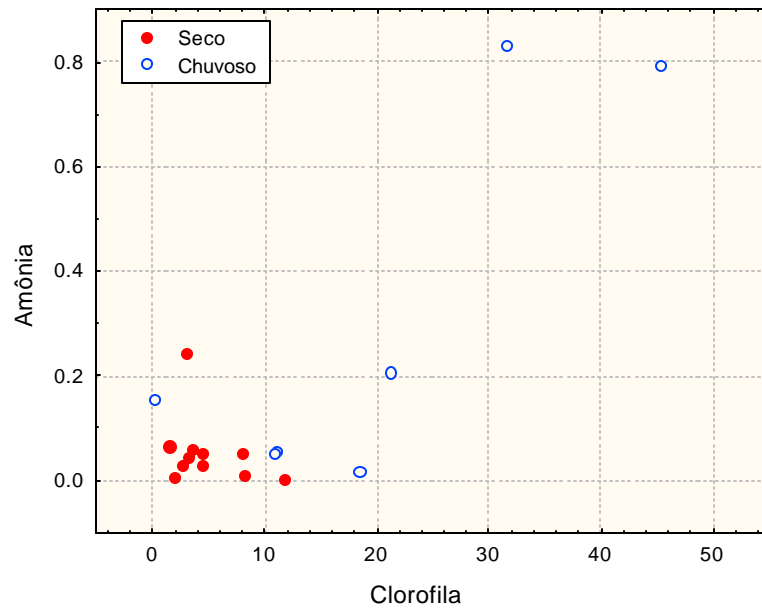


Figura 32 – Diagrama de dispersão para a Clorofila e a Amônia, no Reservatório Tapacurá.

Treinamento

Como previsto no plano de trabalho, todas as atividades contaram com a participação dos técnicos Helena Alves Moreira (CPRH) e Luciano Costa Almeida (SRH), que nesta etapa receberam treinamento nos seguintes procedimentos, no ambiente do programa computacional *Statistica 5.5*:

- Construção de subconjuntos de dados obedecendo a critérios de seleção pré-estabelecidos
- Reordenamento hierárquico de casos de acordo com os valores de variáveis de interesse
- Cálculo de várias estatísticas descritivas
- Construção de gráficos seqüenciais para variáveis e casos (gráficos de linha)
- Construção de gráficos de caixa, gráficos matriciais e gráficos tridimensionais
- Estratificação de conjuntos de dados por variável e por varredura
- Rotulação de gráficos de acordo com os valores de variáveis de interesse
- Modificação dos elementos tipográficos e da aparência de um gráfico qualquer

Recife, 3 de janeiro de 2003.

Maria Fernanda Pimentel

Programa Nacional do Meio Ambiente II – PNMA II

Componente: Desenvolvimento Institucional

Subcomponente: Monitoramento da Qualidade da Água

Projeto: Monitoramento da Qualidade da Água como Instrumento de Controle Ambiental e Gestão de Recursos Hídricos no Estado de Pernambuco

**Análise Estatística de Dados
do Monitoramento da Qualidade das Águas
do Rio Ipojuca e do Reservatório Tapacurá**

Produto 4 – Relatório 3

Maria Fernanda Pimentel

Recife, abril de 2003

Introdução

O presente relatório se refere ao desenvolvimento das atividades 4 e 5 previstas no plano de trabalho, a saber:

4. Identificação de subconjuntos de variáveis e amostras para a realização da análise estatística multivariada.
5. Análise estatística multivariada, através das técnicas de projeções em componentes principais e análise de agrupamentos hierárquica, visando identificar padrões de associação ou semelhança entre amostras ou variáveis.

No que segue passamos a descrever o que foi feito nesses dois itens.

Atividade 4

Para a realização de uma análise multivariada qualquer, os conjuntos de dados não podem conter lacunas. Como consequência, tivemos de restringir o conjunto de dados do Rio Ipojuca de 842 para 331 amostras, representando as nove estações onde a coleta é feita atualmente, e varrendo o período de 1996 a 2002. Dessas amostras, 186 foram coletadas durante o verão, e as outras 145 no período de inverno. A análise multivariada foi realizada com as seguintes variáveis: Temperatura, D. B. O., O. D., Cloreto, pH e Condutividade (esta, após transformação logarítmica). As demais variáveis foram excluídas da análise pela escassez de valores. Tentativas de incluir mais variáveis foram feitas, mas acarretaram numa significativa diminuição do número de amostra válidas, o que por sua vez torna os resultados estatísticos menos representativos. Para efeito de análise gráfica e de possível identificação de zonas homogêneas, as estações de coleta foram agrupadas por zona predominante de atividade, de acordo com a seguinte correspondência:

- Estações 25 e 50: Zona Urbana/Industrial;
- Estações 40 e 55: Zona de Atividade Pecuária;
- Estação 64: Zona de Transição da Pecuária para Agroindústria, com predomínio da agricultura;
- Estações 70, 85, 90 e 95: Zona Agroindustrial.

Para o Reservatório de Tapacurá, as variáveis selecionadas, também com base na quantidade de valores disponíveis, foram as seguintes: Temperatura, O. D., O. D. saturado, Salinidade, pH, Sólidos dissolvidos e Coliformes fecais (esta, mais uma vez, após transformação logarítmica). O número de casos válidos foi de apenas 35, cobrindo os anos de 1999 a 2002.

Atividade 5

A análise de componentes principais (PCA) é uma técnica de compressão da informação que consiste em projetar o máximo de informação no menor número possível de eixos não correlacionados. Esses eixos são combinações lineares das variáveis originais, com pesos escolhidos de forma a maximizar a variância, que em Estatística é tomada como uma medida da informação. Assim, a primeira componente principal (PC1) representa a direção de máxima variância (e portanto de máxima informação) no espaço multidimensional em que o conjunto de dados está representado. A segunda componente (PC2) é ortogonal à primeira e descreve o máximo da informação ainda não descrita por PC1, e assim por diante.

Análise dos dados do Rio Ipojuca

As Figuras 1 – 5 mostram os resultados da análise de componentes principais (PCA) realizada com o conjunto de todas as 331 amostras do Rio Ipojuca. Nesta análise, como o conjunto de dados contém seis variáveis, cada amostra pode ser imaginada, em termos geométricos, como um ponto localizado num sistema de seis dimensões. A PCA visa projetar o máximo da informação contida nesse espaço multidimensional no mínimo de dimensões, o que permite, por inspeção dos gráficos produzidos na análise, detectar padrões de associação entre amostras ou variáveis.

A Figura 1 mostra o gráfico dos pesos nas duas primeiras componentes. Juntas, essas componentes explicam cerca de três quartos de toda a informação. A primeira componente (44,5% da informação) pode ser interpretada como um contraste entre, de um lado, o pH, o Cloreto, a Condutividade e, em menor grau, o O. D., que têm pesos positivos, e de outro a Temperatura e a D. B. O., com pesos negativos. A proximidade dos pontos que representam o Cloreto e a Condutividade indica que estas duas variáveis têm uma alta correlação positiva ($r = +0,86$). Da mesma forma, a localização do O. D. no quadrante oposto ao da temperatura e da D. B. O. indicam correlações negativas ($r = - 0,50$ e $r = - 0,49$), respectivamente. A Temperatura, na verdade, está negativamente correlacionada com todas as demais variáveis, exceto a D. B. O. A segunda componente (28,5% da informação) reflete sobretudo o contraste entre O. D. e D. B. O. Todas essas observações nos ajudam a interpretar o gráfico dos escores (Figura 2), que veremos a seguir, e onde estão representadas as amostras. Em PC1, amostras com altos valores de pH, Cloreto, Condutividade e O. D. e/ou baixos valores de Temperatura e D. B. O. tendem a localizar-se mais à direita do gráfico dos escores, e vice-versa. Analogamente, amostras com alto D. B. O. e/ou baixo O. D. tendem a localizar-se na parte superior, e vice-versa.

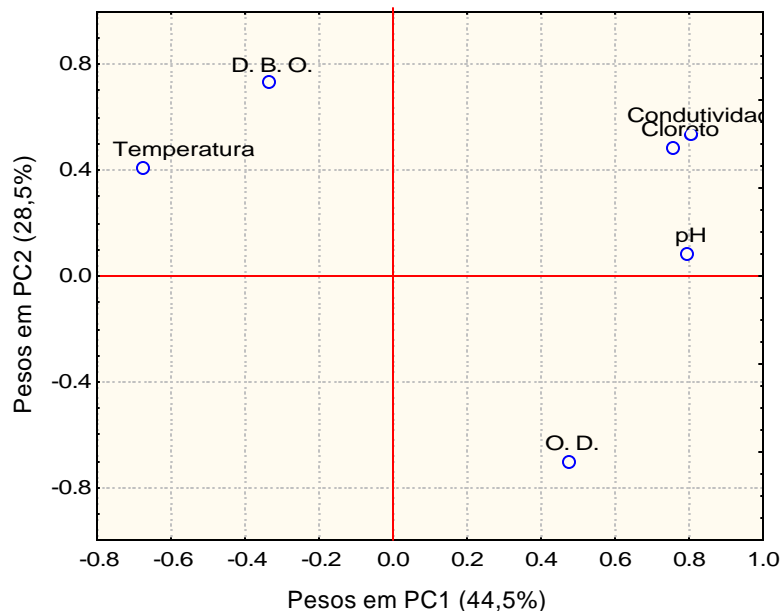


Figura 1 – Pesos nas duas primeiras componentes principais da análise de todos os dados juntos do Rio Ipojuca. As percentagens são as variâncias explicadas por cada PC.

A Figura 2 mostra os escores da PCA feita com todas as amostras. Podemos perceber que parte da variância em PC1 é devida à mudança inverno – verão. Os escores negativos, associados, como vimos, a valores mais elevados da Temperatura, são majoritariamente de amostra coletadas no verão (os pontos em vermelho). É conveniente, portanto, fazer análises separadas para as duas estações climáticas, para eliminar esse efeito.

Análise das amostras coletadas durante o verão

As Figuras 3-5 mostram os resultados da PCA com os dados coletados durante os meses de verão (186 amostras). O gráfico dos pesos (Figura 3) é muito semelhante ao da análise com o conjunto completo, e pode ser interpretado da mesma forma. O gráfico dos escores aparece nas Figuras 4 e 5, porém com estratificações diferentes. Na Figura 4 as amostras são identificadas por estação, enquanto na Figura 5 a identificação é feita por zona, conforme definido no início deste relatório. As quatro zonas ocupam regiões diferentes do gráfico, com alguma superposição parcial, além de alguns pontos anômalos. As zonas urbana e agroindustrial estão quase perfeitamente separadas, ao longo de PC1. A presença de um ponto da zona urbana/industrial entre os da área agroindustrial (indicado pela seta) pode ser explicado pelo baixíssimo valor registrado para o seu pH. Essa amostra foi coletada na Estação 25, em janeiro de 1996. As duas amostras da área agroindustrial que se localizaram entre as da zona urbana/industrial foram coletadas na estação 95, e têm valores

de cloreto de 1270 e 740 mg L⁻¹, (outubro de 1999 e fevereiro de 2000, respectivamente) além de O.D. igual a zero.

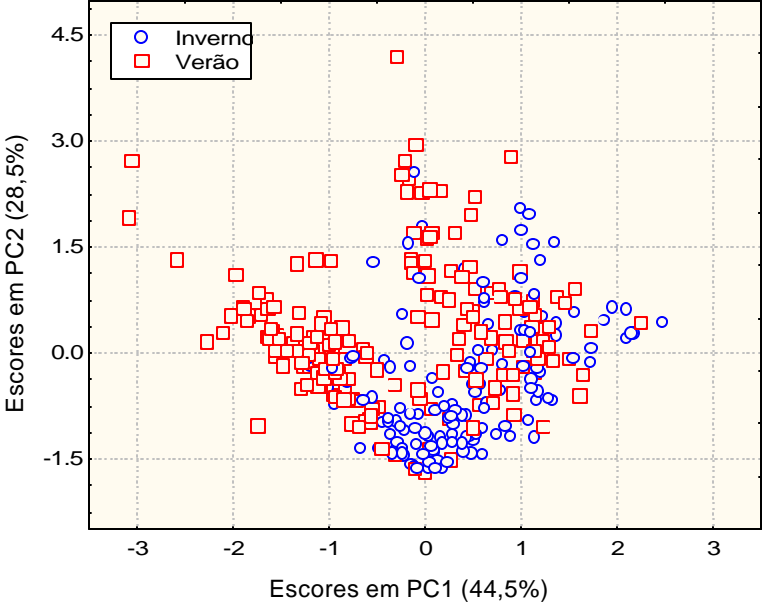


Figura 2 – Escores nas duas primeiras componentes principais da análise de todos os dados do Rio Ipojuca juntos. As percentagens são as variâncias explicadas por cada PC.

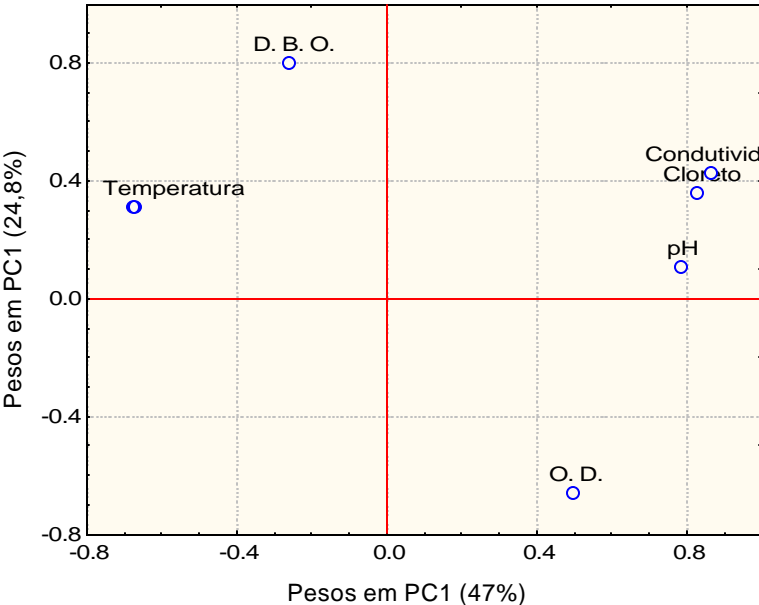


Figura 3 – Pesos nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca coletados no verão.

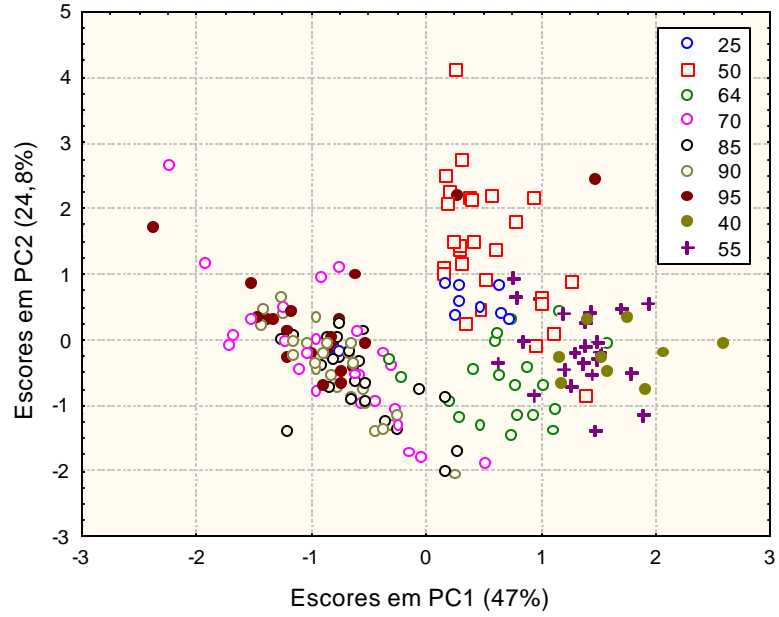


Figura 4 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no verão. Estratificação por estação de coleta

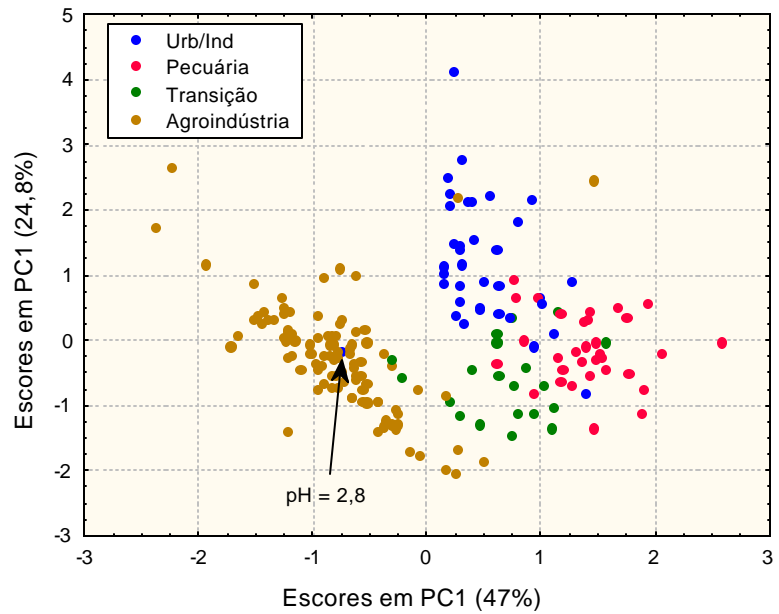


Figura 5 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no verão. Estratificação por zona de atividade principal.

A existência dessas zonas separadas é consequência das diferenças existentes entre os valores das variáveis usadas na PCA, como podemos comprovar pelos gráficos de caixa mostrados a seguir.

As Figuras 6 e 7 mostram o comportamento da D. B. O. e do oxigênio dissolvido, que se correlacionam negativamente, como já vimos. Nas zonas urbana/industrial e agroindustrial a D. B. O. é mais alta (e o O. D. mais baixo) do que nas outras duas zonas. No caso da zona urbana, a distribuição do oxigênio dissolvido chama a atenção pela sua enorme assimetria. A causa desse comportamento pode ser constatada na Figura 8, que mostra a série temporal do O. D. para a zona urbana. Quase todos os valores são iguais a zero, embora em novembro de 1996 tenha sido registrado um valor de $9,6 \text{ mg L}^{-1}$ na Estação 50.

Para a Temperatura (Figura 9), as medianas também são mais elevadas nas zonas agroindustrial e urbana/industrial.

No pH, duas características nos chamaram a atenção. Na progressão da zona urbana/industrial para a agroindustrial, que reflete aproximadamente o curso do Ipojuca em direção a sua foz, a Figura 10 sugere, pela análise das medianas, que o pH, ao menos durante o verão, vai-se tornado progressivamente mais ácido. O segundo aspecto a se notar é que as zonas urbana/industrial e agroindustrial apresentam maiores amplitudes de variação do pH, com pontos isolados de grande acidez, sendo um deles o valor de 2,8 já mencionado para a Estação 25.

Os padrões de variação do Cloreto e da Condutividade com a zona são muito semelhantes, de modo que vamos discutir apenas o comportamento do Cloreto (Figura 11). A maior mediana é a da zona de pecuária, e a menor a da zona agroindustrial, que no entanto apresenta dois valores extremamente altos, ambos registrados na Estação 95, a mais próxima da foz do Rio (Figura 12). Esses valores correspondem justamente às duas amostras que se localizaram entre as da zona urbana na Figura 5.

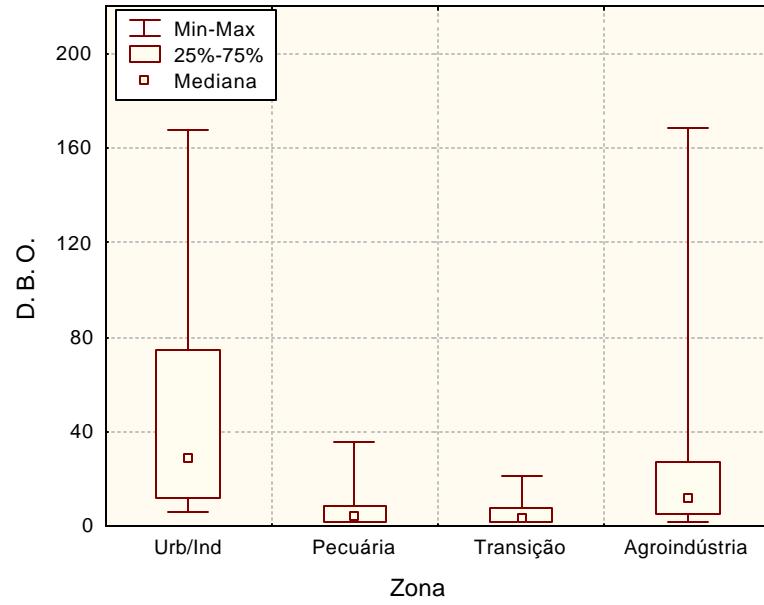


Figura 6 – Gráficos de caixa da D.B.O. no verão, por zona de atividade.

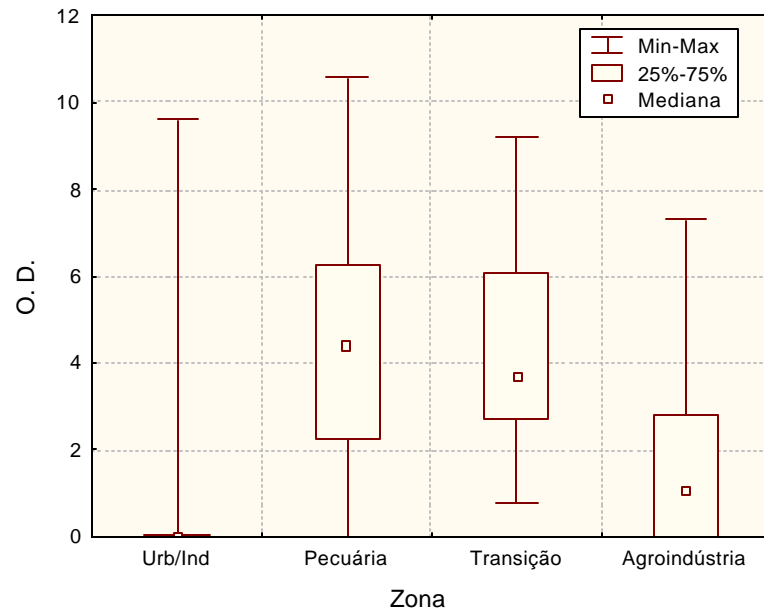


Figura 7 – Gráficos de caixa do O.D. no verão, por zona de atividade.

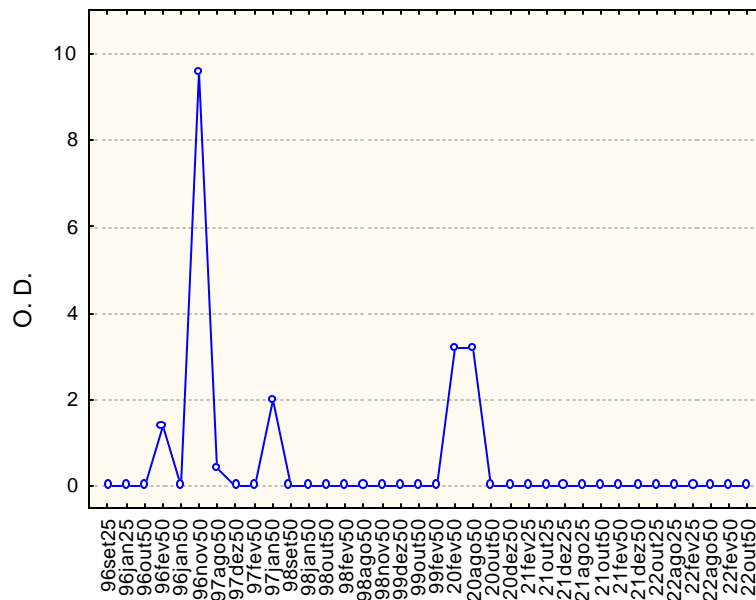


Figura 8 – Série temporal do O.D. no verão, para a zona urbana/industrial.

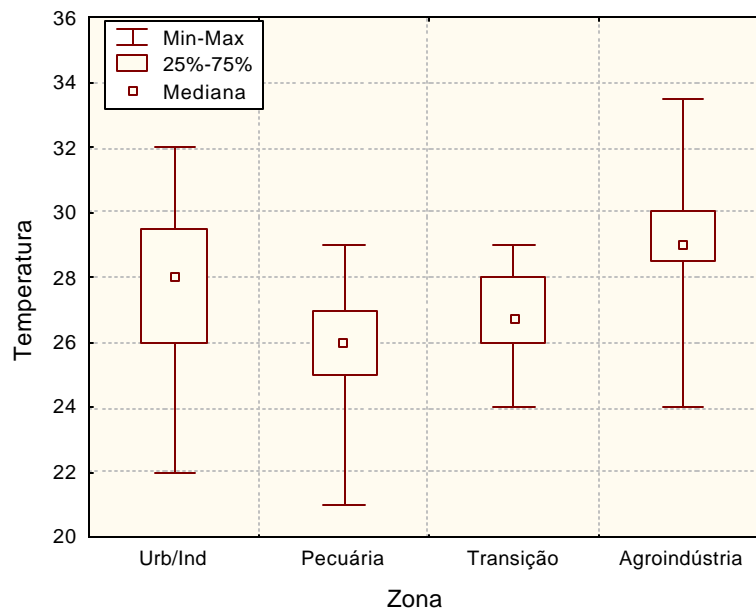


Figura 9 – Gráficos de caixa da Temperatura no verão, por zona de atividade.

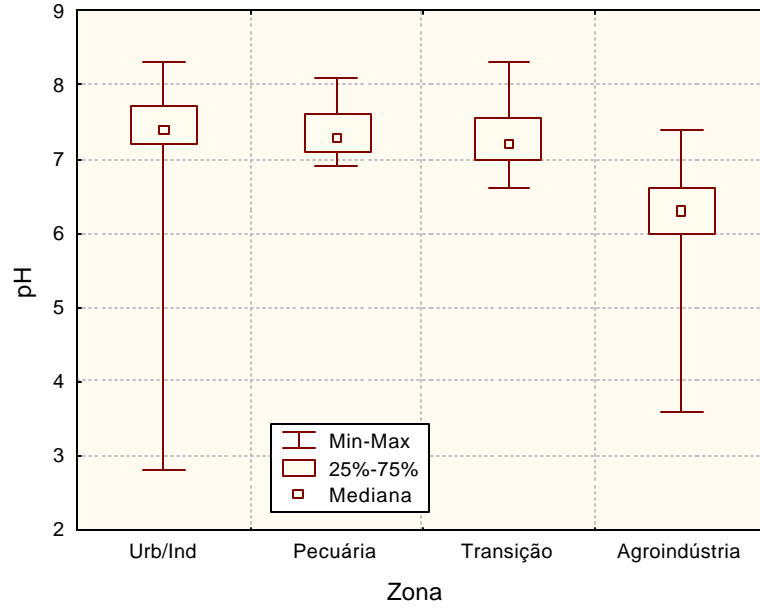


Figura 10 – Gráficos de caixa do pH no verão, por zona de atividade.

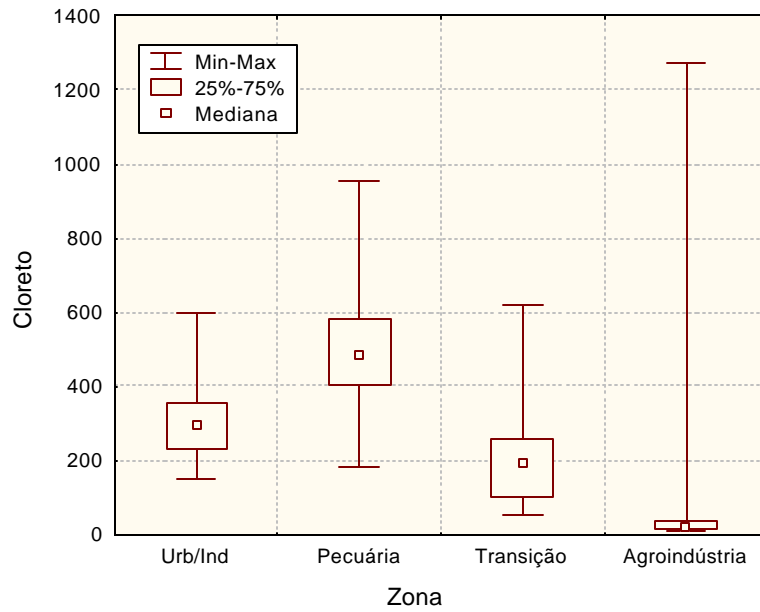


Figura 11 – Gráficos de caixa do Cloreto no verão, por zona de atividade.

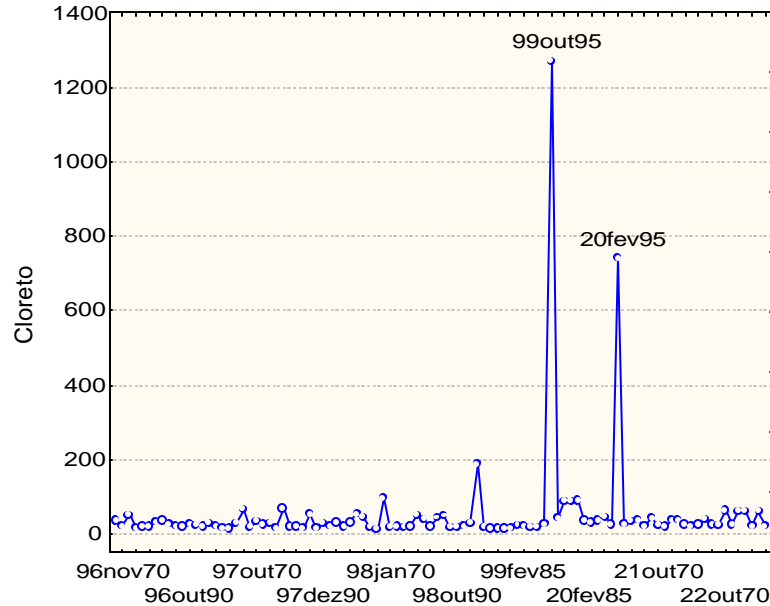


Figura 12 – Série temporal do Cloreto no verão, para a zona agroindustrial.

Análise das amostras coletadas durante o inverno

As duas primeiras componentes representam cerca de 70% da informação total. O gráfico dos pesos tem um padrão semelhante aos que já vimos, devendo-se ressaltar apenas que agora a contribuição da temperatura está praticamente ausente da primeira componente, passando a figurar apenas na segunda. Como consequência, nota-se um certo contraste entre o oxigênio dissolvido, que agora tem peso negativo em PC1, e a dupla Cloreto/Condutividade. Os demais padrões de correlação permanecem os mesmos que já discutimos. Nos gráficos dos escores (Figuras 14 e 15), o que se observa é semelhante ao observado na análise das amostras do verão. Ao longo de PC1 há uma separação quase perfeita entre as zonas agroindustrial, à esquerda, e urbana, mais à direita, com a zona de transição situando-se entre as duas primeiras, e a zona de pecuária ocupando o quadrante inferior direito. Mais uma vez, algumas poucas amostras da zona agroindustrial localizam-se entre as da zona urbana, e pelo mesmo motivo de antes: valores excepcionalmente elevados para o cloreto nas estações próximas à foz do Rio. Comparando as duas séries temporais para o Cloreto na zona agroindustrial, apresentadas nas Figuras 12 e 16, podemos também observar que no inverno os valores foram em geral mais elevados (numericamente: medianas de 24 mgL⁻¹ para o verão e de 50 mgL⁻¹ para o inverno).

Como na análise anterior, os padrões dos escores podem ser interpretados mais detalhadamente levando-se em conta os gráficos de caixa estratificados por zona. Para a D.

B. O. e o O. D., as condições parecem em geral mais favoráveis no inverno do que no verão. Chama a atenção, especificamente, a alta mediana registrada para o O. D. na área agroindustrial.

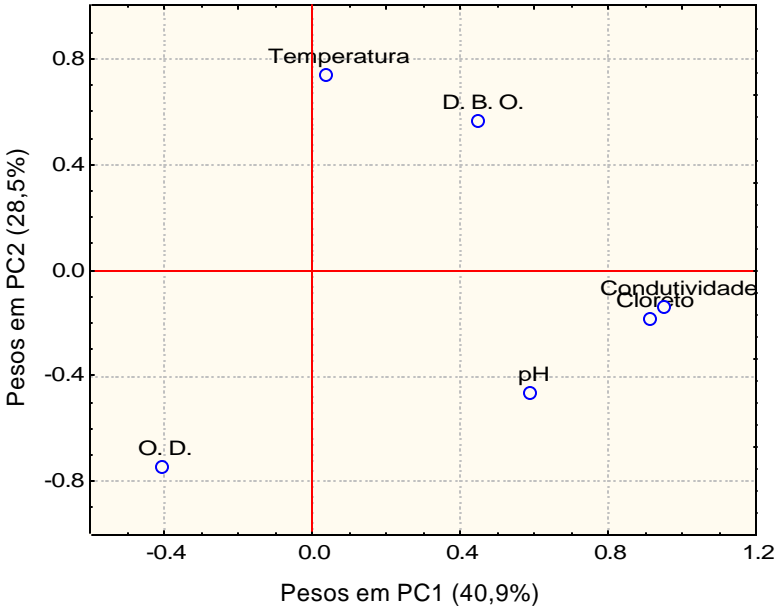


Figura 13 – Pesos nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no inverno.

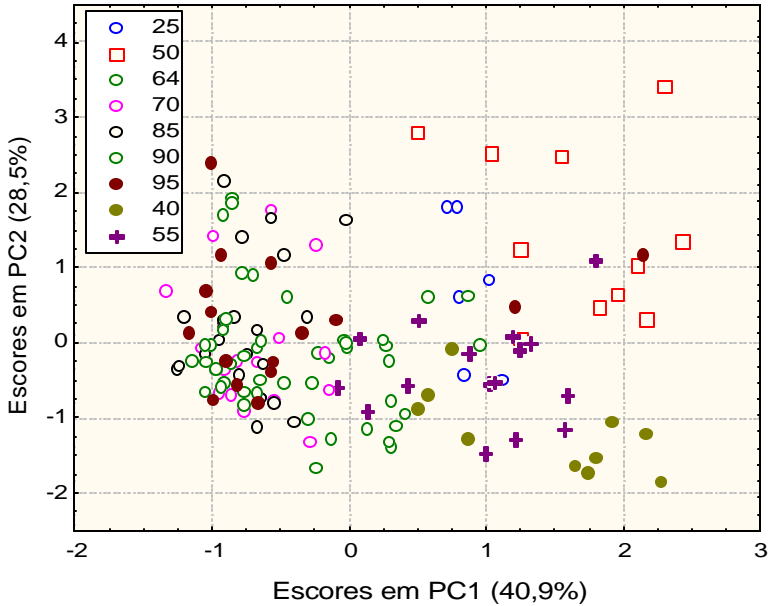


Figura 14 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no inverno. Estratificação por estação de coleta.

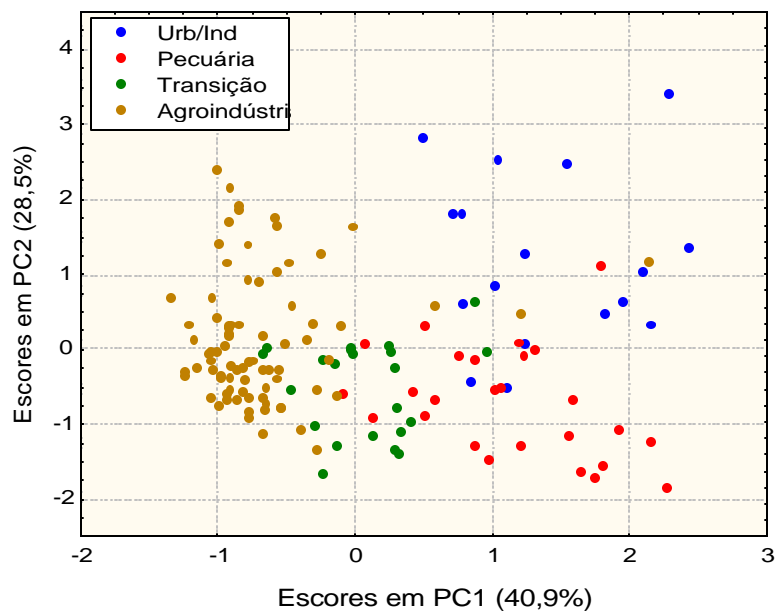


Figura 15 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no inverno. Estratificação por zona de atividade.

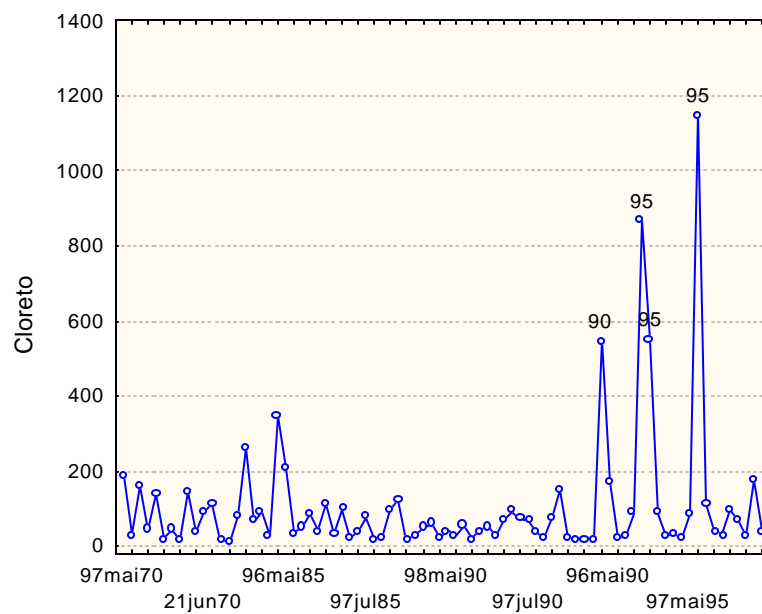


Figura 16 – Série temporal do Cloreto no inverno, para a zona agroindustrial.

O pH comporta-se de forma semelhante ao observado no verão, porém sem nenhum ponto anômalo. Para a Temperatura, o padrão de variação com a área é essencialmente o mesmo do verão. O mesmo se pode concluir para o Cloreto.

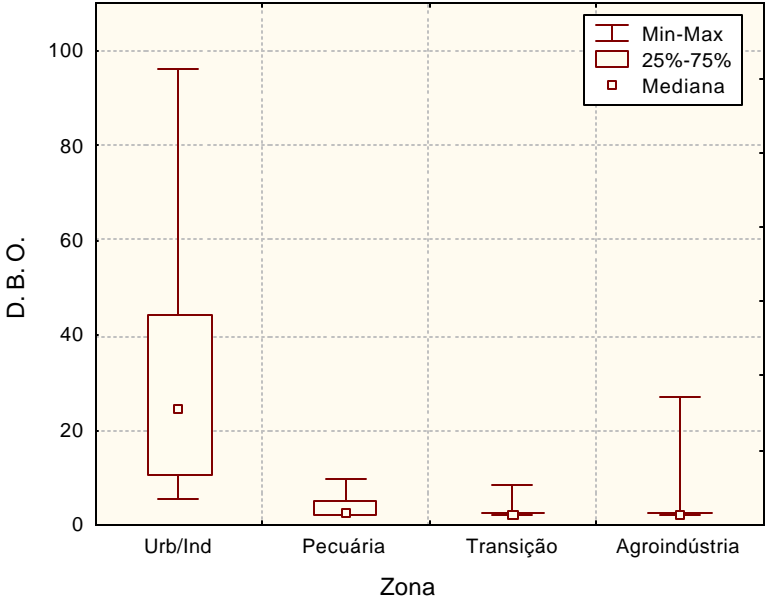


Figura 17 – Gráficos de caixa da D.B.O. no inverno, por zona de atividade.

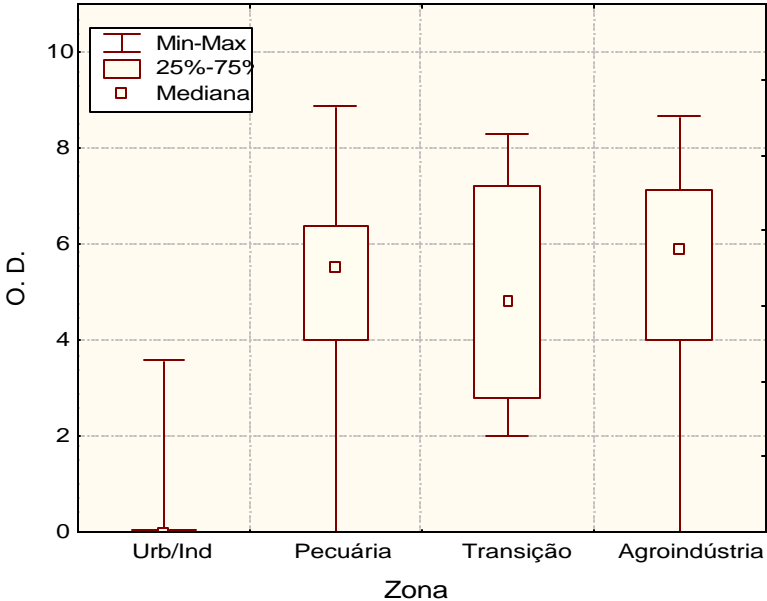


Figura 18 – Gráficos de caixa do O.D. no inverno, por zona de atividade.

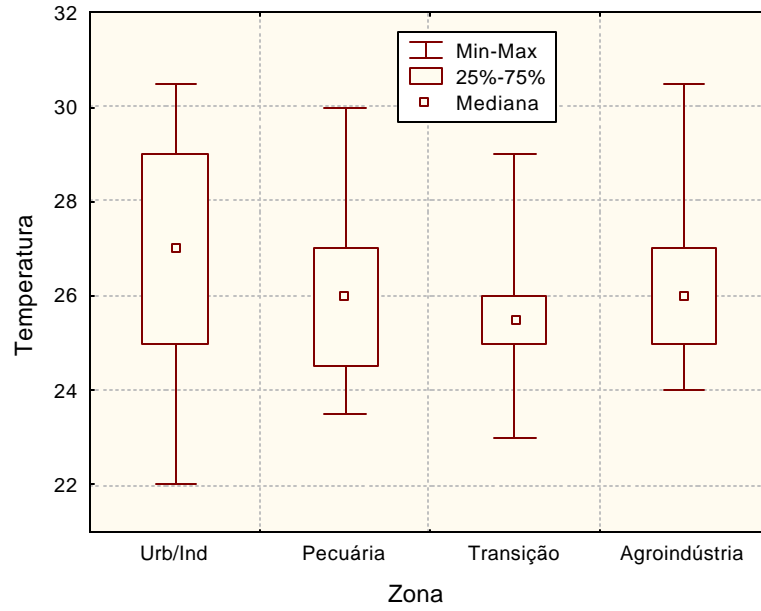


Figura 19 – Gráficos de caixa da Temperatura no inverno, por zona de atividade.

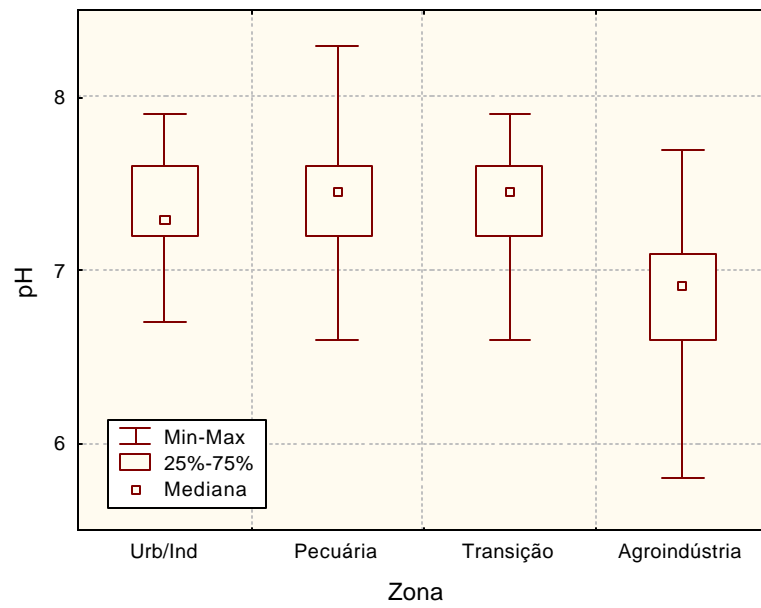


Figura 20 – Gráficos de caixa do pH no inverno, por zona de atividade.

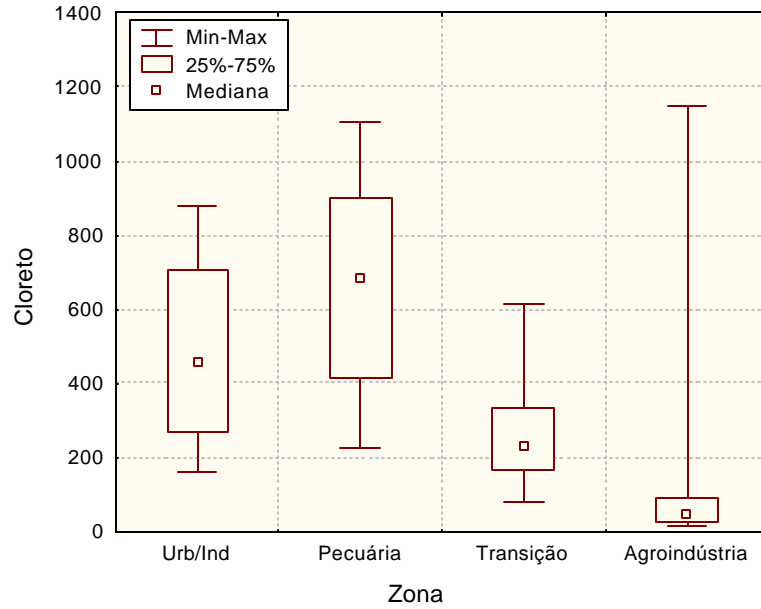


Figura 21 – Gráficos de caixa do Cloreto no inverno, por zona de atividade.

Análise dos dados do Reservatório Tapacurá

Os dados do Reservatório Tapacurá são bem mais simples do que os do Rio Ipojuca. Não há diferentes estações de coleta, e o número de casos contendo valores válidos para todas as sete variáveis selecionadas é apenas 35. A análise, conseqüentemente, também terá resultados menos complexos.

Os dados cobrem o intervalo de junho de 1999 a dezembro de 2002. Como se sabe, 1999 foi um período de grande estiagem, situação que só começou a ser revertida a partir de meados de 2000. A série temporal do volume percentual de água acumulada na barragem ilustra claramente esse fato (Figura 22).

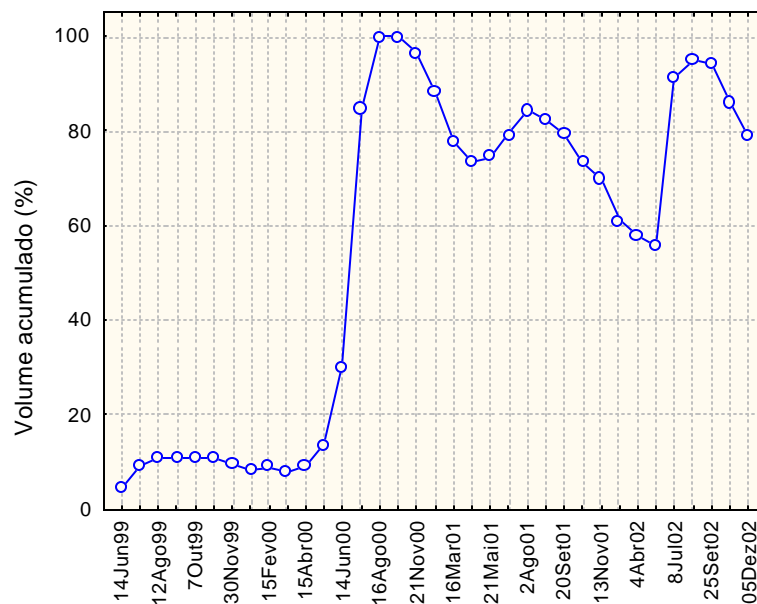


Figura 22 – Série temporal do percentual de água acumulada no Reservatório Tapacurá, de 1999 a 2002.

Na PCA, as duas primeiras componentes principais explicam cerca de 70% da variância total. A Figura 23 mostra o gráfico dos pesos nas duas primeiras PCs. PC1 é uma média ponderada de todas as variáveis (todos os pesos são positivos), mas com maior participação da Temperatura, dos Oxigênios e da Salinidade/Sólidos dissolvidos. A interpretação é bastante simples: Quanto maiores forem os valores dessas variáveis para uma dada amostra, mais positivo será o seu escore em PC1, e mais à direita do gráfico estará ela. A segunda componente é essencialmente um contraste entre dois pares de variáveis altamente correlacionadas entre si: os Oxigênios, com pesos positivos, e Salinidade/Sólidos dissolvidos, com pesos negativos. Conseqüentemente, uma amostra com baixa Salinidade e muito Oxigênio dissolvido tenderá a localizar-se na parte superior do gráfico, e vice-versa.

Os escores, quando estratificados de acordo com a estação (seca ou chuvosa), não mostram nenhum padrão evidente de associação que possa ser atribuído a essa dicotomia (Figura 24). Deve-se observar, porém, que a informação correspondente à estação chuvosa é extremamente limitada (não mais que 11 amostras, ao longo de um período de quatro anos). Isto, naturalmente, dificulta a identificação de qualquer possível padrão.

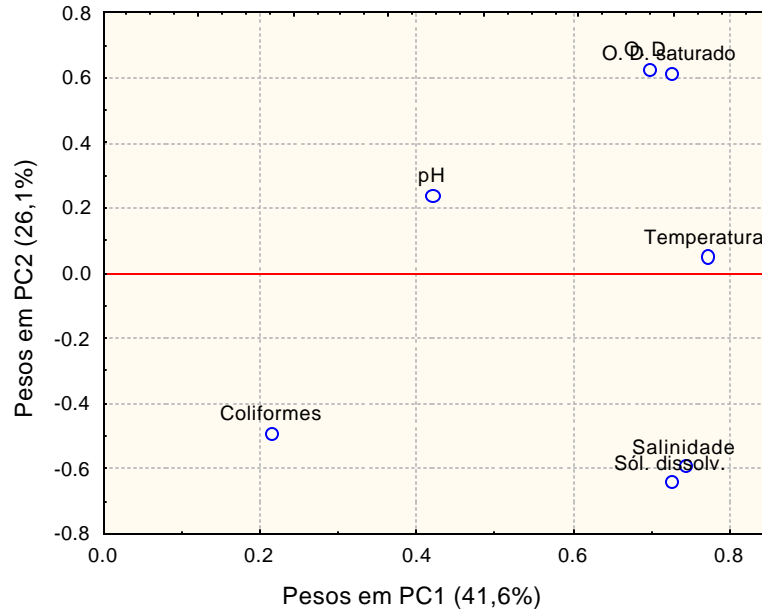


Figura 23 – Pesos nas duas primeiras componentes principais da análise dos dados do Reservatório Tapacurá.

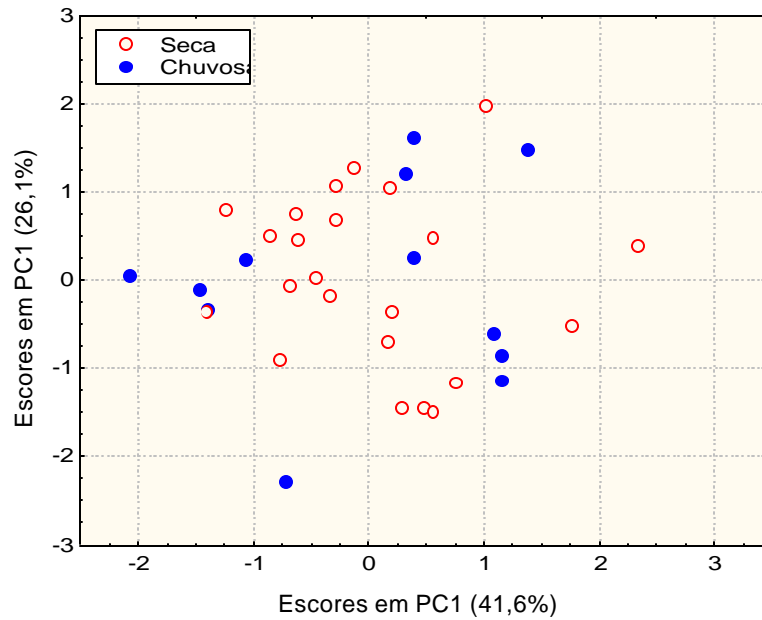


Figura 24 – Escores nas duas primeiras componentes principais da análise dos dados do Reservatório Tapacurá, estratificados por estação.

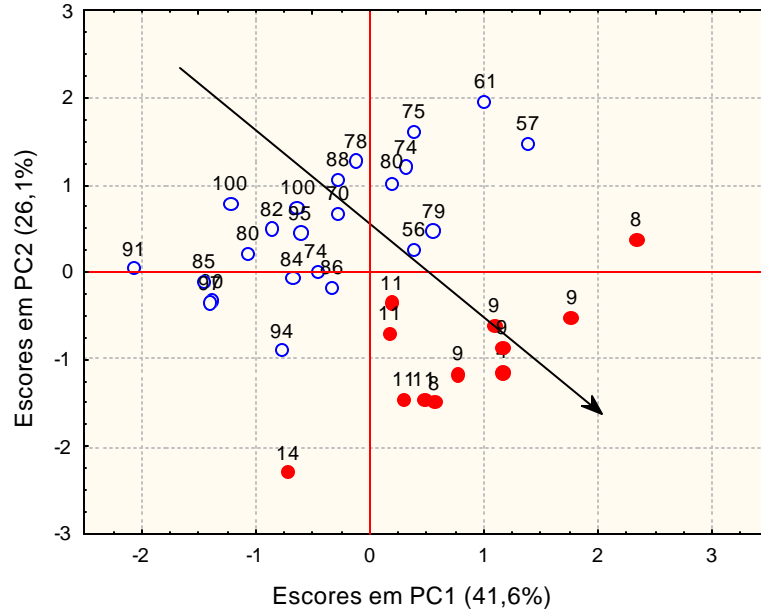


Figura 25 – Escores nas duas primeiras componentes principais da análise dos dados do Reservatório Tapacurá, com os volumes percentuais acumulados. Os pontos em vermelho têm volumes inferiores a 20%.

Quando a estratificação é feita por volume acumulado (Figura 25), podemos observar uma ordenação de acordo com o volume percentual de água acumulada na barragem. Percorrendo-se o gráfico na direção indicada pela seta, isto é, do canto superior esquerdo ao inferior direito as amostras vão correspondendo a níveis de água cada vez mais baixos. A direção da seta corresponde a um aumento do escore em PC1 e uma diminuição simultânea do escore em PC2. As amostras com nível mais baixo estão quase todas no quadrante inferior direito, que corresponde, de acordo com os pesos da Figura 23, a condições de alta Salinidade/Sólidos dissolvidos e baixos teores de Oxigênio. Ou seja: a maior parte da variação observada para os valores registrados na verdade pode ser explicada por uma única variável: o nível de água acumulada na barragem. A principal variável associada com esse fato parece ser a Salinidade, como mostra o diagrama matricial da Figura 26.

Todos os dados, tanto do Ipojuca quanto de Tapacurá, também foram submetidos a análises de agrupamentos hierárquicas, mas como essas análises não trouxeram nenhuma informação nova, deixamos de incluir seus resultados neste relatório.

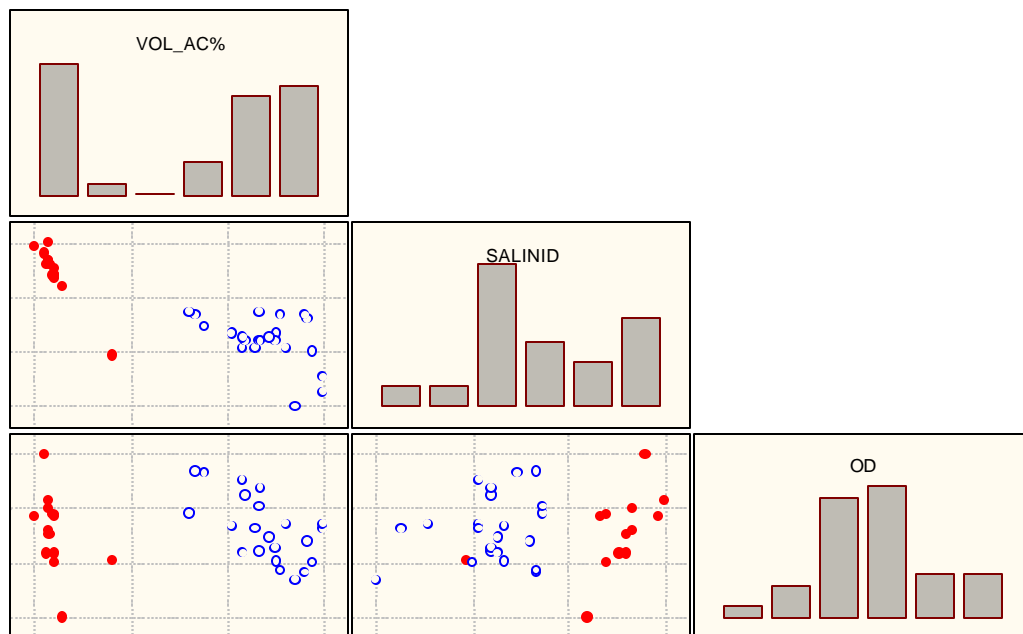


Figura 26 – Diagrama matricial do Volume percentual acumulado, da Salinidade e o do O. D. para o Reservatório Tapacurá. As cores dos pontos são as mesmas da figura 24.

Treinamento

Como previsto no plano de trabalho, todas as atividades contaram com a participação dos técnicos Helena Alves Moreira (CPRH) e Luciano Costa Almeida (SRH), que receberam treinamento nos seguintes procedimentos, no ambiente do programa *Statistica 5.5*:

- Análise e interpretação de componentes principais
- Análise e interpretação de agrupamentos hierárquicos.

Recife, 03 de abril de 2003.

Maria Fernanda Pimentel

Programa Nacional do Meio Ambiente II – PNMA II

Componente: Desenvolvimento Institucional

Subcomponente: Monitoramento da Qualidade da Água

Projeto: Monitoramento da Qualidade da Água como Instrumento de Controle Ambiental e Gestão de Recursos Hídricos no Estado de Pernambuco

**Análise Estatística de Dados
do Monitoramento da Qualidade das Águas
do Rio Ipojuca e do Reservatório Tapacurá**

Produto 5 – Relatório Final

Maria Fernanda Pimentel

Recife, junho de 2003

Introdução

O presente relatório apresenta as conclusões resultantes de uma oficina realizada em 27 de maio de 2003, da qual participaram técnicos da SECTMA e consultores envolvidos na definição de zonas homogêneas de monitoramento na bacia do Rio Ipojuca.

Os participantes da oficina foram divididos em três equipes, a cada uma das quais foi solicitada uma proposta de definição de zonas homogêneas na Bacia do Ipojuca, levando em conta os relatórios apresentados pelos diversos consultores. Ao término da oficina, chegou-se, por decisão consensual, a uma proposta contemplando seis zonas distintas, a saber:

1. Zona de **Interesse Ambiental**
2. Zona **Agroindustrial**
3. Zona de **Pecuária Leiteira**
4. Zona de **Pecuária de Corte**
5. Zona **Urbana /Industrial**
6. Zona da **Policultura**

Zonas propostas - Análise de componentes principais (PCA)

Das seis zonas homogêneas definidas na oficina, as zonas 1 e 3 (Interesse Ambiental e Pecuária Leiteira) não têm correspondentes na análise apresentada no Relatório 3, já que não foram representadas por nenhuma estação de coleta. Para as demais, a oficina concluiu apenas por uma modificação parcial da nomenclatura, de forma a torná-la mais específica. Conseqüentemente, a correspondência entre as zonas propostas e as atuais estações de coleta passou a ser a seguinte:

- Estações **25** e **50**: Zona Urbana/Industrial;
- Estações **40** e **55**: Zona de Pecuária de Corte;
- Estação **64**: Zona da Policultura;
- Estações **70**, **85**, **90** e **95**: Zona Agroindustrial.

Nomenclatura à parte, esta correspondência é a mesma do Relatório 3. Não há, portanto, em decorrência das atividades desenvolvidas na oficina, nada a acrescentar às conclusões das análises de componentes principais já relatadas. Apenas para efeito de atualização da nomenclatura, os gráficos dos escores das amostras passam a ser os que aparecem nas Figuras 1 e 2. Os gráficos dos pesos (Figuras 3 e 4) permanecem exatamente os mesmos de antes.

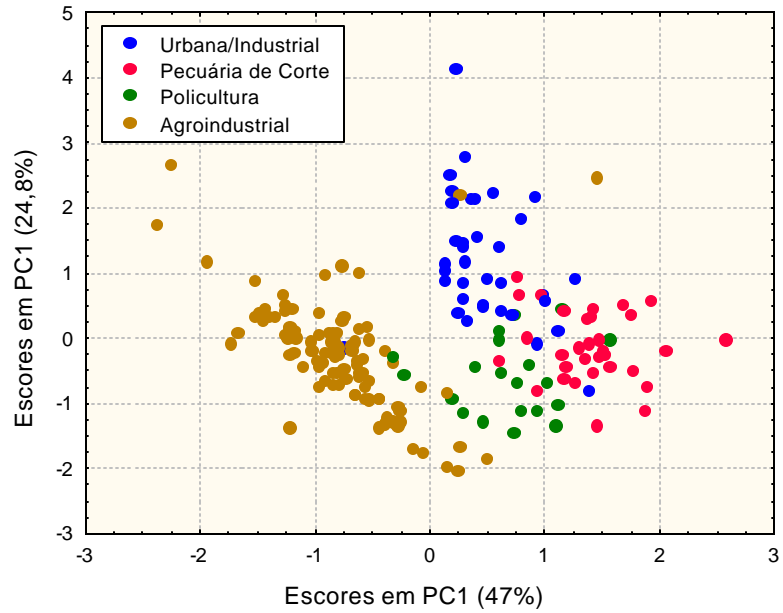


Figura 1 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no verão (186 amostras). Estratificação por zona de atividade principal.

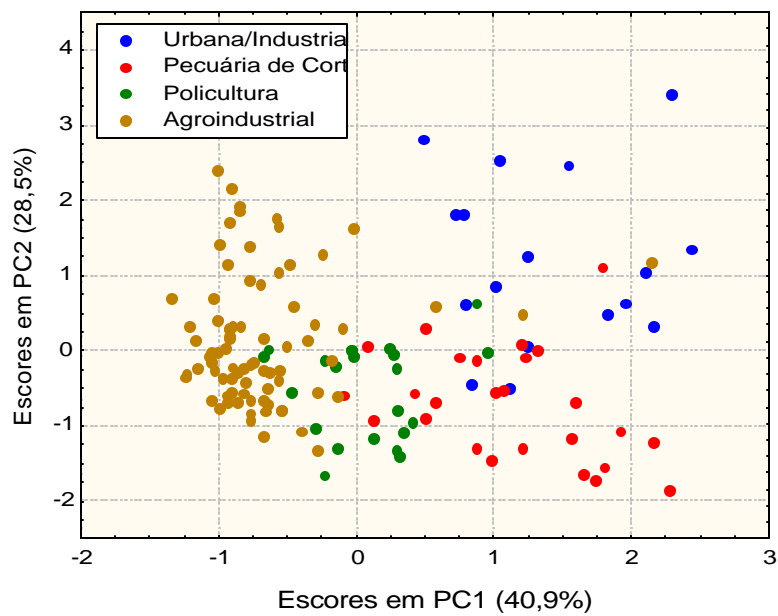


Figura 2 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no inverno (145 amostras). Estratificação por zona de atividade.

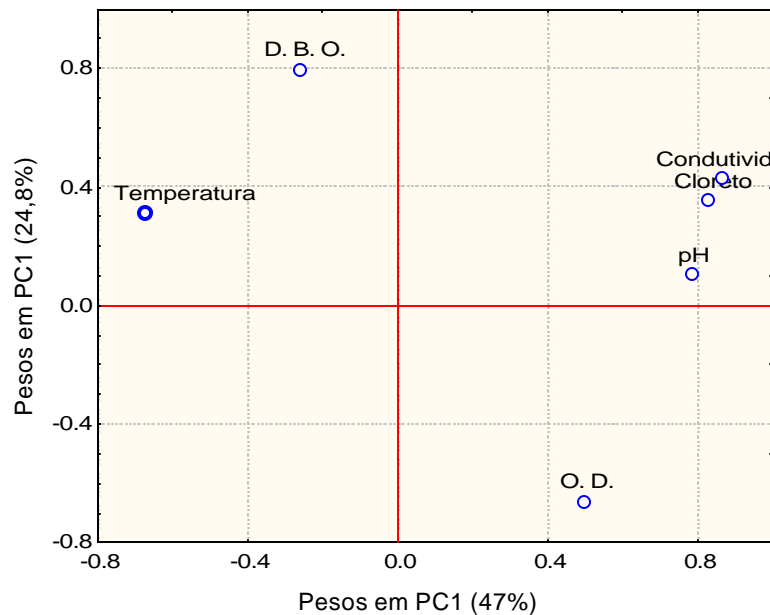


Figura 3 – Pesos nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca coletados no verão. Os escores correspondentes são mostrados na Figura 1.

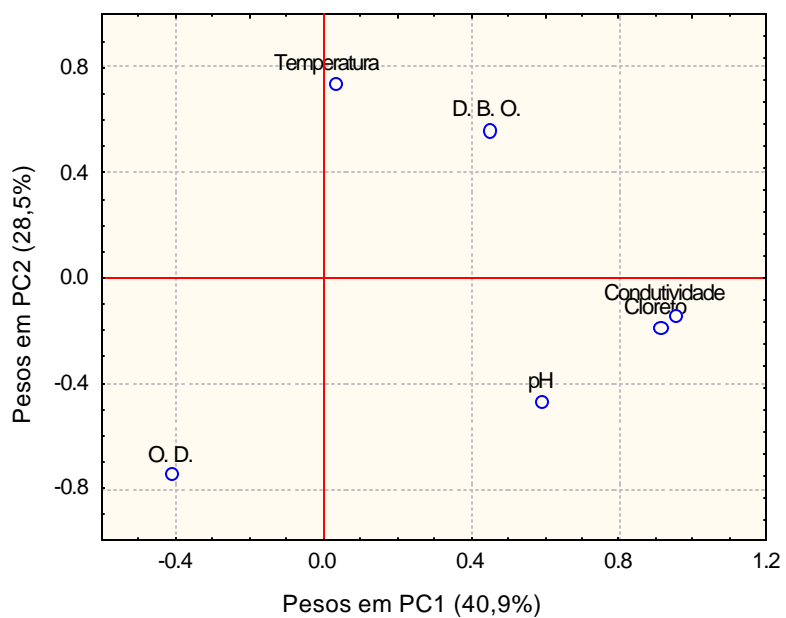


Figura 4 – Pesos nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca colhidos no inverno. Os escores correspondentes são mostrados na Figura 2.

Uma variável muito importante para o monitoramento da qualidade da água é o índice de coliformes fecais. É interessante verificar se a inclusão desse parâmetro na análise agrega algum dado que permita enriquecer as conclusões do trabalho.

As Figuras 5 - 11 mostram os resultados de uma análise de componentes principais em que os coliformes foram incluídos. Como alguns valores de coliformes estão ausentes, a primeira consequência é a diminuição no número de amostras válidas, que passa a ser 155 (102 no verão e 53 no inverno).

Os resultados da PCA com todo o conjunto de dados estão nas figuras 5 - 7. Os pesos nas duas primeiras componentes, que descrevem 71,8 % da variância total, têm a mesma interpretação das análises realizadas até agora, com uma informação adicional: observa-se que os pesos dos coliformes e da temperatura nas duas primeiras componentes são semelhantes (Figura 5). Isto nos permite concluir que maiores índices de coliformes estão associados a temperaturas mais altas. O gráfico dos escores apresenta, ao longo da primeira componente, uma separação parcial entre o verão e o inverno (Figura 6). O conjunto de dados correspondentes ao inverno está deslocado para a direita, o que significa que ele está associado, em média, a índices de coliformes menores e a temperaturas mais baixas. Esta observação confirma a relação observada anteriormente na análise bivariada, e registrada no Produto 3 (página 32).

Quando os escores são estratificados de acordo com a zona de atividade principal (Figura 7), pode-se observar uma separação bem definida entre as quatro zonas representadas na análise. Os padrões de associação entre as amostras permanecem essencialmente os mesmos que já foram destacados na Figura 1, que não inclui coliformes e só se baseia em amostras coletadas durante o verão. Podemos concluir, portanto, que a variabilidade do conjunto de todas as amostras é dominada pela variabilidade dos dados coletados durante o verão. Resultado idêntico ao obtido na análise sem os coliformes (página 50).

Os resultados das análises separadas de acordo com a estação climática são apresentados nas Figuras 8-11. Para avaliar a influência dos coliformes sobre os pesos nas duas primeiras componentes, devemos comparar a Figura 8 com a Figura 3 (para o verão), e a Figura 10 com a Figura 4 (para o inverno).

No verão, podemos observar que o padrão permanece o mesmo ao longo de PC1, com ou sem coliformes. A D. B. O. e a temperatura (e os coliformes) têm pesos negativos, enquanto as demais variáveis têm pesos positivos. Ao longo de PC2, a introdução dos coliformes produz uma inversão de sinais algébricos. Sem os coliformes, o único peso negativo na segunda componente principal é o do O. D. Com os coliformes, o O. D. passa a ter o único peso *positivo*. Tecnicamente, este fenômeno é chamado de mudança de fase da componente. É uma consequência do algoritmo de cálculo, e em nada afeta as conclusões da

análise. Os padrões de associação permanecem os mesmos, mas correspondem a quadrantes diferentes. A análise dos pesos para o inverno (Figuras 10 e 4) leva a conclusões análogas.

Se a entrada dos coliformes não trouxe alteração significativa para os pesos, o mesmo forçosamente terá de verificar-se para os escores. É o que podemos constatar comparando as Figuras 9 e 1 (verão) e 11 e 2 (inverno). O que se observa é apenas a esperada inversão ao longo de PC2, devida à mudança de fase.

Em toda esta análise, porém, cabe uma importante ressalva, que se aplica com mais intensidade aos dados coletados no inverno. Com a entrada dos coliformes, o número de amostras tornou-se ainda mais limitado. Conseqüentemente, a extrapolação dos padrões observados, que já é, por natureza, um procedimento arriscado, ficou mais difícil de justificar. Na Figura 11, por exemplo, a zona urbana/industrial é descrita por apenas três amostras. Trata-se de um número que não só é totalmente inadequado para sustentar qualquer argumentação estatística como neste caso tem a agravante de que essas três amostras deveriam representar uma série temporal de seis anos e dois pontos de coleta.

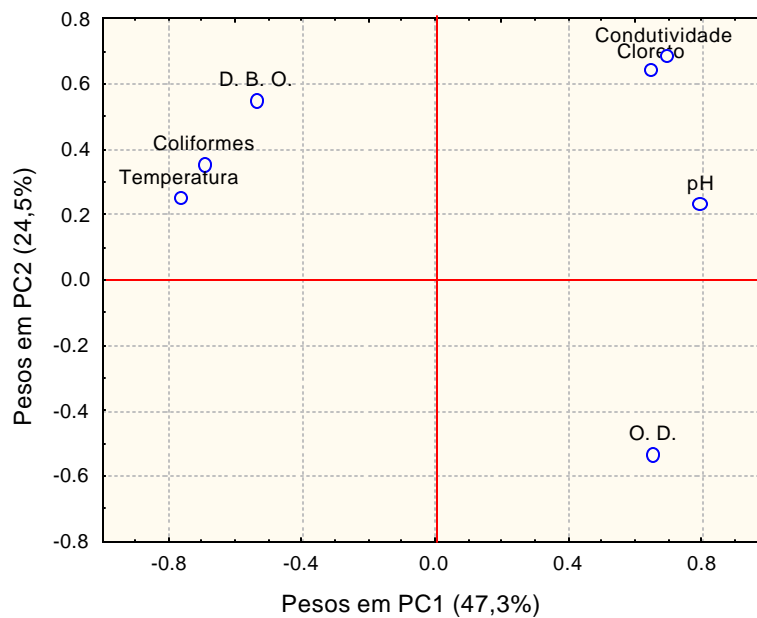


Figura 5 – Pesos nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca, incluindo coliformes fecais.

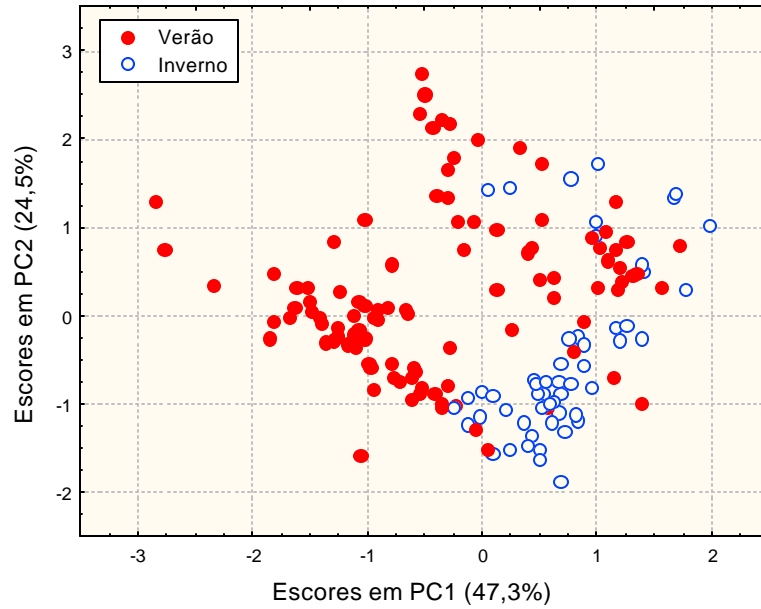


Figura 6 – Escores nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca, incluindo coliformes fecais. Os pesos correspondentes são mostrados na Figura 5.

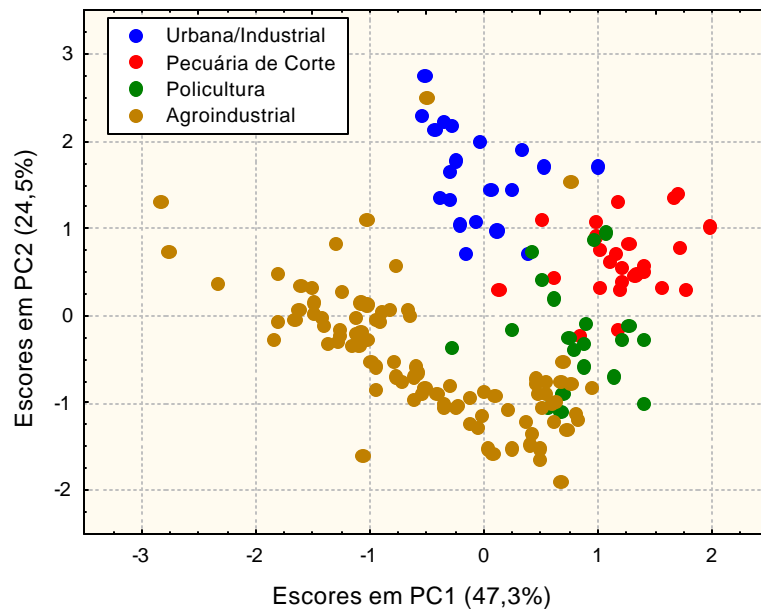


Figura 7 – Escores nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca, incluindo coliformes fecais. Estratificação por zona de atividade. Os pesos correspondentes são mostrados na Figura 5.

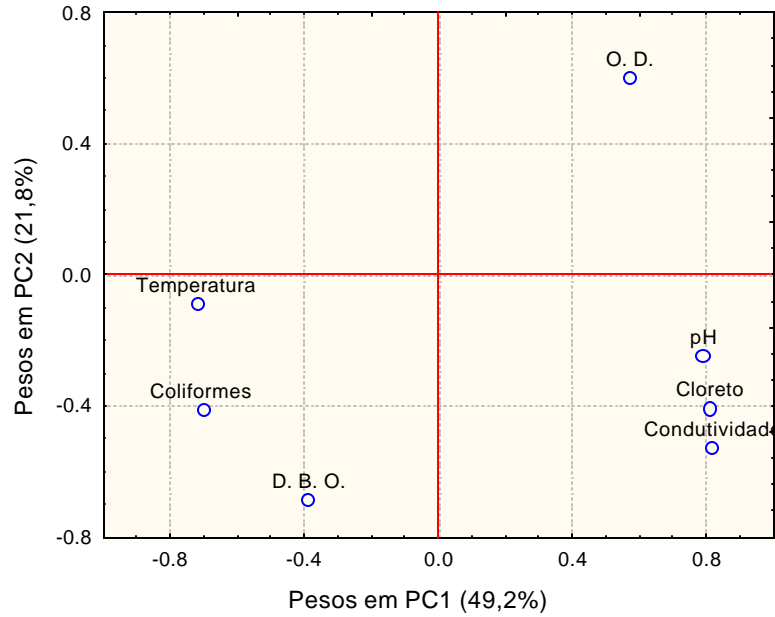


Figura 8 – Pesos nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca, incluindo coliformes fecais, para as amostras coletadas no verão.

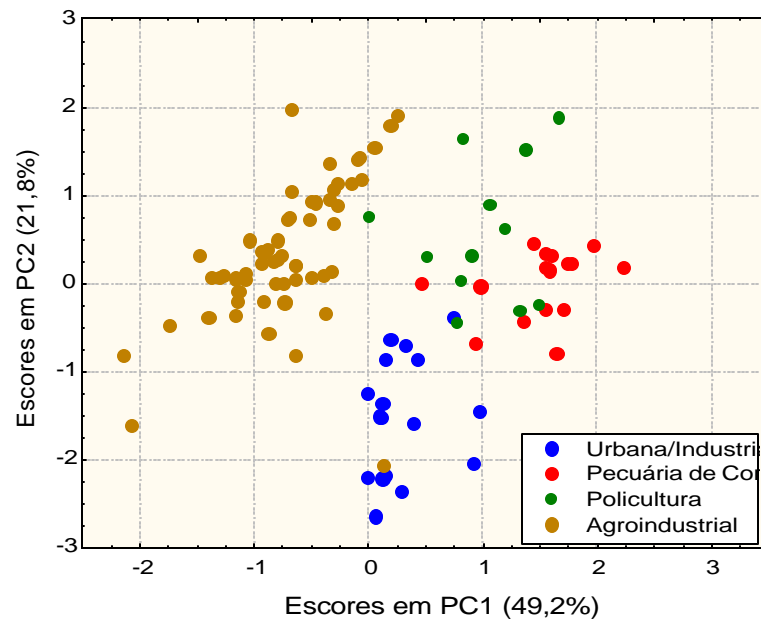


Figura 9 – Escores nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca, incluindo coliformes fecais, para as amostras coletadas no verão (102 amostras). Estratificação por zona de atividade. Os pesos correspondentes são mostrados na Figura 8.

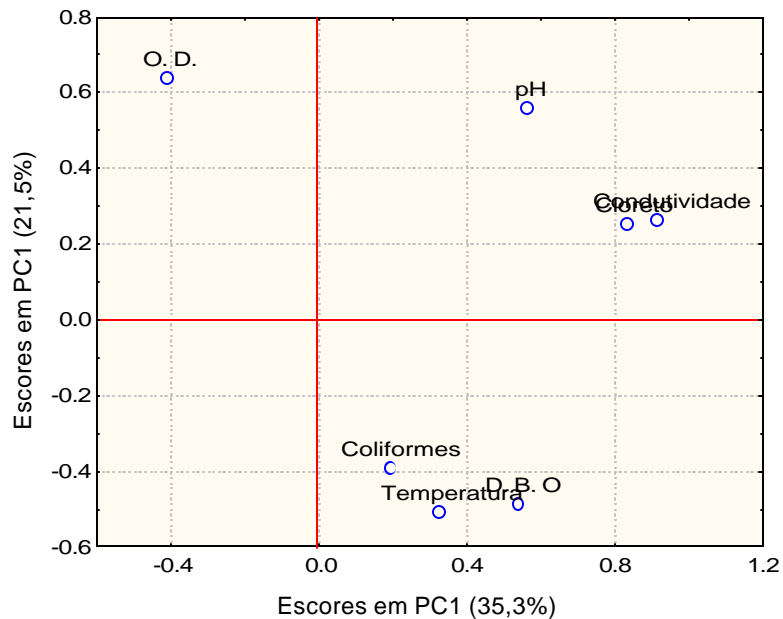


Figura 10 – Pesos nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca, incluindo coliformes fecais, para as amostras coletadas no inverno.

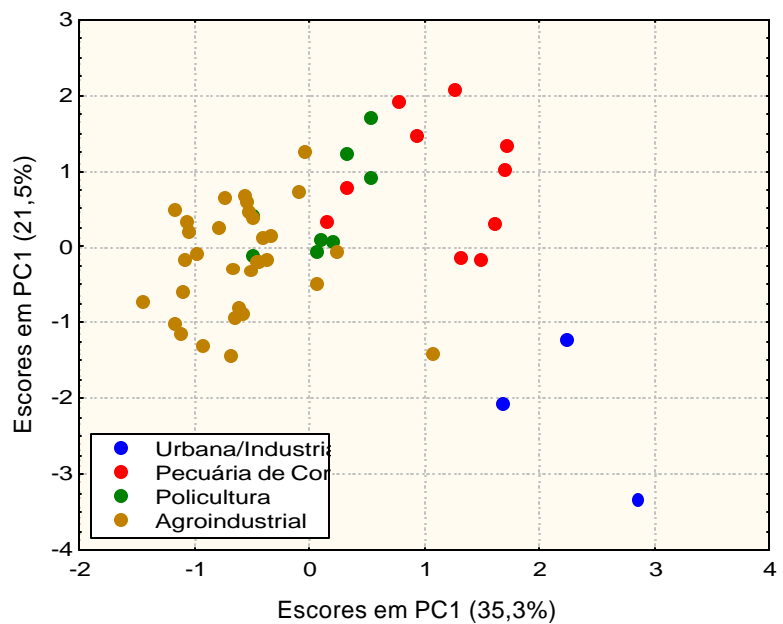


Figura 11 – Escores nas duas primeiras componentes principais da análise dos dados do Rio Ipojuca, incluindo coliformes fecais, para as amostras coletadas no inverno (53 amostras). Estratificação por zona de atividade. Os pesos correspondentes são mostrados na Figura 8.

Redução da frequência de amostragem

Uma interessante questão levantada durante a oficina foi a possibilidade de se reduzir a frequência de amostragem das estações de coleta. Para avaliar que efeito uma possível redução na amostragem teria sobre os padrões de agrupamento, realizamos novas análises de componentes principais com subconjuntos de amostras, simulando análises realizadas bimestralmente, e não mais mensalmente. Para tanto, dividimos os dados disponíveis da seguinte forma:

Verão:

- *Amostragem 1*: meses de agosto, outubro, dezembro e fevereiro (138 amostras)
- *Amostragem 2*: meses de setembro, novembro e janeiro (48 amostras)

Inverno:

- *Amostragem 1*: meses de abril, junho e agosto (105 amostras)
- *Amostragem 2*: meses de março, maio e julho (40 amostras)

Os resultados da PCA das amostras do verão são apresentados nas Figuras 12 e 13 (amostragem 1) e 14 e 15 (amostragem 2). Pode-se observar que tanto os gráficos dos pesos quanto os gráficos dos escores, nas duas amostragens, apresentam padrões muito parecidos com os observados para o conjunto completo (Figuras 1 e 3), em que pese a relativa escassez de pontos na análise da amostragem 2.

A semelhança entre os resultados das análises dos dois subconjuntos e o resultado da análise completa é uma indicação de que a redução da frequência de amostragem pode ser considerada em futuros monitoramentos. Deve-se ressaltar, contudo, que esta conclusão é de natureza estatística, e se baseia apenas nos parâmetros usados para a PCA (O. D., D. B. O., Temperatura, pH, Cloreto e Condutividade). A implantação de um esquema de amostragem reduzida certamente terá de levar em conta, além de questões de logística, outros parâmetros de interesse ambiental que não puderam ser incluídos na análise apresentada aqui.

Para o inverno, os resultados das análises de componentes principais nos dois subconjuntos são mostrados nas Figuras 16-19, que devem ser comparadas com as Figuras 2 e 4. As conclusões que podemos tirar desses gráficos, sobre a possibilidade de uma amostragem menos freqüente, são análogas às da análise dos dados do verão. A elas se aplicam também as mesmas ressalvas, e com maior intensidade.

Recife, 03 de junho de 2003.

Maria Fernanda Pimentel

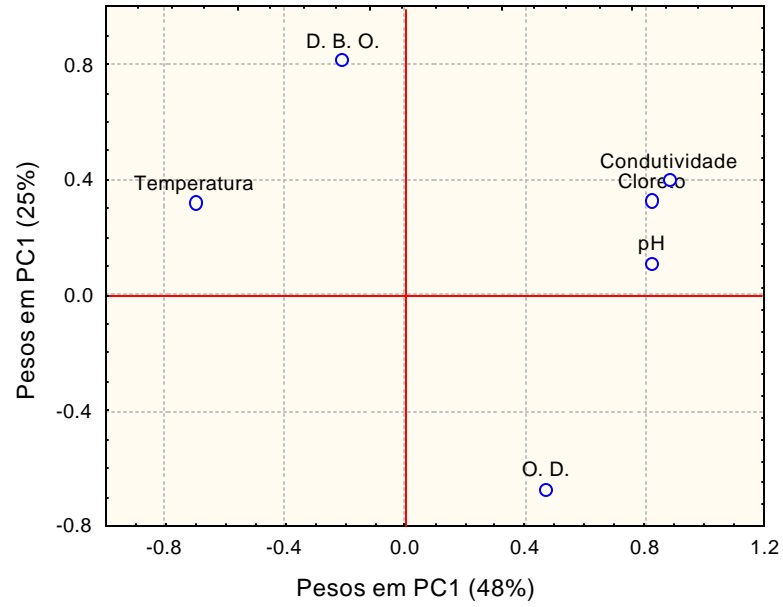


Figura 12 – Pesos nas duas primeiras componentes principais da análise dos dados do verão (amostragem 1, 138 amostras). Os escores correspondentes são mostrados na Figura 6.

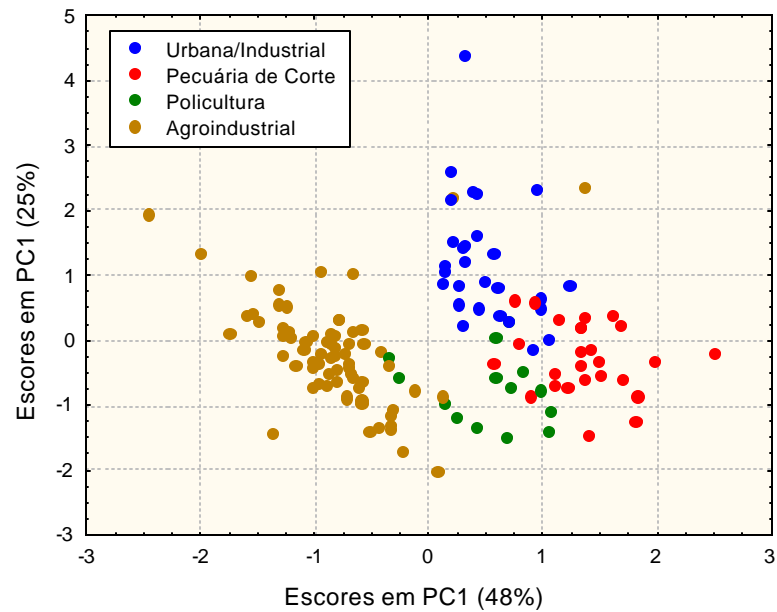


Figura 13 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do verão (amostragem 1, 138 amostras). Estratificação por zona de atividade.

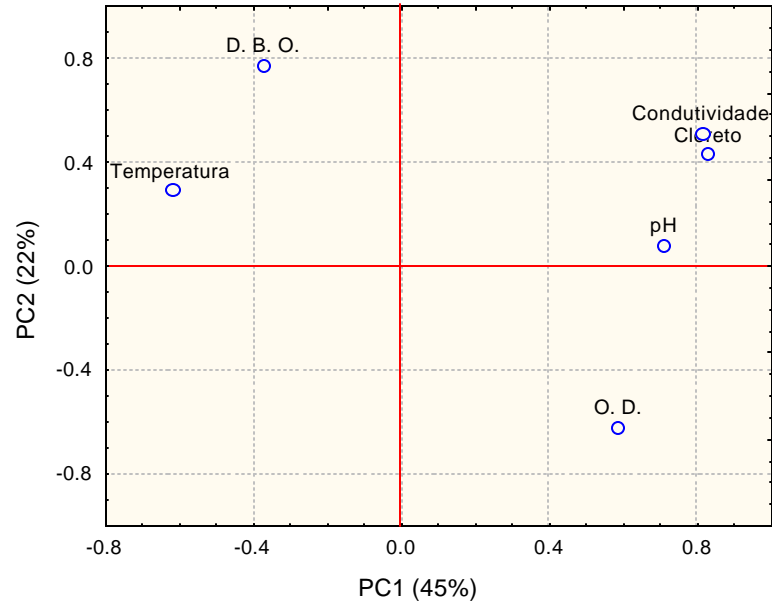


Figura 14 – Pesos nas duas primeiras componentes principais da análise dos dados do verão (amostragem 2, 48 amostras). Os escores correspondentes são mostrados na Figura 8.

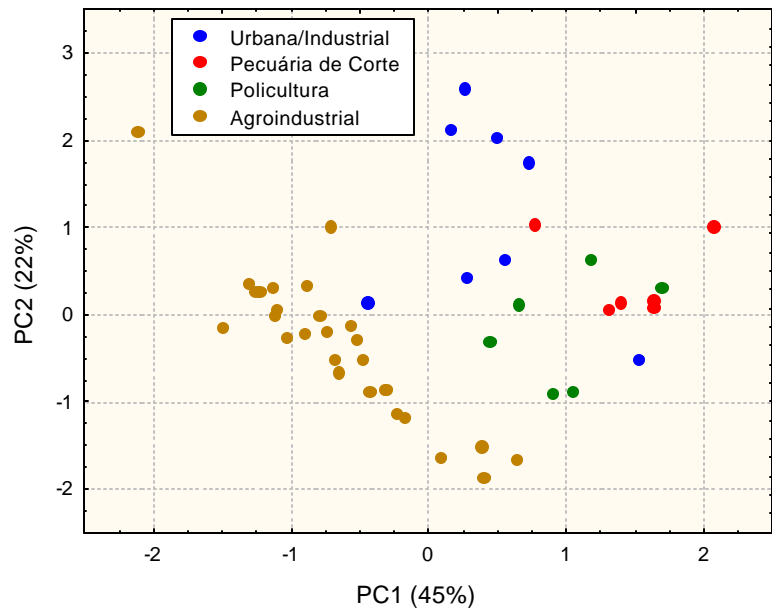


Figura 15 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do verão (amostragem 2, 48 amostras). Estratificação por zona de atividade.

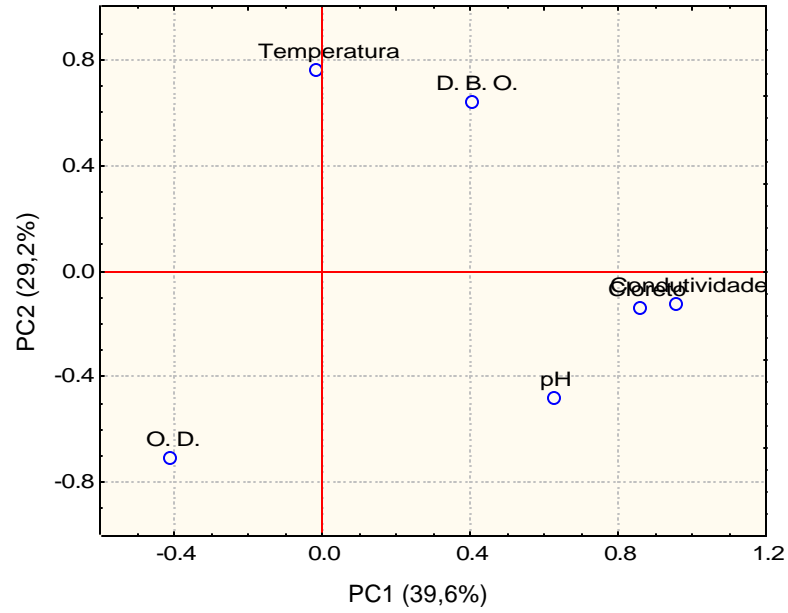


Figura 16 – Pesos nas duas primeiras componentes principais da análise dos dados do inverno (amostragem 1, 105 amostras). Os escores correspondentes são mostrados na Figura 10.

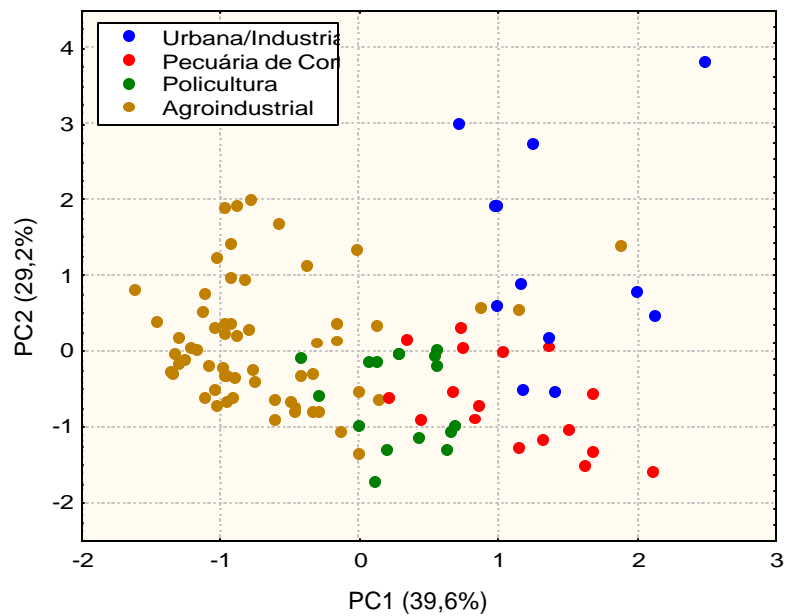


Figura 17 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do inverno (amostragem 1, 105 amostras). Estratificação por zona de atividade.

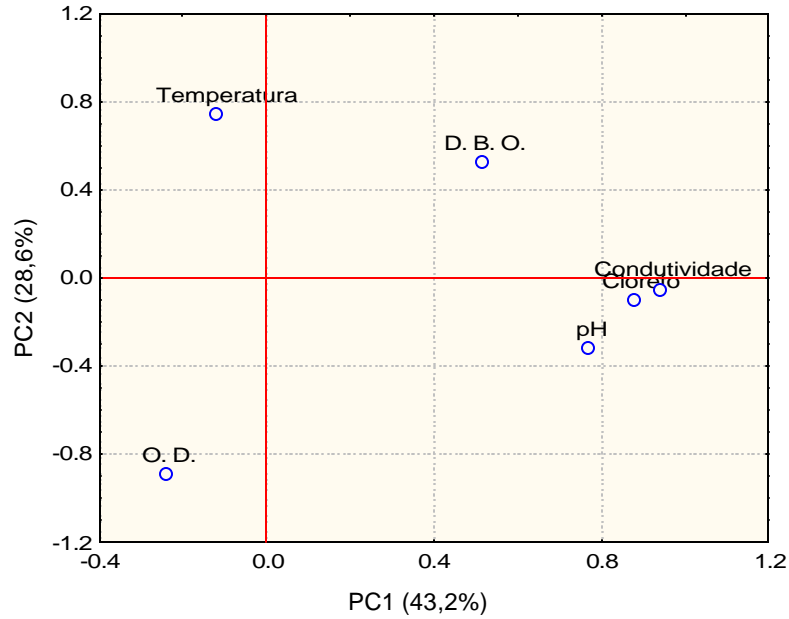


Figura 18 – Pesos nas duas primeiras componentes principais da análise dos dados do inverno (amostragem 2, 40 amostras). Os escores correspondentes são mostrados na Figura 10.

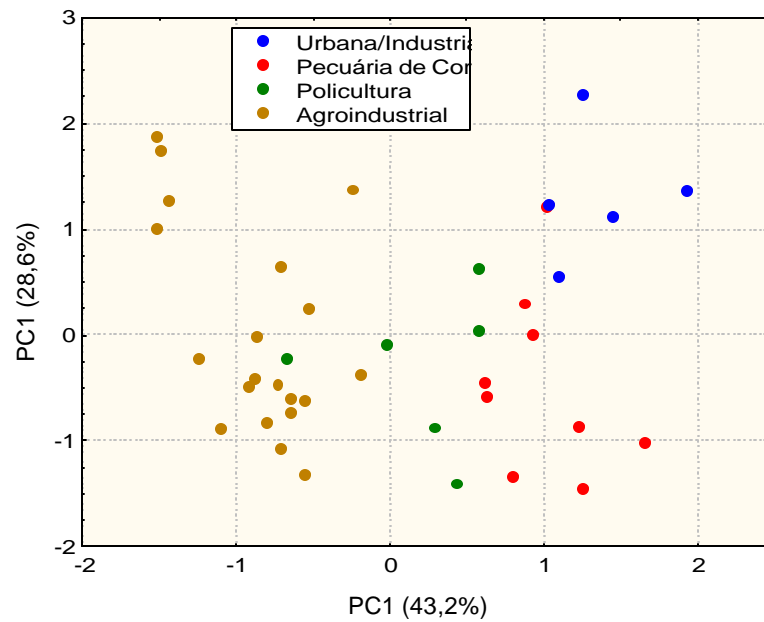


Figura 19 – Escores das amostras nas duas primeiras componentes principais da análise dos dados do inverno (amostragem 2, 40 amostras). Estratificação por zona de atividade.