

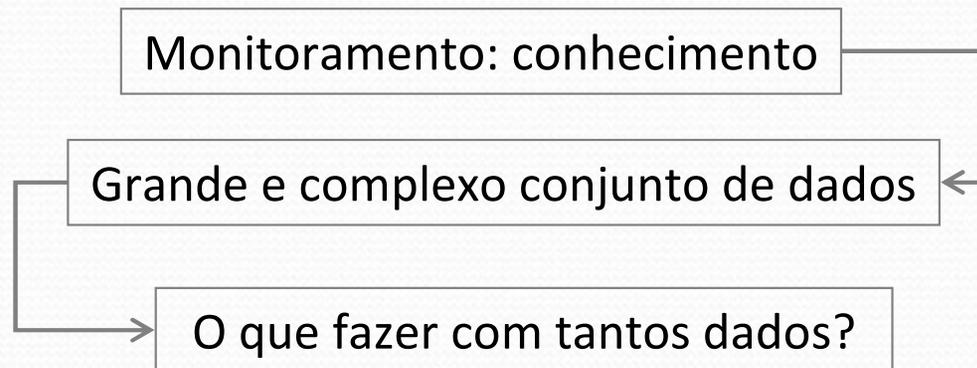
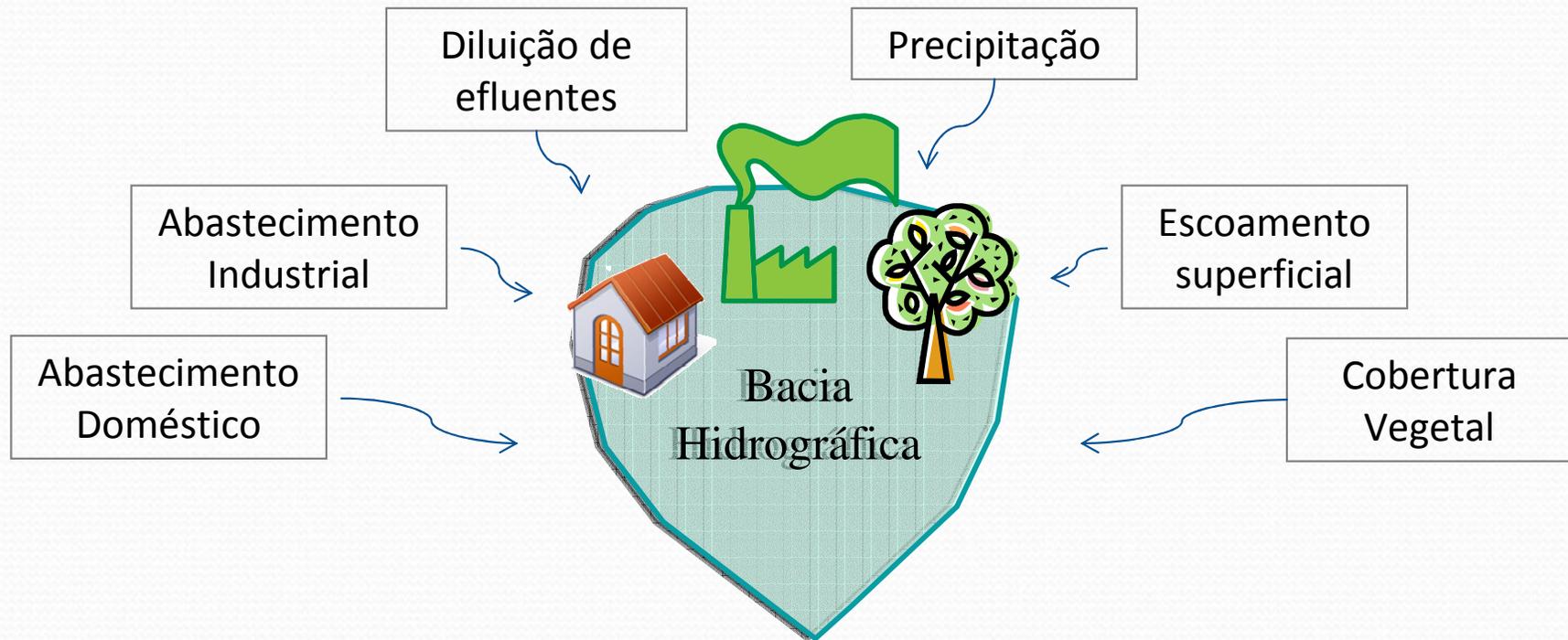
Avaliação da qualidade da água *uma abordagem estatística*

Marianne Schaefer França (M.Sc.)
Cristovão Vicente Scapulatempo Fernandes (Orientador)
Eloy Kaviski (Co-orientador)



Aspectos Conceituais

1 Introdução



Análise Multivariada!

2 Justificativa

- Resultados de KNAPIK (2006) acerca do cálculo do coeficiente de correlação para dados monitorados na mesma bacia.
- Apresentar um procedimento para aplicação deste tipo de análise em dados de monitoramento.
- Transformar dados em informação, os investimentos devem apresentar um retorno.
- Primeiro estudo realizado na RMC quanto à aplicação de técnicas multivariadas na avaliação da qualidade da água da bacia do Alto Iguaçu.

3 Objetivos

1. Realizar a AM dos dados de QA da Bacia do Alto Iguaçu considerando as seguintes técnicas estatísticas: Análise de Componentes Principais, Análise Fatorial e Análise de Agrupamentos, utilizando-se das rotinas estatísticas dos *softwares* MATLAB e STATISTICA.
2. Indicar um conjunto representativo de parâmetros de QA que pudessem eventualmente mostrar a melhor estratégia de monitoramento, visando sua otimização.
3. Identificar os Pontos de Monitoramento mais relevantes para avaliação da QA.
4. Agrupar as coletas de amostras de água do rio, visando encontrar quais refletem a melhor e a pior qualidade do corpo hídrico.

4 Análise Multivariada

- Avaliar diversas variáveis simultaneamente;
- Compreender grandes e complexos conjuntos de dados;
- Simplificar ou facilitar a interpretação do fenômeno em estudo;
- Investigar a relação de dependência entre as variáveis;
- Agrupar variáveis similares;
- Otimizar a rede de amostragem;
- Programar melhor a frequência de amostragem e o número de parâmetros monitorados;
- Redução de custos e tempo.

4.1 Pré-requisitos para aplicação da AM

Quanto à matriz de dados

- nº de observações (linhas) > nº de variáveis (colunas)

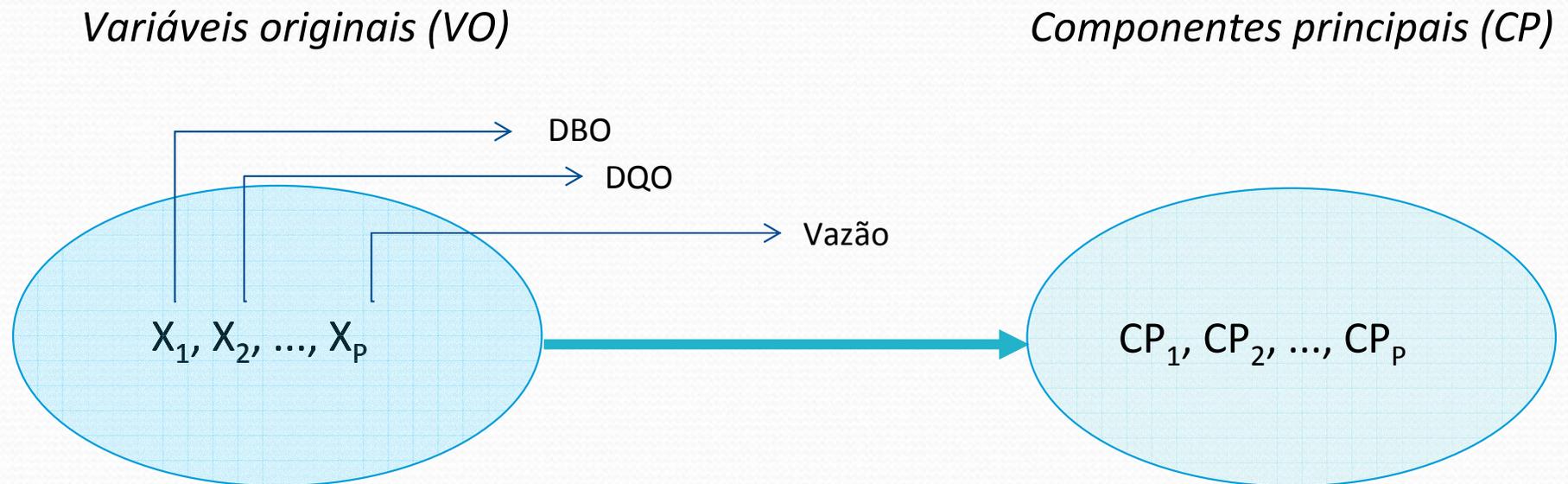
Quanto à aplicação da Análise Fatorial

- Teste de Esfericidade de Bartlett e Medida de Adequacidade de KMO

Quanto à distribuição normal

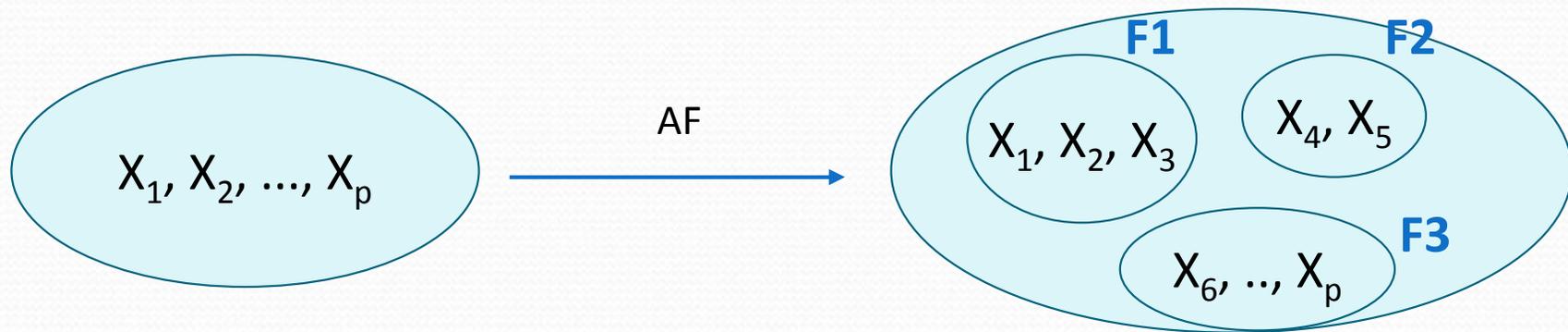
- Método da máxima verossimilhança para estimação dos fatores (AF)
- Teste de Esfericidade de Bartlett (MINGOTI, 2005)

4.2 Análise de Componentes Principais



- As componentes principais são combinações lineares das variáveis originais;
- “p” variáveis \rightarrow “p” CPs, no entanto, a maior parte da variabilidade total pode ser explicada por “k” CPs (onde $k < p$);

4.3 Análise Fatorial



- A AF simplifica ainda mais a estrutura vinda da ACP, diminuindo a contribuição das variáveis com menor significância e aumentando a contribuição das que possuem maior significância através da rotação dos eixos definidos pela ACP.
- Dentro de um fator as variáveis são altamente correlacionadas entre si, de um fator para outro as correlações são baixas.



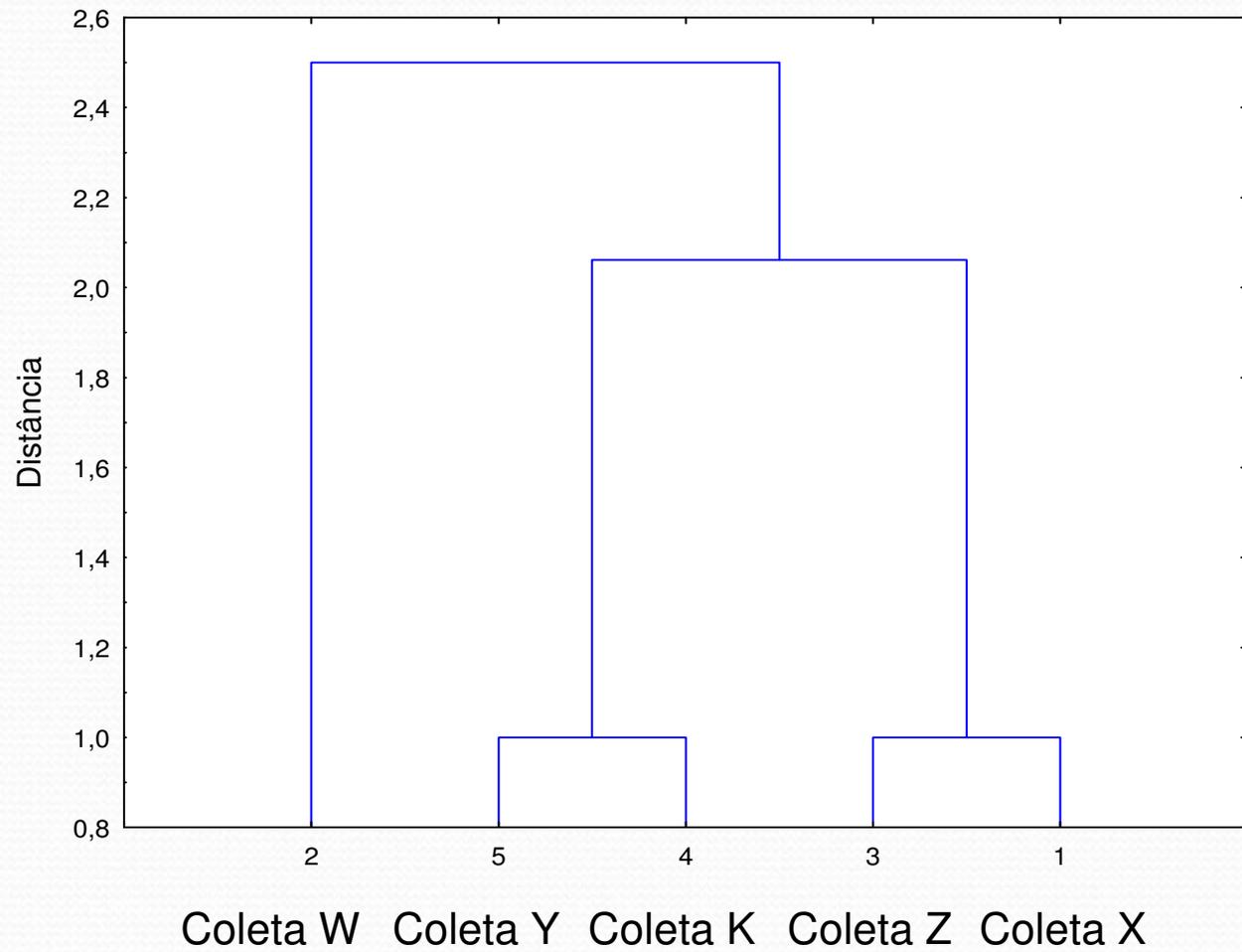
4.4 Análise de Agrupamentos

Dividir os elementos do conjunto amostral em grupos de forma que os elementos de um mesmo grupo sejam similares entre si e os elementos em grupos diferentes sejam heterogêneos.

4.4.1 Métodos de agrupamentos hierárquicos

As técnicas de agrupamento são realizadas por série de junções sucessivas. Os métodos aglomerativos hierárquicos começam com objetos individuais e no início há tantos grupos quanto objetos. Os grupos “fundem-se” de acordo com suas similaridades, até que se tenha um único grupo.

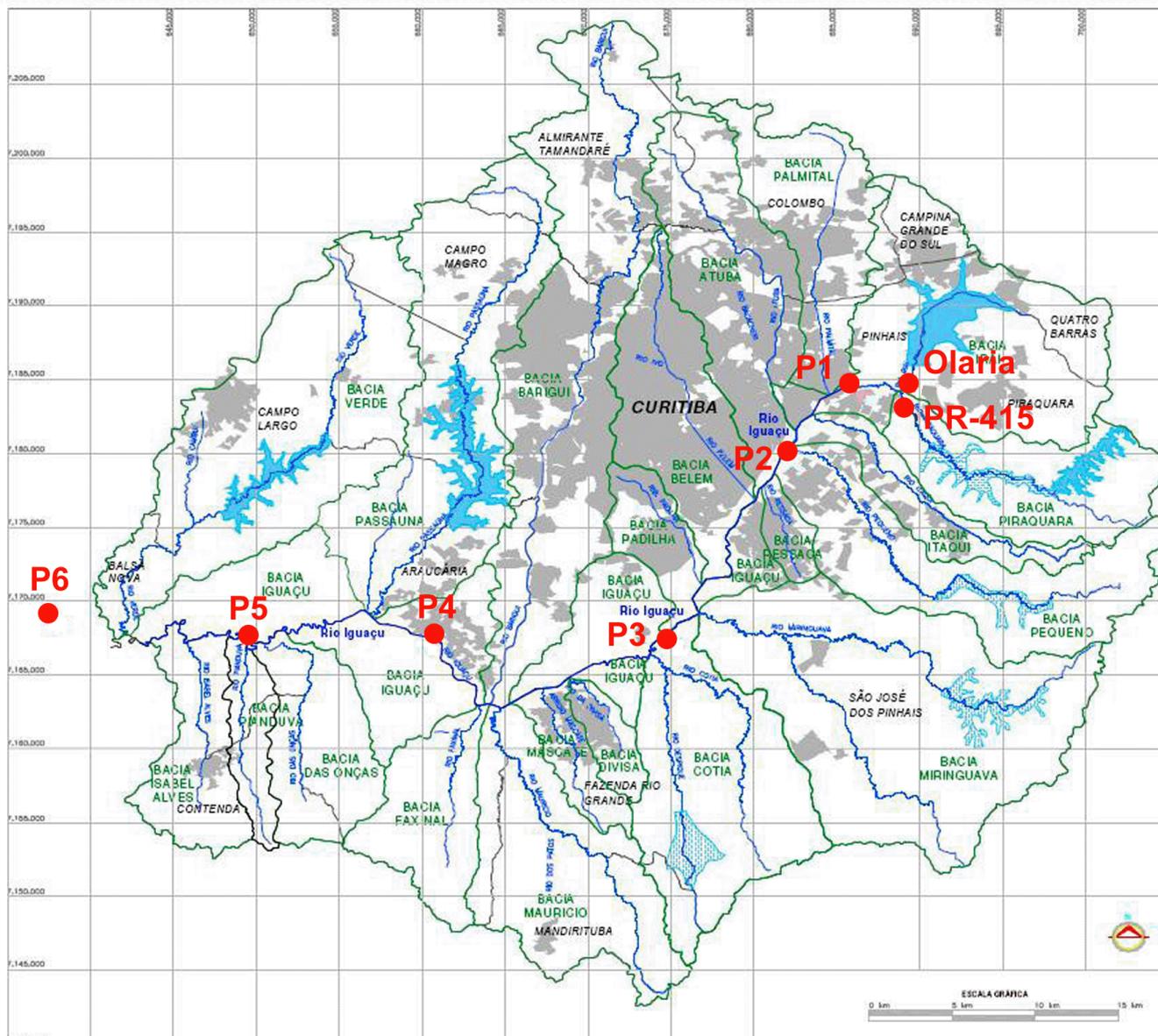
4.4.2 Dendrograma





Estudo de Caso & Aplicação

1 Área de Estudo



2.800 km²
86 km

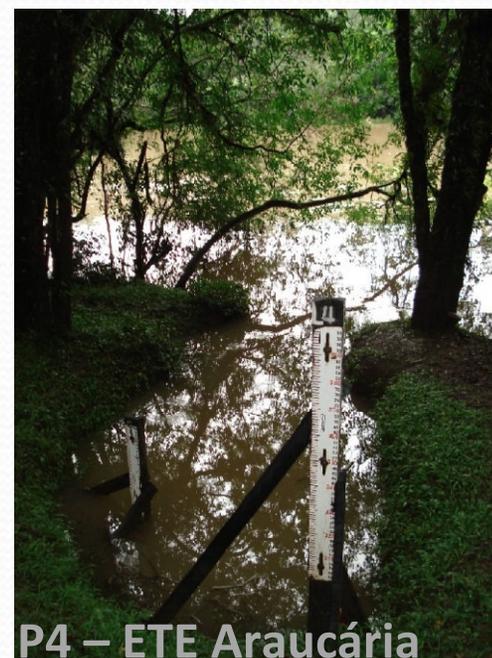
P1 – Início do Canal de Água Limpa



P2 – Ponte BR-277



P3 - Umbarazinho



P4 – ETE Araucária



P5 - Guajuvira

2 Parâmetros de Qualidade de Água Monitorados

Em laboratório

1. DBO₅
2. DQO
3. COT
4. Nitrogênio Orgânico
5. Nitrogênio Amoniacal
6. Nitrito
7. Nitrato
8. Fósforo Total
9. Sólidos Sedimentáveis
10. Sólidos Dissolvidos
11. Sólidos Suspensos

In situ

1. OD
2. Turbidez
3. Condutividade
4. Temperatura da água
5. pH
6. Profundidade Secchi
7. Vazão

3 Atividades de Campo

Ponto de Monitoramento	Nº de campanhas com parâmetros analisados <i>in situ</i>	Nº de campanhas com parâmetros analisados em laboratório	Total de campanhas realizadas
Olaria	21	4	21
P1	24	24	24
P2	24	24	24
P3	24	24	24
P4	24	24	24
P5	23	23	23
P6	21	21	21

4 Aplicação dos Métodos Propostos

Análise de Componentes Principais

MATLAB 5.3 : função “comp2” (ACP a partir da Matriz de Correlação)

Seleção do número de CPs: Critério de Kaiser

Análise Fatorial

MATLAB 5.3: função “normult” e “KMO”

STATISTICA 6.0: *Factor Analysis*

Análise de Agrupamentos

MATLAB 5.3: função “cophenet”

STATISTICA 6.0: *Cluster Analysis*

5 Estratégias de Avaliação

Análise I – Análise Global da Bacia

Variáveis: 18 parâmetros de Q.A.

Matriz: 34 linhas X 18 colunas

Análises: ACP, AF e AA

Análise II – Análise dos Pontos de Monitoramento

Variáveis: 6 pontos de monitoramento (mediana)

Matriz: 18 linhas X 6 colunas

Análises: AF

Análise I – Análise Global da Bacia

	Nº Coleta	Data da Coleta	DQO (mg/L)	DBO ₅ (mg/L)	SDT (mg/L)	SST (mg/L)	SSed (mL/L)	N-A (mg/L)	N-Org (mg/L)	Nitrito (mg/L)	Nitrato (mg/L)	Fósforo (mg/L)	COT (mg/L)	Turbidez (NTU)	Cond (µS/cm)	T (°C)	OD (mg/L)	Secchi (cm)	pH	Vazão (m³/s)
P1	3	20/07/05	64,00	24,00	14	22	0,1	0,11	0,06	0,12	0,27	0,001	6,24	10,75	4,4	13,5	6,60	75	6,09	6,00
	11	14/03/06	12,31	6,12	97	20	0,1	0,65	0,16	0,04	0,35	0,022	7,80	22,31	25,1	22,1	4,88	35	6,67	2,46
	14	26/04/06	11,15	4,44	22	12	0,1	0,11	0,65	0,04	0,45	0,024	6,67	9,13	22,7	19,7	7,06	60	7,01	3,02
	17	07/06/06	14,60	1,98	21	14	0,1	0,11	0,28	0,04	0,38	0,018	5,45	9,94	18,7	16,6	7,44	70	6,70	2,40
	18	21/06/06	13,00	3,32	9	7	0,1	0,16	0,16	0,03	0,10	0,019	8,24	12,32	19,0	15,7	7,20	60	6,72	2,48
	19	19/07/06	6,66	2,58	60	8	0,1	0,33	0,38	0,03	0,05	0,035	3,62	12,10	16,6	15,1	7,52	100	6,56	2,48
P2	4	10/08/05	28,22	23,00	86	87	0,1	6,32	6,32	0,45	0,98	0,200	6,83	74,00	65,1	12,5	7,70	10	6,51	40,50
	13	10/04/06	58,96	40,20	177	34	0,2	11,43	1,47	0,06	0,35	0,123	23,25	10,75	183,6	21,3	1,40	45	7,25	8,60
	16	24/05/06	37,09	13,13	249	20	0,1	8,62	0,55	0,03	0,87	0,518	13,58	4,40	141,7	15,4	1,86	45	7,30	8,60
	17	07/06/06	38,52	17,70	149	32	0,1	8,85	1,34	0,05	0,19	1,820	18,15	12,83	163,1	17,2	2,32	30	7,00	7,76
	18	21/06/06	64,19	24,90	143	18	0,4	1,7	1,26	0,08	0,08	1,500	24,59	14,87	171,1	16,5	2,02	15	7,28	7,55
P3	3	20/07/05	41,60	92,40	181	27	0,1	7,43	0,22	0,44	0,20	0,534	12,22	12,07	126,1	13,3	1,40	50	6,66	12,00
	7	19/10/05	24,80	9,60	81	48	0,4	1,15	1,50	0,34	0,19	0,200	9,74	20,60	72,1	19,1	3,48	45	6,59	24,59
	11	14/03/06	27,69	6,96	120	29	0,1	6,79	4,68	0,05	0,92	0,223	7,70	18,55	114,3	23,4	3,22	20	6,80	30,04
	12	03/04/06	30,28	7,54	175	20	0,1	5,77	1,02	0,02	1,10	0,888	9,78	14,96	73,4	21,1	1,80	40	7,03	20,75
	13	10/04/06	23,90	9,66	133	31	0,3	7,01	1,58	0,04	0,34	0,105	12,53	11,47	125,6	21,4	1,34	50	7,16	9,06
	14	26/04/06	35,06	26,10	119	12	0,1	10,06	1,90	0,07	0,60	1,300	14,96	10,88	139,2	21,2	0,92	40	7,25	8,54
	16	24/05/06	32,26	34,80	101	34	0,5	0,88	0,33	0,03	0,12	0,467	21,22	3,20	120,0	14,6	1,92	35	7,20	8,89
	17	07/06/06	36,43	25,80	147	24	0,1	8,01	1,40	0,06	0,42	1,870	20,83	16,91	163,9	18,0	0,76	40	7,00	9,06
18	21/06/06	35,15	22,35	311	35	0,2	9,54	2,14	0,07	0,09	1,440	45,63	17,79	175,3	16,9	0,56	20	7,33	8,54	
P4	7	19/10/05	9,61	7,92	41	36	0,1	2,31	1,38	0,54	0,19	0,168	6,65	22,38	80,4	19,1	1,82	25	6,62	74,22
	12	03/04/06	12,75	11,20	167	13	0,1	6,05	1,70	0,08	0,44	0,561	11,49	11,34	75,9	20,7	1,76	15	7,06	59,97
	17	07/06/06	31,45	8,53	129	51	0,2	7,56	1,29	0,05	0,95	1,550	21,37	16,87	166,9	18,2	0,66	30	7,10	17,03
P5	7	19/10/05	18,60	9,84	72	41	0,3	2,02	1,10	0,64	0,23	0,141	7,26	25,72	73,4	19,7	2,16	50	6,59	84,08
	13	10/04/06	17,53	7,02	186	41	0,1	5,09	0,85	0,08	0,49	0,054	8,48	21,55	123,3	21,1	1,60	45	7,23	26,70
	14	26/04/06	12,75	7,26	136	11	0,1	4,78	1,58	0,06	0,55	0,820	9,25	16,19	129,9	20,9	2,14	60	7,24	19,30
	16	24/05/06	29,03	6,45	188	47	0,5	0,6	0,05	0,05	0,15	0,232	10,77	13,77	108,1	14,0	1,00	35	7,20	27,00
	17	07/06/06	22,95	8,83	94	35	0,1	6,61	1,18	0,05	0,20	1,230	13,30	12,19	150,8	18,3	0,98	60	7,00	19,98
	18	21/06/06	22,79	13,29	155	15	0,1	8,34	1,98	0,14	0,08	1,010	13,02	13,43	162,1	17,1	0,96	60	7,20	14,60
P6	5	19/10/05	9,34	7,44	81	28	0,1	1,56	1,21	0,77	0,25	0,071	5,71	23,16	62,8	20,0	2,46	40	6,60	103,38
	12	26/04/06	12,54	6,90	135	16	0,1	5,16	0,54	0,13	0,69	0,555	7,17	12,96	113,3	20,8	4,06	40	7,40	20,07
	14	24/05/06	25,80	9,40	241	38	0,1	6,09	0,55	0,03	0,62	0,306	22,65	6,97	99,9	14,7	1,16	40	7,30	30,44
	15	07/06/06	16,85	9,60	53	21	0,1	5,66	1,01	0,06	0,08	0,958	9,12	12,83	135,0	17,8	2,60	50	7,10	8,11
	16	21/06/06	17,65	5,54	255	5	0,1	10,15	1,04	0,09	0,07	0,910	10,40	11,00	149,9	17,6	2,56	60	7,47	16,47

Análise II – Pontos de Monitoramento da Bacia



	P1	P2	P3	P4	P5	P6
DQO (mg/L)	12,66	38,52	32,26	12,75	20,70	16,85
DBO ₅ (mg/L)	3,88	3,88	3,88	8,23	15,42	20,35
Sólidos Dissolvidos Totais (mg/L)	21,50	41,00	41,00	72,79	117,27	145,84
Sólidos Suspensos Totais (mg/L)	13,00	13,00	13,00	17,00	26,00	26,00
Sólidos Sedimentáveis (mL/L)	0,10	0,10	0,10	0,10	0,10	0,10
N-Amoniacal (mg/L)	0,14	0,25	0,25	3,33	7,47	7,47
N-Orgânico (mg/L)	0,22	0,33	0,52	0,47	0,95	1,30
Nitrito (mg/L)	0,04	0,04	0,04	0,04	0,04	0,05
Nitrato (mg/L)	0,31	0,37	0,37	0,37	0,27	0,27
Fósforo (mg/L)	0,02	0,02	0,03	0,08	0,16	0,36
COT (mg/L)	6,46	6,75	6,75	7,54	10,91	15,87
Turbidez (NTU)	11,43	12,21	11,43	11,43	12,21	12,47
Condutividade (μ S/cm)	18,85	20,85	20,85	42,05	103,40	152,40
Temperatura (° C)	16,15	16,15	16,15	15,55	15,55	15,95
OD (mg/L)	7,13	7,32	7,32	7,32	4,76	2,17
Profundidade Secchi (cm)	65,00	60,00	60,00	52,50	45,00	37,50
pH	6,69	6,69	6,71	6,71	6,86	7,13
Vazão (m ³ /s)	2,48	2,48	2,75	5,54	8,18	8,18



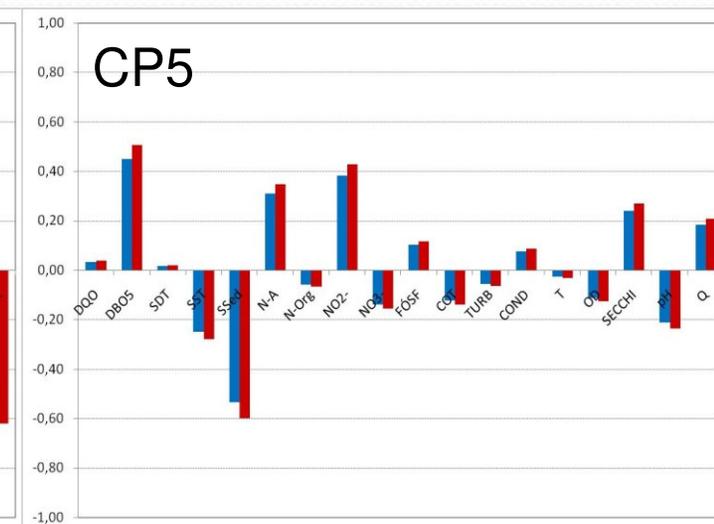
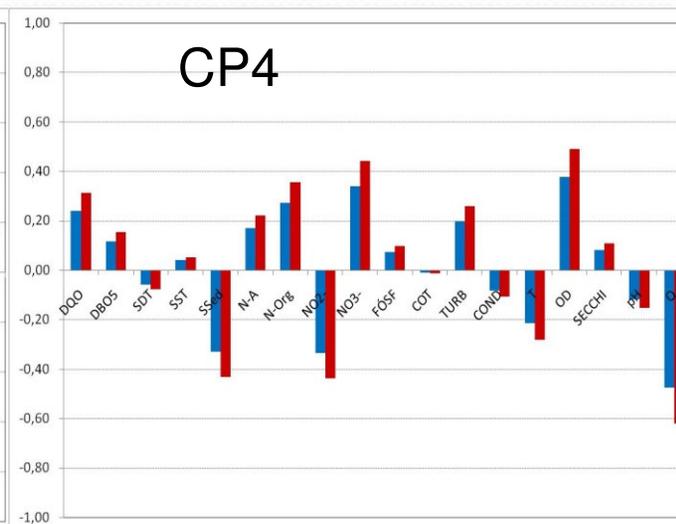
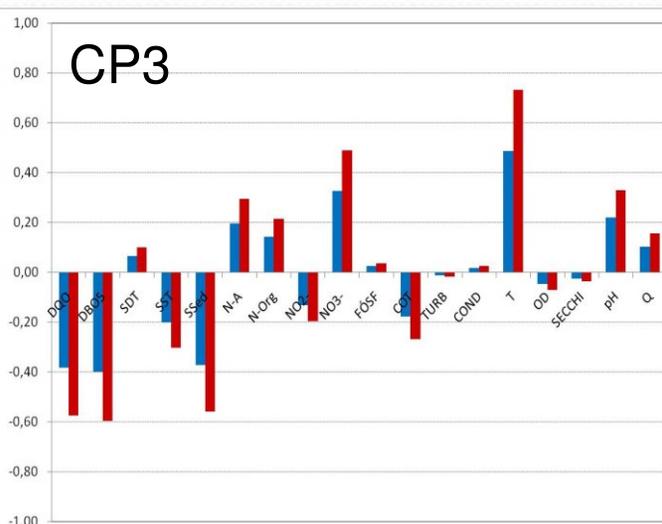
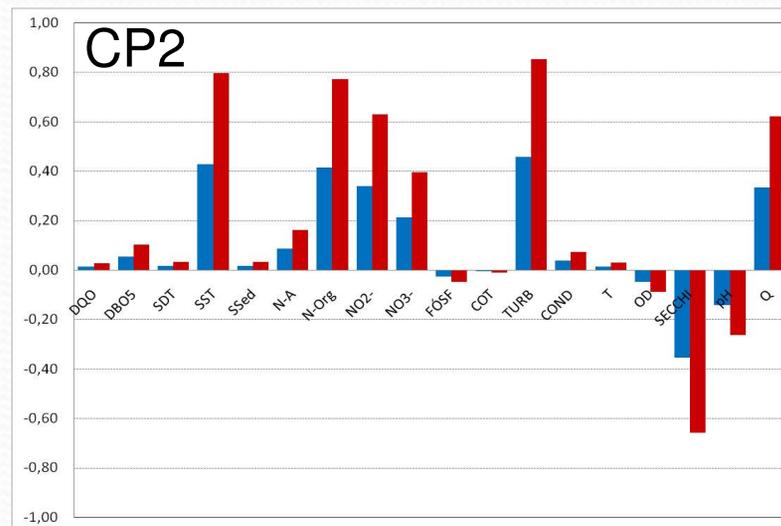
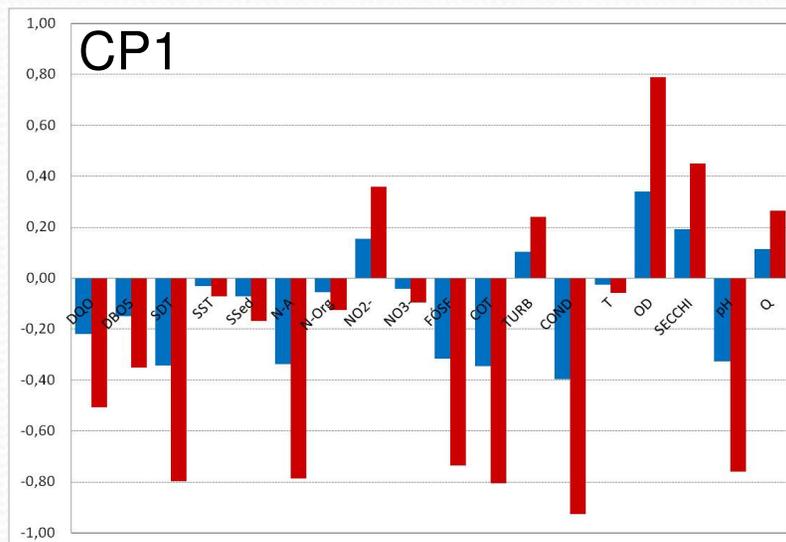
Resultados

1 Análise Global da Bacia

1.1 Análise de Componentes Principais

<i>Componente Principal</i>	<i>Autovalor</i>	<i>Variância Explicada (%)</i>	<i>Variância Explicada Acumulada (%)</i>
1	5,40	30,02	30,02
2	3,46	19,25	49,27
3	2,26	12,53	61,81
4	1,70	9,47	71,28
5	1,26	7,00	78,27
6	0,83	4,59	82,86
7	0,75	4,19	87,05
8	0,53	2,95	90,00
9	0,47	2,60	92,60
10	0,35	1,95	94,55
11	0,29	1,60	96,14
12	0,19	1,04	97,18
13	0,17	0,97	98,15
14	0,16	0,87	99,03
15	0,07	0,40	99,43
16	0,05	0,28	99,70
17	0,04	0,20	99,90
18	0,02	0,10	100,00

Pesos e Correlações das Variáveis



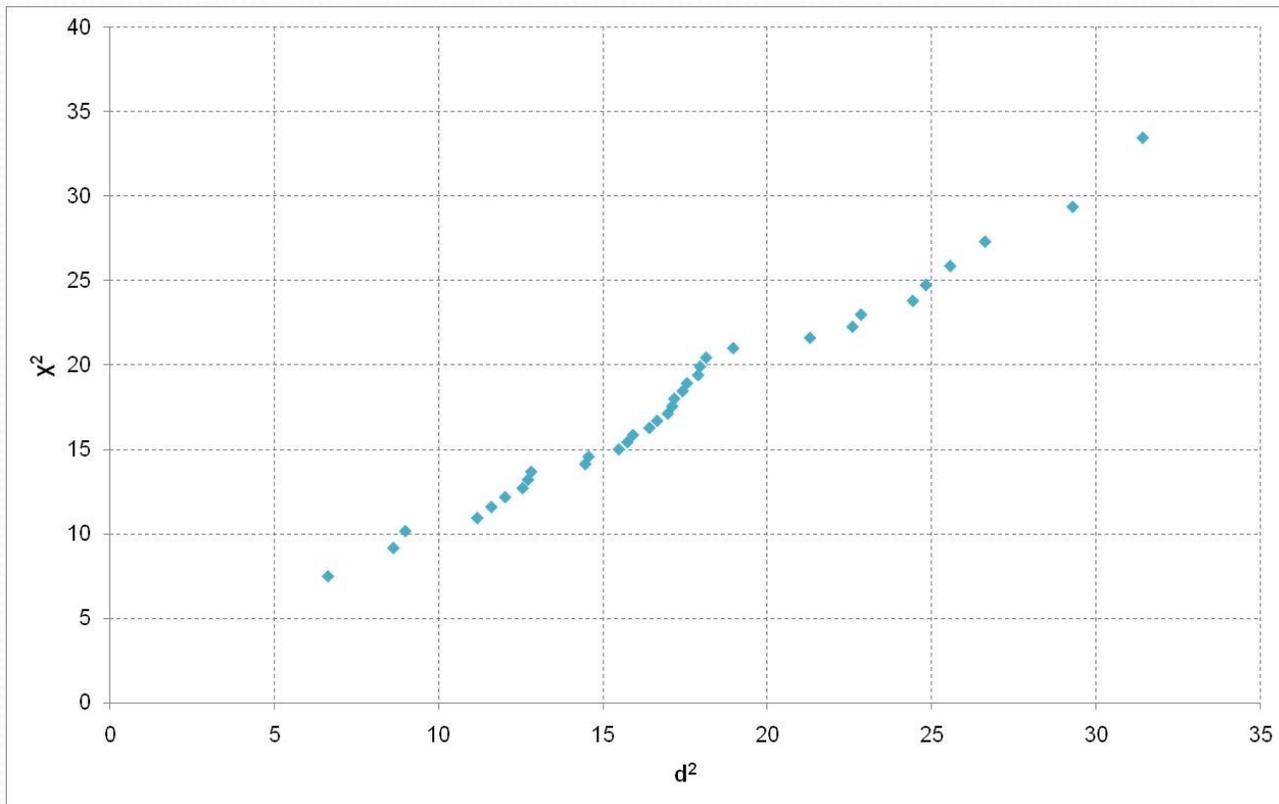
■ Pesos ■ Correlações

CP1: Aspectos da degradação da matéria orgânica e sua interação com a dinâmica de transportes de sólidos

CP2: Caso se aceitasse uma correlação mais baixa, seriam agregados ainda o nitrito e a vazão, destacando a importância do nitrogênio na poluição orgânica

1.2 Análise Fatorial

Verificação da Normalidade Multivariada



Testes prévios

Teste de Bartlett
 $p\text{-valor} < 0.05$, rejeita-se H_0

Medida de KMO
 $MSA > 0.5476$, Ok!

Estimação do Número de Fatores:

Fator	Autovalores	Variância Total Explicada(%)	Variância Total Acumulada (%)
1	5,40	30,02	30,02
2	3,46	19,25	49,27
3	2,26	12,53	61,81
4	1,70	9,47	71,28
5	1,26	7,00	78,27

Matriz dos Pesos das Variáveis nos Fatores

Variável	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
DQO	0,34	0,10	0,70	0,23	-0,14
DBO ₅	0,27	-0,04	0,80	-0,19	0,12
SDT	0,80	0,06	0,03	0,09	-0,02
SST	0,01	0,75	0,23	-0,25	-0,37
SSed	0,10	-0,04	0,13	-0,02	-0,92
N-A	0,80	0,22	0,10	0,07	0,44
N-Org	0,10	0,86	-0,01	-0,09	0,17
NO ₂ ⁻	-0,22	0,18	0,13	-0,92	0
NO ₃ ⁻	0,07	0,64	-0,27	0,23	0,31
FÓSF	0,70	0,02	0,17	0,18	0,11
COT	0,72	0,05	0,31	0,20	-0,29
TURB	-0,26	0,83	0,09	-0,29	0,03
COND	0,93	0,06	0,13	0,04	-0,03
T	0,22	0,02	-0,72	-0,11	0,20
OD	-0,88	0,10	0,05	0,27	0,16
SECCHI	-0,45	-0,63	0,05	0,19	0,29
pH	0,78	-0,13	-0,30	0,33	-0,11
Q	-0,06	0,22	0,30	-0,87	-0,08

Composição dos 5 fatores

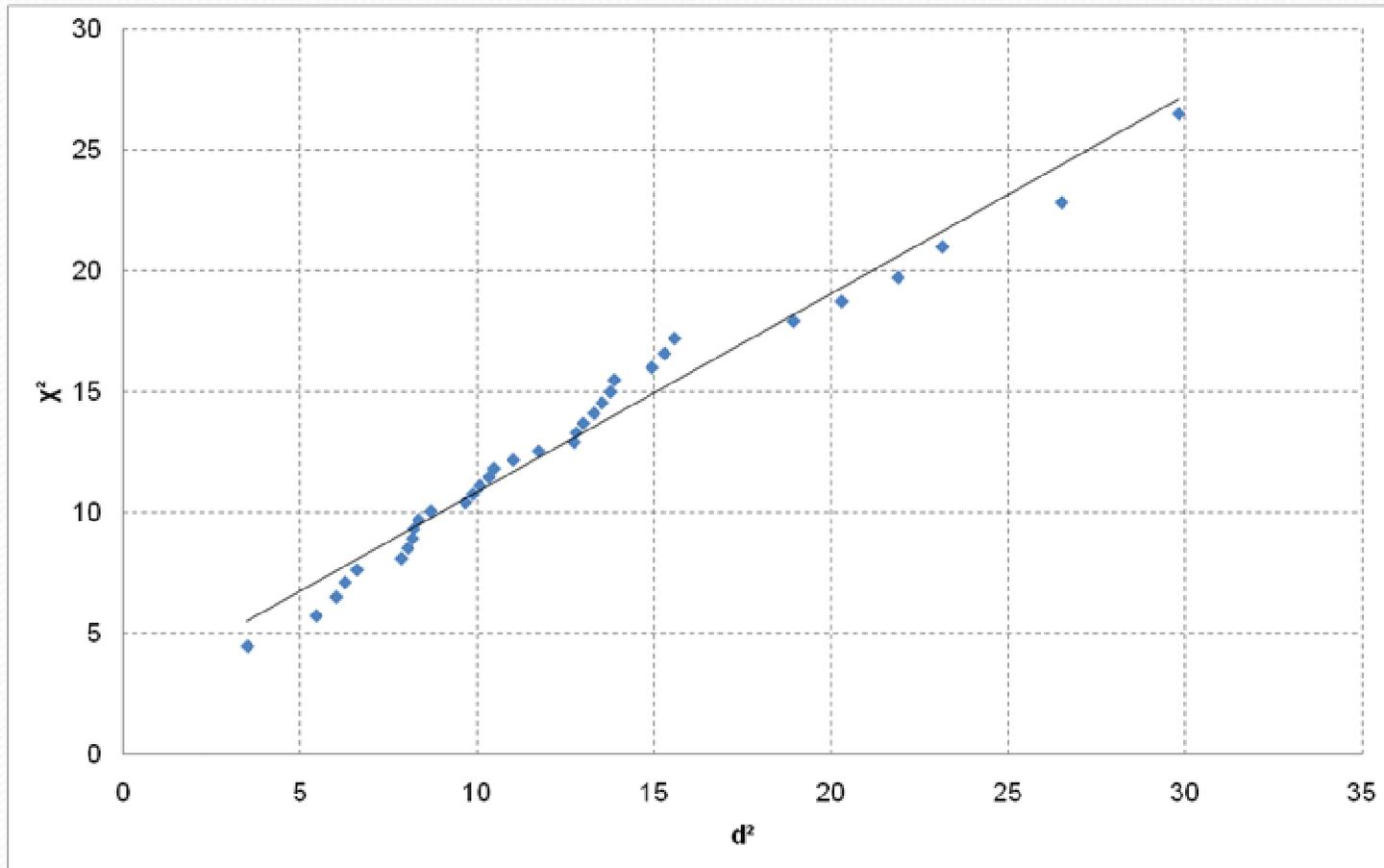
	Variáveis
Fator 1	SDT(+), N-A (+), Fósforo(+),COT(+), Cond (+),pH (+), OD (-)
Fator 2	SST (+), N-Org (+), Turb (+)
Fator 3	DQO(+), DBO ₅ (+), T (-)
Fator 4	NO ₂ ⁻ (-), Q (-)
Fator 5	SSed (-)

Comunalidades

Variáveis	Comunalidades
DQO	0,69
DBO ₅	0,77
SDT	0,65
SST	0,81
SSed	0,88
N-A	0,90
N-Org	0,79
NO ₂ ⁻	0,94
NO ₃ ⁻	0,63
FÓSF	0,57
COT	0,74
TURB	0,85
COND	0,88
T	0,62
OD	0,89
SECCHI	0,72
pH	0,83
Q	0,91

< 0,7 ?
Descarte!

Nova verificação da Normalidade Multivariada



Novos testes de Bartlett e KMO

Método	Variáveis Analisadas	Teste de Esfericidade de Bartlett			Medida de Adequacidade da Amostra de Kaiser-Meyer-Oklin
		χ^2	p-valor	v (grau de liberdade)	MSA
CP	13	328,5001	0	78	0,525

Nova estimação do número de fatores

Fator	Autovalores	Variância Total Explicada(%)	Variância Total Acumulada (%)
1	4,03	31,02	31,02
2	3,35	25,79	56,81
3	1,53	11,77	68,58
4	1,26	9,70	78,29
5	1,14	8,80	87,08

Novos pesos das variáveis nos fatores

Variável	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
DBO ₅	0,18	0,01	0,03	-0,01	0,95
SST	0,03	0,78	0,36	-0,20	0,16
SSED	0,12	0,02	0,94	0,01	0,04
N-A	0,78	0,21	-0,47	0,14	0,21
N-Org	0,12	0,89	-0,25	-0,03	-0,05
NO ₂ ⁻	-0,22	0,24	0,01	-0,87	0,25
COT	0,68	0,11	0,30	0,30	0,25
TURB	-0,27	0,88	-0,08	-0,23	0,02
COND	0,92	0,09	0,01	0,11	0,17
OD	-0,89	0,12	-0,16	0,27	-0,10
SECCHI	-0,47	-0,62	-0,26	0,18	0,04
pH	0,81	-0,16	0,04	0,30	-0,27
Q	-0,01	0,21	0,02	-0,94	-0,19

	Fator	Variância (%)	Variáveis com peso $\geq 10,71$
Fatores com Alta Variância	1	31,02	N-A (+), Cond (+), pH (+), OD (-)
	2	25,79	SST (+), N-Org (+), Turb (+)
Fatores com Baixa Variância	3	11,77	SSed (+)
	4	9,7	NO ₂ ⁻ (-), Q (-)
	5	8,8	DBO ₅ (+)

Novas comunalidades

Variáveis	Comunalidades
DBO ₅	0,94
SST	0,80
SSed	0,90
N-A	0,93
N-Org	0,87
NO ₂ ⁻	0,93
COT	0,72
TURB	0,91
COND	0,90
OD	0,91
SECCHI	0,70
pH	0,85
Q	0,96

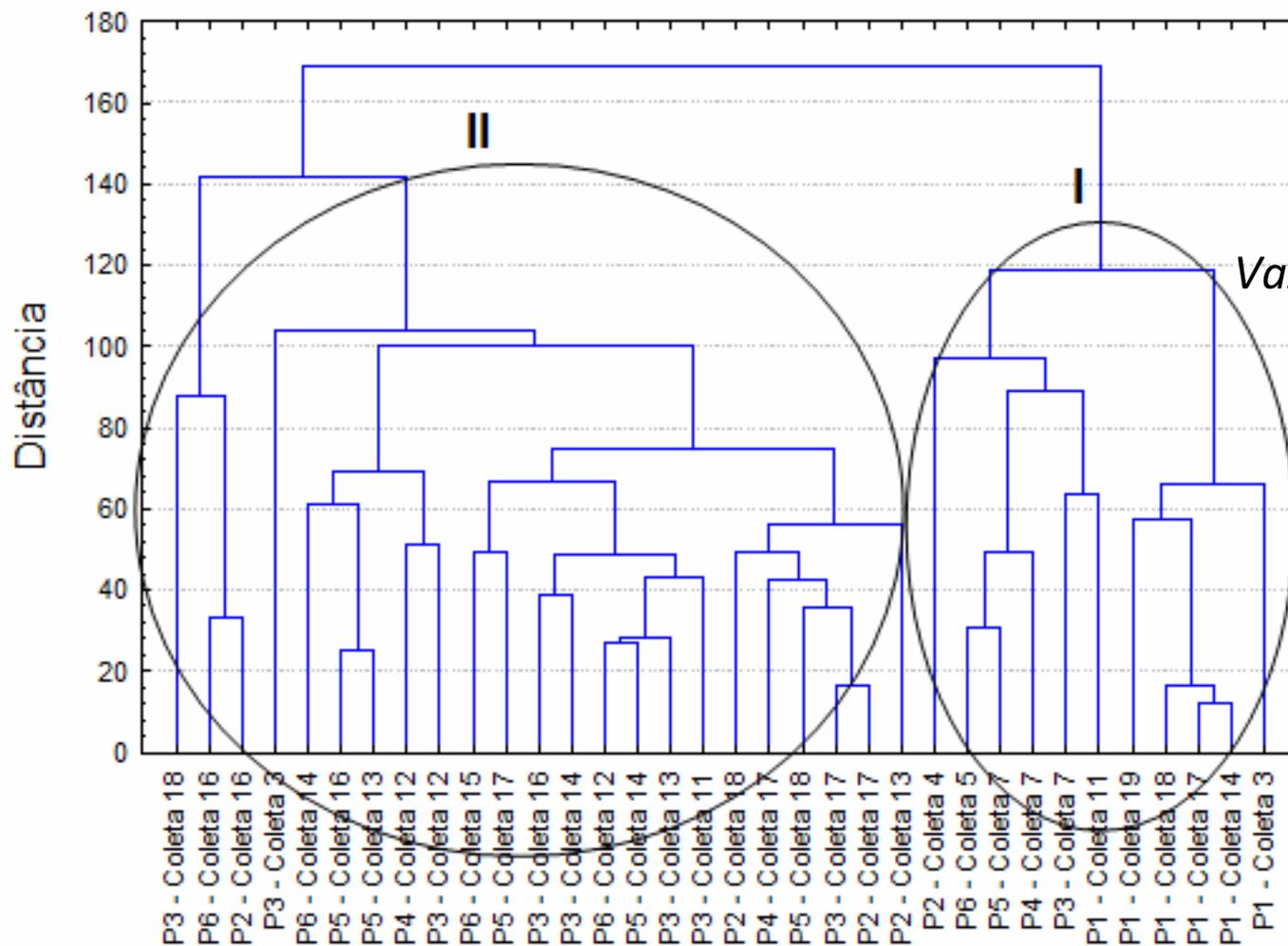
1.3 Análise de Agrupamentos

Variáveis consideradas: Coletas

Correlação cofenética

	LIGAÇÃO				
DISTÂNCIAS	<i>Simples</i>	<i>Completa</i>	<i>Média</i>	<i>Centróide</i>	<i>Ward</i>
<i>Euclidiana</i>	0,6082	0,6568	0,6598	0,6082	0,5148
<i>Quadrado da Dist. Euclidiana</i>	0,5852	0,6160	0,6186	0,6173	0,4807
<i>Cityblock</i>	0,6077	0,5577	0,5971	0,5957	0,5032
<i>Mahalanobis</i>	-	-	-	-	-

Dendrograma



Pior



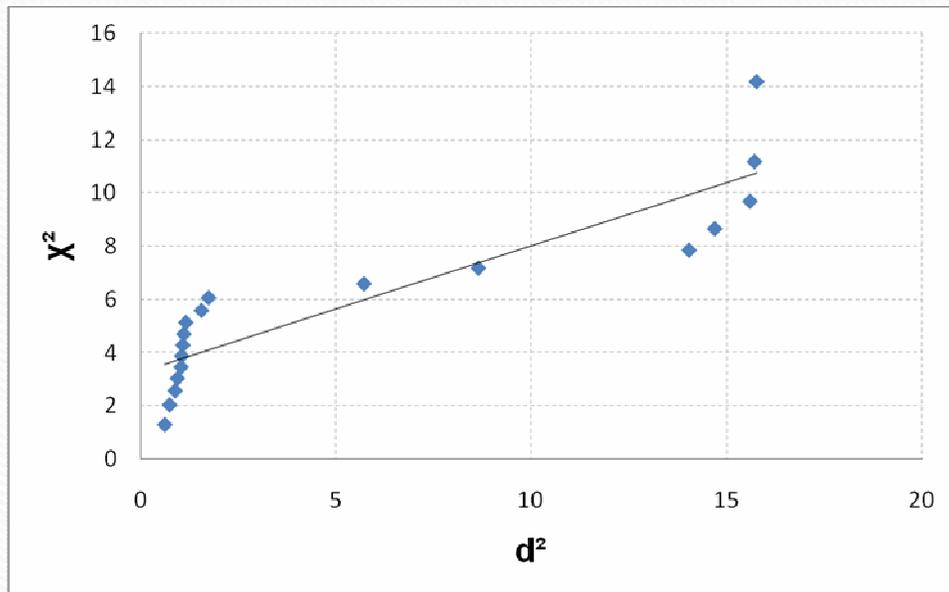
Melhor

2 Resultados da **Análise dos Pontos de Monitoramento**

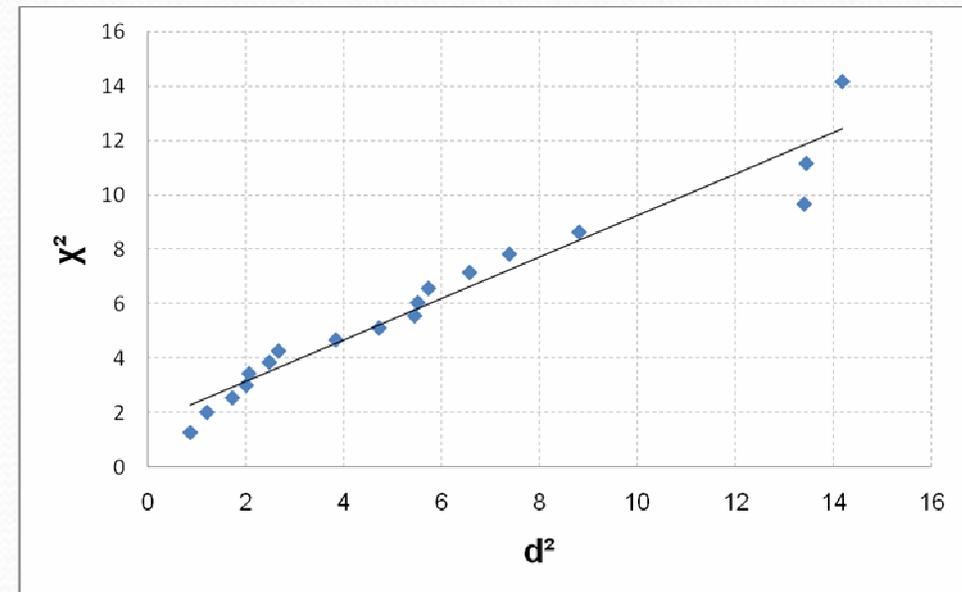
Variáveis consideradas: Pontos de Monitoramento

Técnica: Análise Fatorial

Verificação da Normalidade Multivariada



Dados originais



Log dos dados originais

Testes de Bartlett e KMO

Dados Analisados	Teste de Esfericidade de Bartlett			Medida de Adequacidade da Amostra de Kaiser-Meyer-Oklin
	χ^2	p-valor	v(grau de liberdade)	MSA
Logaritmo dos dados originais	313,9343	0	66	0,7587

Estimação do número de fatores

Fator	Autovalores	Variância Total Explicada(%)	Variância Total Acumulada (%)
1	5,697	94,95	94,95
2	0,2741	4,57	99,52

Pesos das variáveis nos fatores

Variável	Fator 1	Fator 2
P1	0,86	0,5
P2	0,84	0,53
P3	0,84	0,54
P4	0,66	0,74
P5	0,55	0,83
P6	0,48	0,87



Assim, a análise indicou que os pontos de monitoramento considerados mais relevantes foram o P1, P2 e P3 (95% da variância total da amostra).

Estes pontos se encontram na primeira porção da bacia, onde se localizam as sub-bacias mais impactantes em termos de fontes poluição como as bacias do Belém, do Atuba, do Palmital, do Padilha e do Ressaca.



Conclusões e Recomendações

1 Conclusões

- Parâmetros mais relevantes: OD, Nitrogênio Amoniaco, Condutividade, pH, SST, Nitrogênio Orgânico e Turbidez, os quais ilustram os aspectos de degradação da matéria orgânica e sua interação com a dinâmica de transporte de sólidos;
- Os resultados indicaram como relevantes os parâmetros de impacto e não os de efeito como a DBO, DQO e COT.;

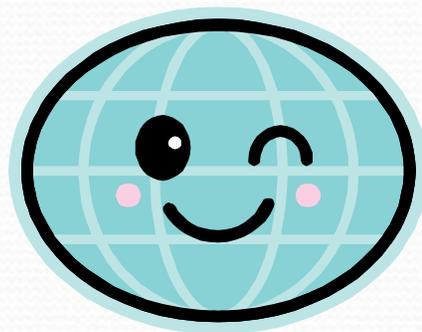
- Grupo de Coletas I: coletas que refletiram melhor qualidade do corpo hídrico, formado principalmente por coletas realizadas no ponto P1, próximo a uma área de manancial da bacia; Grupo de Coletas II: formado por grande parte das outras coletas, as quais refletiram o estado de degradação do rio, evidenciando que em sua totalidade, a qualidade da água da bacia apresenta-se inadequada.
- Pontos de Monitoramento mais relevantes: P1, P2 e P3, os quais se encontram na primeira porção da bacia, onde se localizam as sub-bacias mais impactantes em termos de fontes poluição como as bacias do Belém, do Atuba, do Palmital, do Padilha e do Ressaca

2 Recomendações

- Desenvolver técnicas para análise de consistência de dados de QA similares às técnicas de preenchimentos de falhas para as séries hidrológicas, visando avaliar o impacto de não se desprezar dados obtidos de procedimentos amostrais tradicionais e, que em geral, são caros e complexos;
- Realizar novas análises considerando uma base de dados maior, de forma a complementar os resultados aqui obtidos;
- Incluir a variável *precipitação (mm)* em estudos futuros, buscando sua relação com a vazão e os demais parâmetros de qualidade de água;

- 
- Realizar a ACP e a AF para cada ponto de monitoramento individualmente, visando encontrar quais são os parâmetros de QA mais importantes para cada um dos pontos monitorados;
 - Separar os dados de acordo com os períodos de cheia e estiagem e realizar novas análises para estes 2 conjuntos de dados, procurando investigar se os parâmetros de QA mais relevantes na bacia diferem em razão das características hidrológicas.

Muito Obrigada!



marianne.schaefer@gmail.com