

UNIVERSIDADE FEDERAL DE MINAS GERAIS
CURSO DE ESPECIALIZAÇÃO EM SANEAMENTO E MEIO AMBIENTE

MONOGRAFIA DE FINAL DE CURSO

**CONSTRUÇÃO E TRATAMENTO DO BANCO DE DADOS
DE QUALIDADE DA ÁGUA DA UHE NOVA PONTE**

Déborah Tavares Viana

Belo Horizonte
2011

Déborah Tavares Viana

CONSTRUÇÃO E TRATAMENTO DO BANCO DE DADOS DE QUALIDADE DA ÁGUA DA UHE NOVA PONTE

Monografia apresentada ao Curso de Especialização em Saneamento e Meio Ambiente da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Saneamento e Meio Ambiente

Área de concentração: Controle Ambiental na Indústria
Linha de pesquisa: Impacto Ambiental, Qualidade da água

Orientadora: Profa. Dra. Mônica M^a Diniz Leão

Belo Horizonte

Escola de Engenharia da UFMG

2011

RESUMO

A Usina Hidrelétrica de Nova Ponte está localizada na bacia do rio Paranaíba. Como medida de controle ambiental a Companhia Energética de Minas Gerais – CEMIG realiza monitoramentos periódicos da qualidade da água do reservatório e cursos d'água. O gerenciamento de lagos é de extrema importância já que o empreendimento altera a condição física da água de lântica para lótica. Os monitoramentos são realizados durante longos períodos e com frequência pré definida. Dessa forma, o número de informações que deve ser armazenada é muito grande. O armazenamento e análise das informações obtidas são etapas que necessitam de estudos e organização. Este trabalho procurou compilar os dados de qualidade da água da UHE Nova Ponte em forma de banco de dados. Posteriormente, foi realizado um estudo e aplicação de algumas formas de tratamento de dados. O número de dados faltantes, censurados e *outliers* foi levantado para a análise prévia dos dados. Ficou claro que o planejamento e organização prévia para a construção de um banco de dados com qualidade é uma etapa essencial e não pode ser realizada de maneira superficial. Os gráficos Boxplot construídos mostram que a qualidade da água de ambientes lóticos varia de forma diferente da qualidade da água de ambientes lânticos, sugerindo uma análise individual para cada ambiente. Percebe-se, ainda, que a qualidade da água não apresenta uma variação negativa perceptível. São necessárias, portanto, análises estatísticas para levantar as conclusões e relacionar os pontos de forma mais eficiente. Nesse trabalho é realizada a análise preliminar, não podendo tomar como conclusivas as hipóteses levantadas. Conclui-se que a análise preliminar é uma etapa importante para uma avaliação adequada e, principalmente, para o efetivo tratamento do banco de dados.

Palavras Chave: Tratamento de dados; monitoramento; lagos; análise preliminar.

SUMÁRIO

1 INTRODUÇÃO	ERRO! INDICADOR NÃO DEFINIDO.
2 OBJETIVOS	ERRO! INDICADOR NÃO DEFINIDO.
2.1 OBJETIVO GERAL.....	ERRO! INDICADOR NÃO DEFINIDO.
2.2 OBJETIVOS ESPECÍFICOS	ERRO! INDICADOR NÃO DEFINIDO.
3 REVISÃO BIBLIOGRÁFICA	ERRO! INDICADOR NÃO DEFINIDO.
3.1 A QUALIDADE DA ÁGUA E AS INSTITUIÇÕES TRABALHADAS	ERRO! INDICADOR NÃO DEFINIDO
3.2 ESTUDOS APLICADOS À QUALIDADE DA ÁGUA DE RESERVATÓRIOS	ERRO! INDICADOR NÃO DE
3.2.1 Métodos estatísticos utilizados na análise da qualidade da água	Erro! Indicador não def
3.3 O MONITORAMENTO DA QUALIDADE DA ÁGUA ...	ERRO! INDICADOR NÃO DEFINIDO.
3.4 O MONITORAMENTO COMO FERRAMENTA DE GESTÃO	ERRO! INDICADOR NÃO DEFINIDO.
3.5 TRATAMENTO DE DADOS: QUALIDADE DOS DADOS	ERRO! INDICADOR NÃO DEFINIDO.
3.6 DADOS CENSURADOS	ERRO! INDICADOR NÃO DEFINIDO.
4 METODOLOGIA	ERRO! INDICADOR NÃO DEFINIDO.
4.1 CARACTERIZAÇÃO DA ÁREA	ERRO! INDICADOR NÃO DEFINIDO.
4.2 CRIAÇÃO DO BANCO DE DADOS.....	ERRO! INDICADOR NÃO DEFINIDO.
4.3 DEFINIÇÃO DE PARÂMETROS TRABALHADOS.....	ERRO! INDICADOR NÃO DEFINIDO.
4.4. DEFINIÇÃO DOS PONTOS TRABALHADOS	ERRO! INDICADOR NÃO DEFINIDO.
4.5 VALIDAÇÃO DO BANCO E ANÁLISE PRELIMINAR....	ERRO! INDICADOR NÃO DEFINIDO.
4.5.1 Análise preliminar.....	Erro! Indicador não definido.
4.5.2 Outliers.....	Erro! Indicador não definido.
4.5.3. Dados faltantes.....	Erro! Indicador não definido.
4.5.4 Dados censurados.....	Erro! Indicador não definido.
5 RESULTADOS E DISCUSSÃO	ERRO! INDICADOR NÃO DEFINIDO.
5.1 A CONSTRUÇÃO DO BANCO	ERRO! INDICADOR NÃO DEFINIDO.
5.2 ANÁLISE PRELIMINAR DA QUALIDADE DA ÁGUA ...	ERRO! INDICADOR NÃO DEFINIDO.
5.2.1 Considerações iniciais	Erro! Indicador não definido.
5.2.2 Gráficos Boxplot.....	Erro! Indicador não definido.
6 CONCLUSÕES	ERRO! INDICADOR NÃO DEFINIDO.
REFERÊNCIAS	ERRO! INDICADOR NÃO DEFINIDO.

LISTA DE TABELAS E FIGURAS

- Tabela 1 - Descrição dos Pontos de amostragem **Erro! Indicador não definido.**
- Tabela 2 - Cálculos do total de coletas **Erro! Indicador não definido.**
- Tabela 3- Porcentagens de faltantes, censurados e outliers por parâmetro **Erro! Indicador não definido.**
- Tabela 4- Configuração geral dos dados selecionados **Erro! Indicador não definido.**
-
- Figura 1- Aspectos para definição de uma rede de monitoramento **Erro! Indicador não definido.**
- Figura 2 - Organização do Banco de Dados **Erro! Indicador não definido.**
- Figura 3- Gráfico de nº de coletas por ano **Erro! Indicador não definido.**
- Figura 4- Número de coletas por estações **Erro! Indicador não definido.**
- Figura 5- esquema da localização dos pontos no reservatório de Nova Ponte **Erro! Indicador não definido.**
- Figura 6 - Gráficos Boxplot para os parâmetros selecionados **Erro! Indicador não definido.**

LISTA DE SIGLAS E ABREVIações

UHE – Usina Hidrelétrica

LO – Licença de Operação

IGAM – Instituto Mineiro de Gestão das Águas

CEMIG – Companhia Energética de Minas Gerais

SISÁGUA – Sistema de Informação de Qualidade da Água dos Reservatórios da Cemig

OD- Oxigênio Dissolvido

AA – Análise de Agrupamentos

AC – Análise de *Cluster*

ACP – Análise dos Componentes Principais

ANOVA – Análise de Variância

MANOVA – Análise Multivariada de Variância

CP – Componentes Principais

DL – Dados de Laboratório

LD – Limite de Detecção

1 INTRODUÇÃO

A Usina hidrelétrica (UHE) Nova Ponte pertencente à Bacia Hidrográfica do rio Paranaíba, em setembro de 1993 obteve aprovação da Licença de Operação (LO) do empreendimento (ÁGUA E TERRA, 2006). Desde os estudos prévios para implantação da barragem, campanhas de monitoramento da qualidade da água do reservatório e entorno são realizadas. O presente trabalho trata sobre a compilação das informações da qualidade da água do reservatório para organizá-las em formatação de banco de dados.

Segundo Lemos (2011), o gerenciamento de lagos e reservatórios deve ser realizado, visando à proteção e manutenção de uma boa qualidade da água, a vida aquática e todos os outros usos preconizados. Visando o melhor gerenciamento, o monitoramento deve ser bem planejado e seus objetivos devem estar claros desde sua implantação.

Um grande problema no âmbito do monitoramento é a armazenagem e análise eficiente dos dados obtidos (CARLETON; DAHLGREN; TATE, 2005). Dessa forma, o desenvolvimento de banco de dados e novas metodologias de análises devem ser desenvolvidas. Nesse trabalho é feito somente a avaliação preliminar, etapa essencial para assegurar a qualidade de todas as fases seguintes, como o tratamento de dados, validação dos dados, análise estatística, etc.

Durante a construção do banco de dados e sua análise são encontrados muitos desafios. O grande número de dados faltantes, dados censurados e *outliers* tornam a análise preliminar uma etapa indispensável para o tratamento e “limpeza” do banco de dados. Os erros e problemas encontrados nos bancos de dados são vistos como ruins, mas também servem como uma rica fonte de informação para o controle da qualidade dos dados e, principalmente, para prever e evitar problemas futuros.

O objetivo principal do trabalho é mostrar a importância do planejamento e organização prévia para a construção de um banco de dados de qualidade. Será apresentado um estudo aplicado, onde foi realizada a compilação de dados, criação de um banco de dados e análises preliminares. São, então, apontadas falhas organizacionais que trazem muitas dificuldades nos processos de

análise e leitura de dados. As informações de qualidade de água consideradas nessa monografia não são definitivas, já que se trata de um estudo preliminar. Análises e tratamento de dados mais aprofundados ainda devem ser realizados para chegar a conclusões e discussões confiáveis.

2 OBJETIVOS

2.1 Objetivo Geral

Fazer uma análise preliminar dos dados de qualidade da água do reservatório de Nova Ponte, com o foco em tratamento de dados na criação de um banco de dados.

2.2 Objetivos específicos

- Realizar uma revisão bibliográfica sobre monitoramento da qualidade da água e análises de qualidade da água;
- Realizar uma revisão bibliográfica sobre o tema “tratamento de banco de dados”;
- Reunir e organizar em forma de banco de dados as informações referentes à qualidade da água do reservatório de Nova Ponte junto à Cemig;
- Fazer um levantamento dos dados faltantes, censurados e *outliers* do banco de dados;
- Aplicar um pré tratamento ao banco de dados;
- Realizar uma avaliação preliminar dos dados de qualidade da água do reservatório de Nova Ponte.

3 REVISÃO BIBLIOGRÁFICA

3.1 A Qualidade da Água e as Instituições Trabalhadas

A qualidade das águas vem se tornando um assunto em discussão desde a década de 70 no Brasil. A escassez de água e os altos índices de poluição encontrados tornam o recurso cada dia mais valioso. Nesse contexto, surge a necessidade de estudos e investimentos para o acompanhamento da qualidade das águas no nosso país. Só é possível preservar e gerir aquilo que se conhece, portanto pesquisas são essenciais para atingir o desenvolvimento sustentável. O estado de Minas Gerais lançou o projeto Águas de Minas em 1997, que monitora e disponibiliza os dados da qualidade das águas superficiais do estado. O Instituto Mineiro de Gestão das Águas – IGAM é responsável pela realização desse projeto, o qual contribui significativamente para o melhor entendimento de como e quais fatores podem interferir na qualidade da água dos cursos d'água.

Ainda visando o melhor conhecimento da qualidade da água, surge a necessidade de entender como os reservatórios de hidrelétricas afetam a qualidade e o ecossistema aquático. O Brasil tem como fonte de energia principal a energia hidrelétrica, portanto o número de reservatórios continua aumentando para atender a demanda crescente. A Companhia Energética de Minas Gerais (CEMIG) possui em Minas Gerais 40 reservatórios de usinas hidrelétricas (CEMIG, 2010) e a formação dos reservatórios pode afetar todo o ecossistema do local. Para a obtenção das licenças prévias, de instalação e operação das usinas hidrelétricas, são necessários estudos e monitoramentos constantes da qualidade da água. Os estudos são devidamente realizados e devem ser utilizados como ferramenta de gestão da qualidade das águas, porém muita informação acaba se perdendo no universo de tantos dados. Para melhor aproveitamento desses dados a Cemig lançou o Sistema de Informação da Qualidade das Águas de Reservatórios - Siságua que visa contribuir para o conhecimento científico dos ecossistemas represados. Este trabalho portanto, reuni informações dessas duas instituições (CEMIG e IGAM), que são muito importantes no cenário mineiro de conservação dos ambientes aquáticos.

3.2 Estudos Aplicados à Qualidade da Água de Reservatórios

Estudos em reservatórios contribuem para conservação e gestão apropriada do meio ambiente entorno. Natale e Leis (2008), em um estudo realizado no reservatório de Yacyretá na Argentina, mostram como a qualidade da água de reservatórios pode ser alterada pelas atividades humanas, como agricultura, ocupação e atividades da indústria. A qualidade pode diminuir com as atividades humanas e os parâmetros físico-químicos podem ser ótimos indicadores dessas mudanças (NATALE e LEIS, 2008; NADIM *et al*, 2007 e BRANCO *et al*, 2009). Em alguns estudos também é observado que as alterações dos parâmetros físico-químicos podem vir de efeitos naturais. Pesquisas demonstram que a evolução dos cursos d'água pode ter mudanças que são naturais e acontecem por motivos sazonais (MWAURA, 2003 e RANGEL-PERASA *et al*, 2009). As hipóteses levantadas inicialmente podem, muitas vezes, ser rejeitadas por estudos de longo termo. LORANTOU e colaboradores (2007) mostram que a qualidade da água em um reservatório na Bélgica é melhor do que previamente esperado. Também não é descartada a necessidade de estudos contínuos para que afirmações como essas possam ser feitas. De qualquer forma, é preciso o constante monitoramento e estudos completos para que as hipóteses possam ser comprovadas ou descartadas definitivamente (LORANTOU, A.; THOMÉ, J. e GOFFART, A., 2007).

A usina hidrelétrica de Nova Ponte começou sua operação em 1994 e assim surgiram diversas alterações no ecossistema do local. Alguns estudos sobre a comunidade de peixes e seu comportamento no reservatório de Nova Ponte já foram realizados (RÊGO *et al*, 2008; FELTRAM *et al*, 2004). Estudos relacionados à qualidade da água do reservatório são mais raros, porém já foram realizados. Em 2003, GOMES e SOUZA e VON SPERLING realizam um estudo na bacia do Araguari que relaciona a qualidade da água com as comunidades zooplânctônicas. Esse levantamento mostra que são necessários estudos temporais dos parâmetros físico-químicos para preencher essa lacuna em relação aos dados da UHE Nova Ponte.

3.2.1 Métodos estatísticos utilizados na análise da qualidade da água

Em geral, um programa de monitoramento inclui coletas freqüentes nos mesmos pontos de amostragem e análise em laboratório de grande número de parâmetros, resultando em matriz de grandes dimensões e complexa interpretação. (ANDRADE *et al*, 2007).

As análises temporais e espaciais são ferramentas que possibilitam a avaliação de séries de dados longos de forma resumida e completa. Essas ferramentas são boas formas de verificar possíveis mudanças que o crescimento e desenvolvimento acelerado estão causando nos ecossistemas aquáticos. Sundaray e colaboradores, (2006), através de análises multivariadas, defendem que as alterações de concentrações nos nutrientes das águas são, além de sazonais, muito ligadas a ocupação do solo. Esse estudo mostra que a poluição altera diretamente as concentrações de nutrientes, OD e pH. Na mesma linha de pesquisa, em estudo conduzido por BU e colaboradores (2010) foi verificado também diferenças temporais e principalmente espaciais no curso d'água, essas disparidades estão relacionadas com a ocupação do solo. Também foi possível estabelecer agrupamentos de acordo com o nível de poluição, os pontos no grupo de alta poluição estavam sempre próximos a grandes centros ou indústrias, o que já era esperado. Diante dessa realidade, surge a necessidade de estudos temporais para acompanhamento e predição da evolução da qualidade da água objetivando que as devidas medidas preventivas e emergenciais sejam tomadas e a manutenção do equilíbrio do ecossistema seja garantida.

Para garantir um bom estudo temporal e espacial é importante definir os objetivos da pesquisa, para que fique claro objetivo principal. O próximo passo seria levantar quais variáveis vão ser utilizadas para o estudo. Por último, é feita a seleção da ferramenta estatística que vai ser utilizada para o tratamento dos dados. Para iniciar um estudo estatístico, VON SPERLING (2007), defende a utilização de estatísticas básicas para a descrição inicial das variáveis. Esse conhecimento inicial da matriz dos dados pode ajudar na escolha do modelo estatístico ideal para o estudo. A caracterização dos dados através de estatísticas básicas dá ao leitor uma visão imediata das médias e variações, por isso devem ser incluídas em series históricas (VON SPERLING, 2007). Para a análise dos dados, devem ser utilizadas ferramentas específicas de estatística, como as análises multivariadas. Elas permitem analisar uma série de dados, de forma a facilitar o entendimento do todo. As análises multivariadas são muito usadas para caracterização e avaliação da qualidade de rios e lagos, pois evidenciam variações temporais causadas por processos naturais ou poluentes (JAYAKUMAR E SIRAZ, 1997; VEGA *et al.*, 1998; HELENA *et al.*, 2000; SINGH *et al.*, 2004).

Para estudos de series temporais as análises do tipo: Análise de Agrupamentos (AA), também conhecida como Análise de *Cluster* (AC), Análise dos Componentes Principais (ACP), Análise

de Variância (ANOVA) e Análise Multivariada de Variância (MANOVA), são muito utilizadas e cada uma apresenta uma série de vantagens e desvantagens que devem ser estudadas para cada tipo de estudo.

A análise de agrupamentos constitui uma metodologia numérica multivariada, com o objetivo de propor uma estrutura classificatória, ou de reconhecimento da existência de grupos, objetivando, mais especificamente, dividir o conjunto de observações em um número de grupos homogêneos, segundo algum critério de homogeneidade (VICINI apud REGAZZI, 2001). SIMEONOVA e EINAX (2002) salientam a utilidade da AA como uma forma de detectar similaridades e dissimilaridades em um grupo grande de objetos.

O método de ACP é usado para identificar quais variáveis explicam a maior variância, portanto contém informações sobre um pequeno grupo de dados que englobam grande número de variáveis (LINDMAN, 1974). A ACP reduz o número de variáveis agrupando diversas em um só grupo, as componentes principais (CP). Essa transformação, em outro conjunto de variáveis, ocorre com a menor perda de informação possível, sendo que esta também busca eliminar algumas variáveis que possuem pouca informação. (VICINI, 2005).

Por último, a análise de variância- ANOVA que é usada para comparar duas variáveis em situações que uma variável influencia a outra (BARAN; WARRY, 2008). Por ser um teste paramétrico, o grupo deve ter distribuição normal ou próxima do normal. É importante destacar que o método estatístico deve ser definido de acordo com a pesquisa, já que nem todos os testes são aplicáveis para o caso.

3.3 O Monitoramento da Qualidade da Água

O monitoramento dos dados ambientais, especificamente da qualidade da água, se faz necessário em função do número de informações diretas e indiretas que ele pode gerar. A existência de séries históricas e sua interpretação, permitem uma melhor compreensão da evolução temporal da qualidade ambiental e sua correlação com outros fenômenos, climáticos ou econômicos. (LAMPARELLI, 2004).

Segundo Barbosa (199?), a principal importância de um programa de monitoramento de qualidade da água é identificar as mudanças nas variáveis bióticas e abióticas, as quais afetam a estrutura e funcionamento dos ecossistemas. Além disso, o monitoramento ambiental surge para responder a alguns questionamentos. Assim, é necessário que o objetivo do monitoramento esteja claro antes de iniciá-lo. Os objetivos podem ser baseados em requisitos operacionais ou gestacionais; podem responder a variadas demandas, já que são ferramentas para estabelecer padrões de qualidade, determinar tendências de qualidade, identificar trechos semelhantes ou não e identificar fontes e causas de degradação da qualidade.

As redes de monitoramento já foram, inicialmente, desenvolvidas a partir das experiências da equipe que está planejando a rede, sem nenhuma metodologia identificável (PARKA et al., 2006). Com o desenvolvimento técnico-científico, há um crescente número de redes sendo planejadas a partir de objetivos específicos pré estabelecidos, utilizando modelos espaciais, matemáticos, neurais, genéticos, entre outros. Para Barbosa (199 os estudos de qualidade da água devem ter o monitoramento planejado, com uma política específica de controle de qualidade que possa assegurar a confiança pública. Os resultados desses monitoramentos devem ser publicados e apresentarem uma frequência estabelecida.

Como as redes de monitoramento iniciaram sem um objetivo claro e bem definido muitas delas são ineficientes. Além disso, é possível perceber que os diversos setores (órgãos ambientais, empresas, instituições públicas, etc) instalam sua rede de monitoramento sem o conhecimento prévio de redes já existentes. Esse fato levou a uma sobreposição de redes e o mal planejamento dessas, gerando grande número de dados que não podem ser analisados devidamente por falta de planejamento prévio . Esse cenário ocasiona a perda do grande objetivo das redes de monitoramento ambientais, que devem dar informações de curto e longo prazo, variações espaciais e se tornarem bases para avaliações de qualidade ambiental e para o estabelecimento de medidas de proteção e controle ambiental (PARKA et al., 2006).

O ideal é que os programas de monitoramento analisem o ecossistema em escala global. Como nem sempre essa escala pode ser atingida, Barbosa (199?) diz que é importante que pelo menos quatro aspectos sejam considerados: saúde, qualidade físico-química da água, biomonitoramento e ecotoxicologia.

Para definir as variáveis analisadas é importante que elas contenham o maior número de informações possíveis, sejam simples de medir e analisar e tenham a maior relação custo/benefício (BARBOSA, 199?). Já para a definição do local PARKA e colaboradores (2006) propõem que leve-se em conta, além dos objetivos iniciais (representatividade, monitoramento fontes de poluição, verificação das mudanças de qualidade, etc), a facilidade de acesso ao local (figura 1).

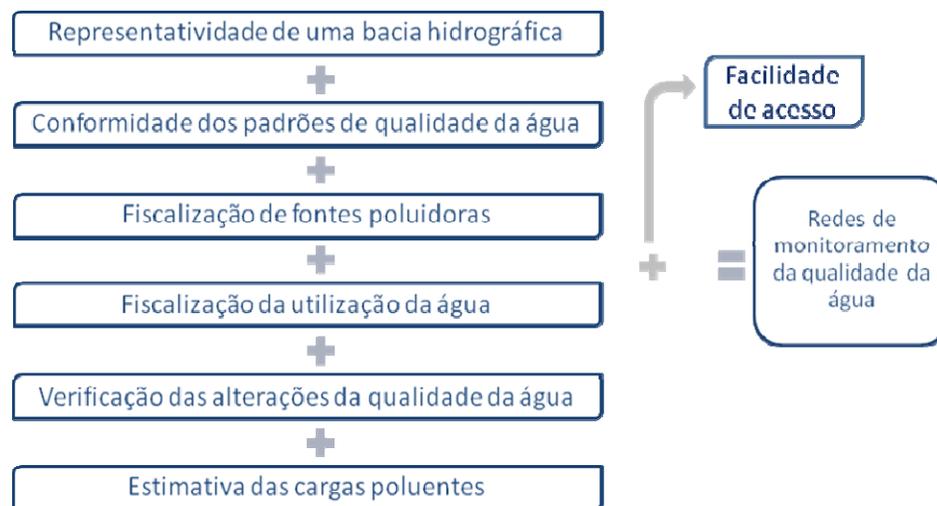


Figura 1- Aspectos para definição de uma rede de monitoramento

Fonte: Adaptado de Parka *et al.* (2006)

3.4 O Monitoramento Como Ferramenta de Gestão

Os monitoramentos são planejados de forma regional, mas segundo Barbosa (199?), seus produtos devem ter linguagem simples para atingir o maior número de pessoas possível e visar técnicas de pesquisa unificadas. Se bem planejados, os monitoramentos contribuem para uma gestão efetiva. O planejamento busca minimizar problemas através de uma preparação para possíveis transtornos futuros. Já a gestão é efetivação das condições que o planejamento feito no passado ajudou a construir. Assim, os dois são complementares.

Para verificar e acompanhar as alterações no ecossistema aquático causadas pelo homem, o monitoramento limnológico surge como uma das mais utilizadas ferramentas de gestão ambiental. Com a devida frequência, o monitoramento pode fornecer um diagnóstico ambiental, o

qual é importante para subsidiar as tomadas de decisão que permitam evitar ou reverter prejuízos ecológicos. As pesquisas subsidiadas pelo monitoramento, se devidamente difundidas e adaptadas às peculiaridades de cada local, resultam em tecnologias e conhecimentos que podem ser aplicados em distintas áreas (JORGENSEN; BENDORICCHIO, 2001).

As informações obtidas podem compor um banco de dados que pode ser usado como referência para tomada de decisões futuras. As informações armazenadas no banco de dados devem sempre estar disponíveis para referências e utilização pública (BARBOSA, 1997).

3.5 Tratamento de Dados: Qualidade dos dados

Durante a compilação de dados em um banco, podem surgir diversos contratemplos. Essas adversidades devem ser tratadas com cautela; para que os resultados permaneçam confiáveis, não é aconselhada a alteração dos dados sem análise e critérios previamente elaborados. O desenvolvimento de métodos para o tratamento de dados deve sempre visar maior eficiência do uso dos resultados existentes, independente da qualidade. Para que as informações sejam confiáveis, elas devem ser validadas e acompanhadas de elementos que indiquem o nível de confiança.

Qualidade aplicada a dados tem várias definições, um bom conceito proposto por Chrisman (1986) que é comumente utilizado, é o “potencial de uso” ou “adequação ao uso”. Esses conceitos definem o que é uma informação que pode ser usada com confiança, mostrando que os dados sem qualidade não tem potencial nem adequação. Muitas vezes os dados são utilizados sem a avaliação crítica de que possam conter erros. Esse fato pode levar a resultados errôneos, dando informações enganosas, decisões ambientais falhas e custos maiores (CHAPMAN, 2005). A baixa qualidade nos dados pode levar a conseqüências desastrosas, como a inutilização de séries inteiras por falta de qualidade.

As fases, durante um monitoramento ambiental, que podem levar a dados errôneos são:

- Anotação dos dados no momento de coleta;
- Manipulação dos dados nos momentos de análises laboratoriais;
- Anotação dos dados em relatórios;

- Arquivamento dos dados;
- Uso dos dados para análises estatísticas.

Todos esses momentos têm um impacto grande na qualidade ou adequação dos dados. Segundo English (1999), em um banco de dados os valores registrados não têm qualidade ou valor, eles têm, na verdade, um valor potencial, que só pode ser percebido quando usam-se os dados para algo útil. A qualidade da informação é relacionada com a habilidade de satisfazer a necessidade do consumidor final, no caso das pesquisas, o leitor.

A visão que se tem de erros e incertezas é que eles são ruins. Segundo Chapman (2005) isso não é necessariamente correto pois, essa informação é útil para se verificar como os erros ocorrem e como eles podem ser manejados e reduzidos. O mesmo autor diz, ainda, que um bom entendimento dos erros e suas propagações levam ao controle de qualidade. Compartilhando o mesmo ponto de vista, Oliveira e colaboradores (2010) defendem que um objetivo fundamental na validação e limpeza de dados é a identificação das causas dos erros detectados e o desenvolvimento de normas que evitem, ou pelo menos mitiguem, a repetição desses erros.

Validação é um processo utilizado para determinar se os dados estão incorretos, imprecisos ou exorbitantes. Este processo pode ser feito através de uma verificação de formatos, verificação de integralidade, verificação de razoabilidade e limites, revisão para identificar outliers e uma análise por especialistas (CHAPMAN, 2005). Uma parte importante no processo de validação é a identificação das causas dos erros detectados e maneiras de prevenir que eles se repitam (REDMAN, 2001).

A limpeza dos dados é o processo de correção dos erros encontrados na validação. Durante esse processo deve-se ter cuidado para que nenhum dado seja perdido e as modificações sejam realizadas com muita atenção (CHAPMAN, 2005). É mais seguro manter todos os dados salvos antes e depois das alterações, pois caso haja alguma falha durante esse processo, os dados ainda podem ser recuperados.

Redman (1996) reconhece que existem dois tipos de consistência em um banco de dados: Semântica e Estrutural. A consistência semântica é que o dado deve ser claro, sem ambigüidade e consistente. Já a consistência estrutural é em relação à estrutura e formato em que esses dados

são apresentados. Os dois tipos de consistência são importantes e possibilitam que os usuários desses dados possam saber quais testes foram empregados e como eles foram feitos. Essas informações são partes importantes na interpretação dos dados.

A detecção de *outliers* pode ser uma ferramenta muito útil na detecção de erros. É importante que esses *outliers* não sejam deletados sem uma boa análise crítica. Dados ambientais são conhecidos por apresentarem muitos dados que parecem *outliers*, mas são dados corretos (CHAPMAN, 2005). Apagar os dados sem uma boa análise pode levar a resultados incorretos e tomadas de decisões baseadas em premissas falsas. Os *outliers* ambientais muitas vezes são dados importantes e devem ser estudados, não ignorados.

3.6 Dados Censurados

Os dados que não são detectados ou dados com valor incerto por ser abaixo ou acima do limite de detecção do método analítico, são chamados de dados censurados. Quando os métodos de medição são considerados imprecisos para reportar um valor tão baixo, ele é então reportado como menor que o limite de detecção do método, por exemplo, “<1”. Esse tipo de dado complica a análise descritiva dos dados, testes de diferença entre grupos, correlações, regressões, entre outros. Os dados não detectados são também chamados de censurados a esquerda. Já os dados que se apresentam com limites de detecção variáveis são chamados de censurados múltiplos (HELSEL, 2006 & HELSEL e GANSER 2007). Segundo Cristofaro (2009) e Baccarelli e colaboradores (2005), existem quatro abordagens para o tratamento de dados censurados: substituição, métodos paramétricos, métodos robustos e métodos não-paramétricos. O método da substituição é a simples substituição do valor censurado por um abaixo dele, geralmente a metade. Os métodos paramétricos baseiam-se na distribuição próxima ao normal para estimar um valor e são muito sensíveis a outliers. Os métodos robustos, também conhecidos como imputação, consideram uma distribuição assimétrica para extrapolar um valor para o dado censurado. Por ultimo, o método não-paramétrico considera a ordem de classificação dos dados, possibilitando testes de hipóteses e não sofrendo interferências pela presença de outliers.

4 METODOLOGIA

4.1 Caracterização da Área

A UHE Nova Ponte é a primeira hidrelétrica de Minas Gerais a ter o processo de licenciamento ambiental concluído de acordo com os novos requisitos da legislação e a cumprir todo o Plano de Controle Ambiental que compreendeu ações relativas aos meios físico, biótico e sócio-cultural (ÁGUA E TERRA, 2006).

A Usina Hidrelétrica Nova Ponte, localizada no município de Nova Ponte, é um barramento do rio Araguari, um dos afluentes do rio Paranaíba. A UHE Nova Ponte iniciou sua operação no ano de 1994, com capacidade de geração de 510 MW e possuindo um reservatório com uma área de 443 km² de espelho d'água (ARAXÁ AMBIENTAL, 2010), sendo a maior usina do rio Araguari. As estações de verão e inverno são bem definidas na região, sendo úmidas e secas respectivamente. As chuvas são concentradas nos meses de outubro e março.

Com relação aos principais usos dos solos na bacia, estudados por Carrijo, (2003), na região predomina a agropecuária, principalmente a cultura anual de soja e milho. Segundo a autora, o grau de antropização é alto e existem poucas áreas representativas da vegetação natural dos cerrados. O uso de maquinários agrícolas é intenso e de modo geral, a estrutura fundiária é marcada pelas grandes propriedades bem capitalizadas, com práticas agrícolas caracterizadas pelo elevado consumo de insumos e defensivos.

4.2 Criação do Banco de Dados

Para realização deste trabalho os dados do monitoramento da qualidade da água foram obtidos através da Companhia Energética de Minas Gerais. O Sistema de Informação da Qualidade da Água dos Reservatórios da Cemig – SISÁGUA- foi acessado no endereço www2.cemig.com.br/sag. Os dados de laboratório (DL) foram salvos para a futura compilação desses em Excel. O relatório de dados de laboratório traz informações sobre os parâmetros físico-químicos analisados, os quais são o foco deste estudo.

Através dos relatórios salvos (743 arquivos) foram obtidos os resultados para cada um dos parâmetros estudados. Todos os dados foram organizados em uma planilha Excel de forma organizada e previamente estipulada. Obteve-se, dessa forma, o banco de dados utilizado para as análises estatísticas.

A matriz de dados foi definida da seguinte forma:

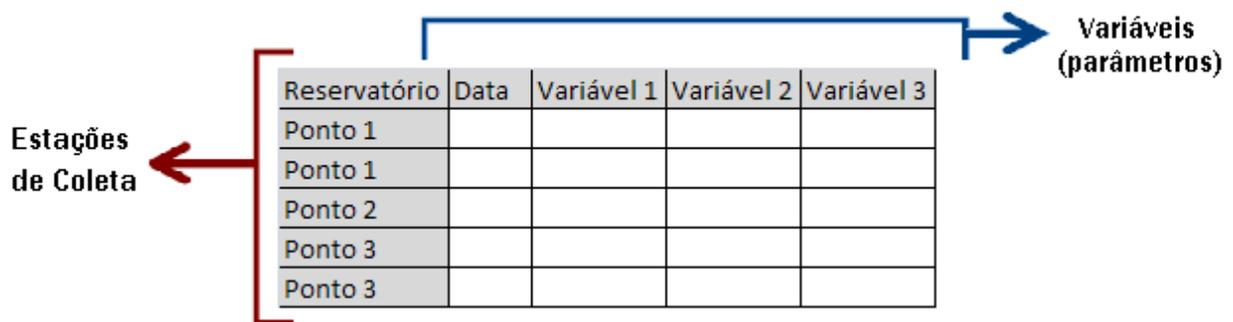


Figura 2 - Organização do Banco de Dados

O banco de dados criado tem dados desde 1986 até 2011. Possui no total 483 coletas distribuídas em 26 pontos de forma não uniforme. A descrição dos pontos pode ser vista na tabela a seguir:

Tabela 1 - Descrição dos Pontos de amostragem

Estação	Descrição	Condição física
BA171*	Braço do rio Quebra Anzol	Lótico
BA172*	Reservatório braço do rio Quebra Anzol	Lêntico
BA173*	Reservatório braço do rio Quebra Anzol	Lêntico
BA174*	Reservatório braço do rio Quebra Anzol	Lêntico
NP021	Rio Quebra Anzol na ponte da BR-146	Lótico
NP025	Rio Capivara na estrada relocada que conduz a Perdizes	Lótico
NP200 S	Reservatório, no braço do Rio Quebra Anzol à jusante do braço do rio Capivara	Lêntico
NP200 1/2ZF	Reservatório, no braço do Rio Quebra Anzol à jusante do braço do rio Capivara	Lêntico
NP200 F	Reservatório, no braço do Rio Quebra Anzol à jusante do braço do rio Capivara	Lêntico
NP180 S	Reservatório, no braço do ribeirão Santo Antônio.	Lêntico
NP180 1/2ZF	Reservatório, no braço do ribeirão Santo Antônio.	Lêntico
NP180 F	Reservatório, no braço do ribeirão Santo Antônio.	Lêntico
NP170 S	Reservatório, na ponte velha do rio Quebra Anzol, a jusante do braço do ribeirão Santo Antônio.	Lêntico

(continuação)

Estação	Descrição	Condição física
NP170 1/2ZF	Reservatório, na ponte velha do rio Quebra Anzol, a jusante do braço do ribeirão Santo Antônio.	Lêntico
NP170 F	Reservatório, na ponte velha do rio Quebra Anzol, a jusante do braço do ribeirão Santo Antônio.	Lêntico
BA160 1/2 ZF*	Braço do reservatório que recebe o ribeirão Santo Antônio e rio Quebra Anzol	Lêntico
BA160F*	Braço do reservatório que recebe o ribeirão Santo Antônio e rio Quebra Anzol	Lêntico
NP006	Rio Araguari na usina de Pai Joaquim	Lótico
NP150 1/2ZF	Reservatório, no braço do rio Araguari.	Lêntico
NP150 F	Reservatório, no braço do rio Araguari	Lêntico
NP140 1/2ZF	Ponto no corpo principal do reservatório, entre o barramento e o braço do rio Quebra Anzol.	Lêntico
NP140 F	Ponto no corpo principal do reservatório, entre o barramento e o braço do rio Quebra Anzol.	Lêntico
NP140 1/2Prof T	Ponto no corpo principal do reservatório, entre o barramento e o braço do rio Quebra Anzol.	Lêntico
NP140 S	Reservatório da UHE Nova Ponte no Corpo Principal entre o Barramento e o Braço do Rio Quebra Anzol, próximo as bóias de segurança.	Lêntico
NP110	Rio Araguari a montante do rio Claro, na mata de Jacob.	Lótico
NP101	Jusante da UHE Nova Ponte - na rampa que desce barcos - passando atrás da subestação	Lótico

*estações desativadas

4.3 Definição de Parâmetros trabalhados

O banco de dados possui 73 parâmetros, que são analisados com frequência indefinida. Para essa análise preliminar dos dados foram selecionados onze parâmetros. Estes parâmetros foram definidos de acordo com suas metodologias de análise, buscando os mais representativos, também pelo número de dados mais completos. O objetivo foi garantir uma análise mais completa e representativa.

Os parâmetros selecionados foram: pH, condutividade elétrica, turbidez, oxigênio dissolvido, nitrogênio amoniacal, fósforo total, demanda química de oxigênio, sólidos em suspensão, sólidos dissolvidos, ferro solúvel e manganês solúvel

4.4. Definição dos Pontos Trabalhados

Para a análise realizada nesse trabalho foram selecionados os pontos que possuíam pelo menos uma coleta. Seis pontos foram desconsiderados, pois se encontram desativados pela equipe da Cemig. Totalizam, portanto, 20 pontos trabalhados.

4.5 Validação do Banco e Análise preliminar

Para a análise preliminar foi necessária a avaliação de alguns pontos que são comuns a qualquer validação de banco de dados, principalmente secundários. Como a série de dados é extensa e não apresentou nenhuma frequência sistematizada, foi necessário estabelecer alguns critérios que facilitam a análise dos dados e pretendem alcançar a formatação mais próxima do ideal possível. Segue algumas das etapas para validação do banco e análise preliminar.

4.5.1 Análise preliminar

Para a análise preliminar dos dados foram realizados gráficos do tipo Boxplot no software de cálculos estatísticos StatSoft, Inc. 2007 - STATISTICA (Data Analysis Software System), versão 8.

Também foi realizada as estatísticas preliminares com o auxílio do Excel (anexo 1). Para essa análise foram calculados: Média, desvio padrão, mediana, média geométrica, percentil 10% e percentil 90%.

4.5.2 Outliers

Para identificação de *outliers* foram utilizadas as fórmulas a seguir. Estas foram calculadas para cada parâmetro e para cada estação de coleta.

Outlier superior : $> \{ \text{Quartil superior} + 1,5 (\text{Quartil superior} - \text{Quartil inferior}) \}$

Outlier inferior: $< \{ \text{Quartil inferior} - 1,5 (\text{Quartil superior} - \text{Quartil inferior}) \}$

Os outliers foram realçados do banco de dados, porém para determinar a exclusão ou não desses dados do banco (considerados erros), é exigida uma análise por especialistas da área. Para esse

trabalho os dados não foram excluídos, o que pode justificar possíveis alterações dos valores encontrados nas análises preliminares.

4.5.3. Dados faltantes

A análise de dados faltantes foi realizada a partir de um número ideal de dados. O ideal seria que todas as coletas analisadas (477) apresentassem resultado para todos os 11 parâmetros abordados. Dessa forma, teriam 5247 dados no total ($477 \times 11 = 5247$). A partir desse valor ideal, foi calculada a porcentagem de dados faltantes.

4.5.4 Dados censurados

De acordo com Cristofaro (2009), o método mais comumente utilizado para tratamento da censura consiste na simples substituição dos valores não detectados por um valor constante abaixo do limite de detecção. Com base nesse conceito foi definido que, por ponto, os parâmetros que tinham abaixo de 25% de dados censurados seriam substituídos pela metade do limite de detecção. Os que tinham acima de 25% não foram alterados, levando-se em conta o valor do limite de detecção. É necessário, ainda, a melhor definição do método para tratar os dados acima de 25%.

5 RESULTADOS E DISCUSSÃO

5.1 A Construção do Banco

A construção do banco de dados teve início em julho de 2011, utilizando os dados disponibilizados pela Cemig. Durante a compilação dos dados, em planilhas Excel, foram encontradas inúmeras falhas. Como a análise estatística dos dados depende de um banco consistente e completo, algumas das análises apresentadas podem não representar a situação real do local.

O banco de dados apresentou muitas falhas, já que as coletas não apresentaram uma frequência definida. Como as coletas iniciaram em 1986 e foram coletados os dados de monitoramento até junho de 2011, o ideal seria obter dados de 100 coletas, distribuídas em 20 pontos. Isso, assumindo a situação ideal das coletas trimestrais, como esperado. Então, somando-se todas as coletas de todos os pontos seriam 2000 coletas (tabela2).

Tabela 2 - Cálculos do total de coletas

Nº de anos	Nº de coletas por ano	Total de coletas (por ponto)
25	4	100
Nº de pontos	Nº de coletas	Total de coletas (em todos os pontos)
20	100	2000

O número de dados obtidos foi imensamente inferior ao esperado, totalizando 422 coletas em 17 anos e não 25 anos. Esse fato pode ser devido ao período de enchimento do reservatório que compreendeu os anos de 1988 a 1994, durante o qual não houve monitoramento. Além do número de coletas estar abaixo do esperado, também houveram problemas com as frequências. O ideal era ter quatro coletas por ano, por ponto, somando 100 coletas (por ano) em 20 pontos. O número de coletas verificadas por ano oscilou de 64 a 2 por ano, conforme pode ser observado na figura 3. É importante verificar que em alguns anos (1988 a 1995, 1997 e 2004) não houve nenhuma coleta realizada ou armazenada no banco de dados.

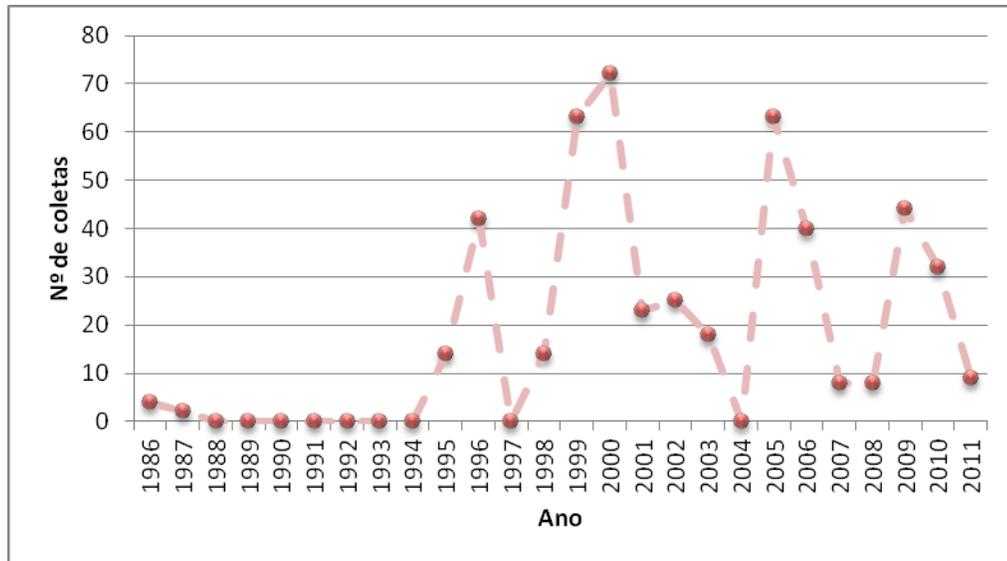


Figura 3- Gráfico de nº de coletas por ano

Outro problema encontrado na compilação do banco de dados foi o número escasso de coletas para alguns dos pontos. O baixo número de dados ou a inconstância deles ocasiona dificuldades em análises estatísticas que utilizam dados pareados ou até a impossibilidade de realização da análise. O número de coletas por ponto pode ser observado na figura 4.

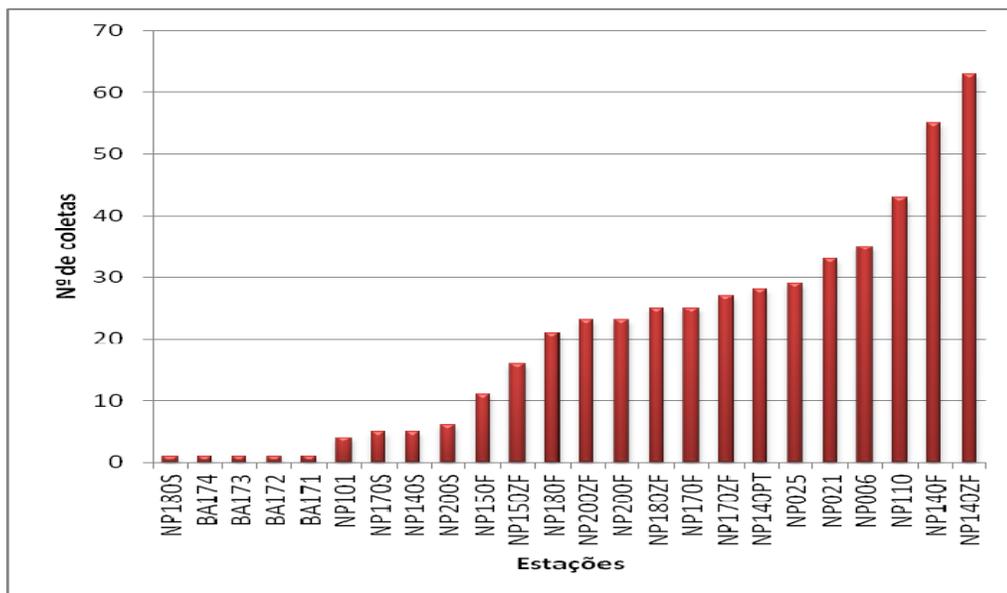


Figura 4- Número de coletas por estações

Após a estruturação final do banco de dados foi possível perceber, através de observação e bases estatísticas, a presença de *outliers*. A análise dos *outliers* para dados ambientais deve ser extremamente cuidadosa e fundamentada em conhecimentos técnicos. Alguns valores podem parecer discrepantes do grupo, mas representarem uma situação real. Um exemplo desse caso são os altos valores de turbidez, que aparentemente tratam-se de *outliers* (valores fora do padrão), mas que se analisados em conjunto com outros dados da coleta, representam uma situação característica da estação chuvosa. A tabela 2 mostra o número de outliers encontrados por parâmetros. Esses valores devem ser analisados e, após testes estatísticos, excluídos do banco de dados, se necessário.

Outro fator que pode alterar a análise dos dados é o número de dados censurados. Esses dados, que são computados como menor que o limite de detecção, podem indicar relações e padrões na verdade inexistentes. É importante definir uma forma de reduzir ao máximo o número de dados censurados. Dependendo do método utilizado no seu tratamento, os resultados podem sofrer alterações consideráveis, conduzindo a uma interpretação prejudicada (CRISTOFARO, 2009). Para este trabalho, foi utilizado o método da substituição. Os três tipos mais comuns de substituição são: substituição pelo valor do limite de detecção (LD), LD dividido por dois e LD dividido pela raiz quadrada de dois (HEWETT; GANSER, 2007).

Para cada um dos parâmetros selecionados foram calculadas as porcentagens de *outlier*, faltantes e censurados (tabela 3). De acordo com esses cálculos, foi possível perceber que os parâmetros com a maior porcentagem de faltantes são: Mn solúvel, Fe dissolvido e Sólidos em Suspensão. Os dados que possuem maior porcentagem de dados censurados são: fósforo total, DQO e Fe dissolvido. Já os parâmetros que apresentaram maiores porcentagens de *outliers* foram: sólidos suspensos, fósforo total e nitrogênio amoniacal.

Tabela 3- Porcentagens de faltantes, censurados e outliers por parâmetro

Parâmetro	Nº de dados	Dados faltantes (%)	Dados censurados (%)	Outliers (%)
pH <i>in loco</i>	476	0,2	0,0	4,8
Condutividade elétrica <i>in loco</i>	471	1,3	0,0	5,1
Sólidos dissolvidos totais - SDT	394	17,4	1,5	8,9
Turbidez	475	0,4	0,0	8,0
Sólidos suspensos totais - SST	357	25	0,0	9,5
Fósforo total	410	14,0	9,8	10,5
Nitrogênio amoniacal	417	12,6	6,0	9,6
Oxigênio dissolvido - OD	476	0,21	0,0	2,1
DQO	383	19,7	10,7	8,6
Manganês solúvel	235	50,7	4,3	8,9
Ferro dissolvido	294	38,4	8,2	7,8

5.2 Análise preliminar da qualidade da água

Quando represadas, as águas do rio podem sofrer uma série de alterações em sua qualidade. O maior impacto nesse ecossistema é a alteração da característica lótica para lântica do curso d água. A avaliação da qualidade da água, portanto, deve ser feita de forma separada. Já as reações e concentrações, são específicas para cada tipo de ambiente. A CONAMA 357/05 propõe diferentes valores para o mesmo parâmetro, de acordo com o tipo de ambiente (lótico, intermediário, ou lântico), em alguns casos.

A seguir, será realizada uma análise preliminar de 5 pontos localizados em zonas lânticas (divididos em coletas de superfície, metade da zona fótica e fundo) e 5 pontos coletados em zonas lóticas (sempre na superfície). A localização desses pontos pode ser vista na figura 5.

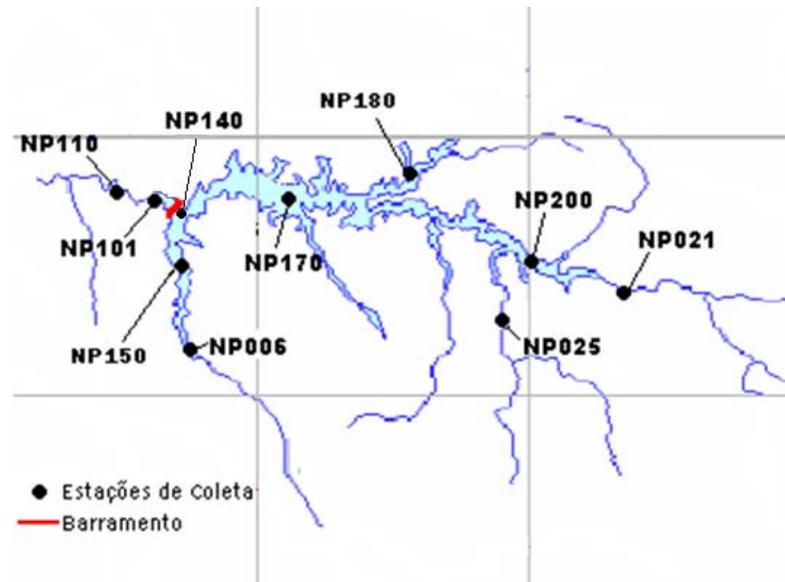


Figura 5- esquema da localização dos pontos no reservatório de Nova Ponte

5.2.1 Considerações iniciais

- Para a análise realizada nesse trabalho foi considerada a legislação CONAMA 357 de 2005. Os valores foram comparados com os padrões estabelecidos para águas de classe 2.
- Em alguns gráficos os eixos foram alterados para melhor visualização das variações, já que o banco apresenta valores de *outliers* em muitos parâmetros.
- Ambientes lânticos compreendem aos pontos localizados no corpo principal do reservatório. De acordo com a CONAMA 357/05, ambiente lântico é o ambiente que corresponde à água parada ou lenta.
- Ambientes lóticos correspondem aos pontos localizados nos rios tributários do reservatório. De acordo com a CONAMA 357/05, ambiente lótico é relativo a águas correntes.
- Os valores de ferro dissolvido e total apresentam-se acima do permitido na legislação em várias das coletas analisadas. O fato não foi pontualmente discutido, pois é conhecido que as águas de Minas Gerais apresentam valores elevados para esse parâmetro, uma vez que os solos da região contêm íons de Fe que podem ser solubilizados na água.

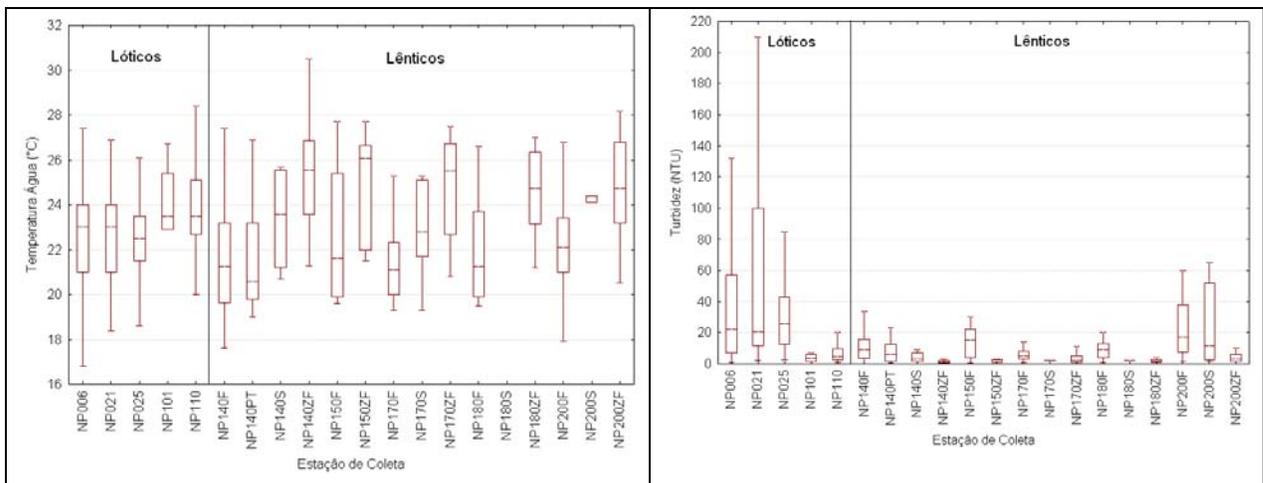
- 1/2ZF ou metade da zona fótica é a profundidade calculada a partir da transparência medida no local. Para o cálculo da zona fótica multiplica-se o valor obtido no disco de Secchi (transparência) por três e, em seguida, para obter a metade da zona fótica, divide-se por dois.
- Para a análise preliminar foram considerados 20 estações de coleta e 11 parâmetros, conforme descrito na metodologia. A configuração geral dos dados selecionados pode ser vista na tabela 4.

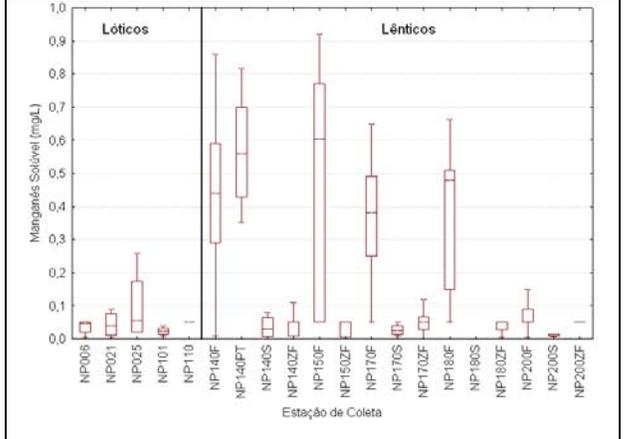
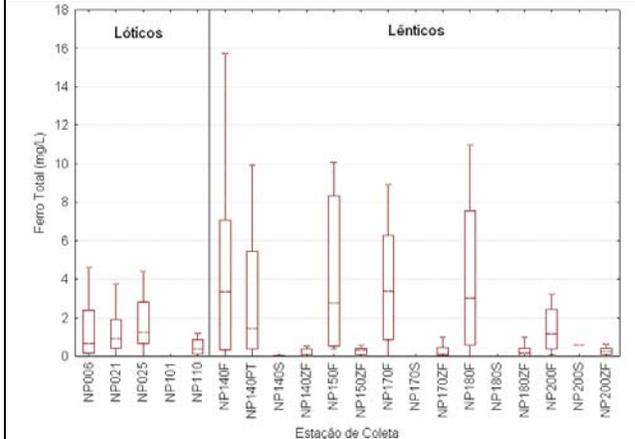
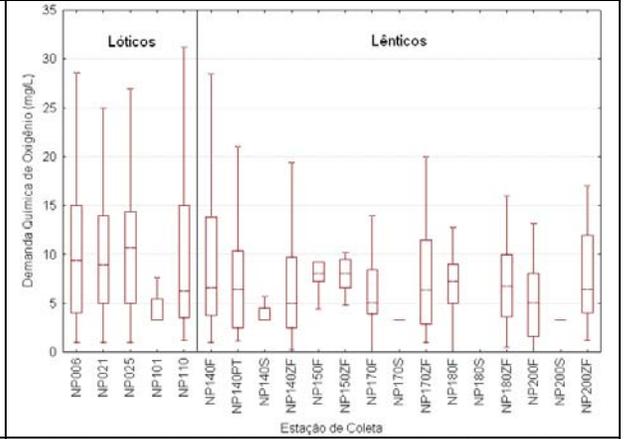
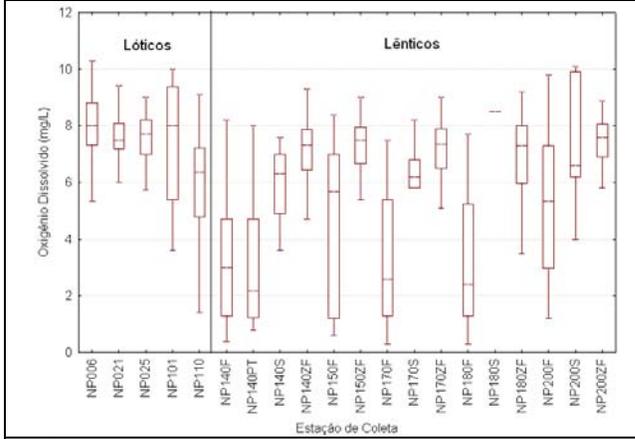
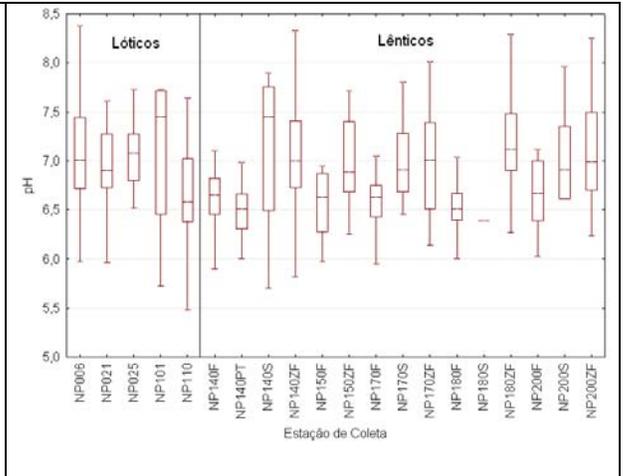
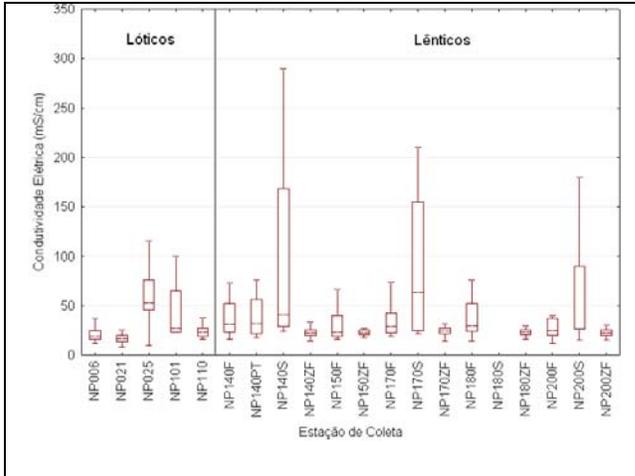
Tabela 4- Configuração geral dos dados selecionados

Configuração Geral	
Período de monitoramento	1986 a 2011
Nº de parâmetros	11
Frequência das análises	irregular
Nº de observações	4.388
Dados faltantes (%)	16
Dados censurados (%)	9
Outliers (%)	7

5.2.2 Gráficos Boxplot

Nos gráficos abaixo podem ser visualizadas as variações de doze parâmetros selecionados por estação de coleta.





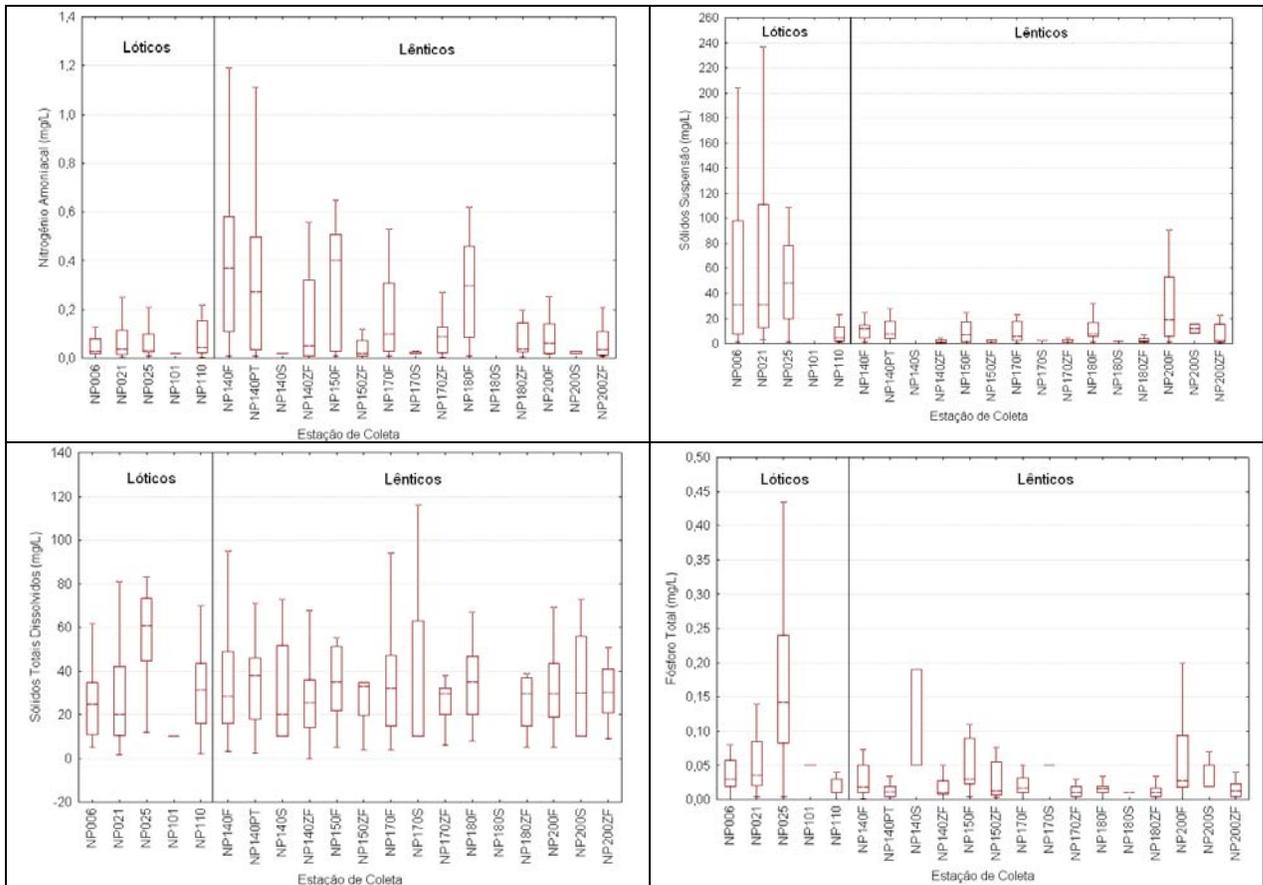


Figura 6 - Gráficos Boxplot para os parâmetros selecionados

Nos gráficos acima observa-se que a turbidez varia mais nos pontos de rio que no reservatório. As variações mais altas nos pontos lóticos podem ser explicadas em função do maior gradiente de mistura e também em função da menor profundidade, o que ocasiona um maior valor de sólidos em suspensão na água. O ponto NP200, que também apresentou variação na turbidez, é no final do corpo do reservatório, próximo às estações NP021 e NP025. A elevada variação e a proximidade dos pontos podem indicar uma fonte de poluição comum entre eles. Essa poluição ou alteração está afetando a qualidade da água do reservatório.

É possível ver no Boxplot do OD por estação de coleta que as variações estão dentro do esperado. Os pontos de profundidade apresentam os valores mais baixos e os pontos de superfície os valores mais altos. Alguns dos valores altos nos pontos profundos podem ser atribuídos a erros de digitação quando do cadastramento da coleta no sistema Siságua.

Os valores apresentados no gráfico para DQO mostram que há grandes variações. Percebe-se que os pontos lóticos apresentam variações maiores e valores mais altos. É possível que haja algum

despejo industrial próximo do local, o que justifica os picos da DQO, mas também é possível se tratar de causas naturais, já que não são valores extremos. No ponto NP110 uma possível alteração da qualidade pode ser justificada pela proximidade com a cidade de Nova Ponte, o que gera despejos em maior número.

Os valores de Fe dissolvido estão claramente mais altos que o padrão de classe 2 da legislação ($0,3 \text{ mgL}^{-1}$). Esses valores podem ser explicados pela composição do solo local, rico em compostos férricos que dissociados elevam os níveis de ferro nas águas.

Os valores de fósforo total apresentam uma alta variação nos pontos de rio. Os pontos NP200F e NP150F também apresentam elevadas variações. No ponto NP025 os valores estão muito altos, o que pode indicar uma fonte de poluição perto do local. Exemplos de fontes de poluição podem ser indústrias ou despejo de esgoto ou também lixiviação dos solos das plantações próximas que contêm fertilizantes compostos de fósforo.

Os valores de sólidos em suspensão foram mais altos para os pontos de rio, como mostram os gráficos na figura 6. Esse fato provoca a alta turbidez, como discutido anteriormente. É esperado que pontos lóticos tenham níveis mais altos de turbidez, já que estão submetidos à menor profundidade e menor movimentação da água. A variação de turbidez no ponto NP200 pode ser explicada, possivelmente, pelo fato de ser o ponto mais a montante do corpo do reservatório, o que justificaria a presença de características intermediárias entre lótico e lêntico.

Os valores de sólidos dissolvidos não seguem a tendência dos sólidos suspensos. Como por ser visto no gráfico acima, os sólidos dissolvidos apresentam menor variação e uma tendência menos perceptível. O ponto NP025 apresenta a média do parâmetro mais elevada, provavelmente por causa dos altos índices de fósforo.

6 CONCLUSÕES

A criação do banco de dados se mostrou uma das tarefas mais importantes para a análise e tratamento de dados. Tanto dados secundários como os dados primários, devem ser organizados e tratados de forma minuciosa para se evitar futuros problemas. No caso de banco de dados secundários, é necessário maior atenção ao tratamento de dados para que ele seja feito de forma imparcial, mas também de forma a encontrar todos os erros possíveis, evitando-os no futuro.

O tratamento dos dados muitas vezes não é realizado da devida forma, e em muitos trabalhos ele não sequer mencionado. Esse trabalho buscou esclarecer a importância dessa etapa e, principalmente, as alterações que ela pode causar aos resultados. Um tratamento feito de forma errada pode levar a conclusões e até a decisões baseadas em análises irreais.

A respeito da qualidade da água, foi possível perceber que os pontos de monitoramento lótico e lênticos mostram algumas diferenças, devendo ser tratados de maneiras específicas e individuais. Maiores conclusões sobre a qualidade da água do reservatório não podem ser alcançadas sem tratamentos mais aprofundados que os desta análise preliminar. Análises estatísticas detalhadas são necessárias para alcançar uma avaliação histórica adequada da qualidade da água.

A conclusão principal desse trabalho foi, portanto, a importância de análises preliminares, para que se possa obter um tratamento adequado dos dados. Com essa análise preliminar foi possível detectar erros que afetariam qualquer tipo de análise posterior, como os dados *outliers* e censurados.

REFERÊNCIAS

- ÁGUA E TERRA PLANEJAMENTO AMBIENTAL LTDA. **Relatório Semestral de Acompanhamento de Condicionantes da UHE Nova Ponte do Período de Janeiro a Junho/2006**. Pato de Minas, 2006.193 p.(Relatório).
- ÁGUA E TERRA. **Monitoramento da Qualidade das Águas de Nova Ponte**. Patos de Minas, 2009. 129 p.
- ANDRADE, E. *et al.* Seleção dos Indicadores da Qualidade das Águas Superficiais pelo Emprego da Análise Multivariada. **Revista de Engenharia Agrícola**, v.27, n.3, p.683-690, 2007.
- ARAXÁ AMBIENTAL LTDA. **Relatório do Monitoramento da Qualidade das Águas – UHE Nova Ponte**. Araxá, 2010. 83p. (Relatório de Monitoramento)
- BARAN, E. e WARRY, F. Simple Data Analysis for Biologists. Cambodia: **WorldFish Center and the Fisheries Administration.**, 2008.
- BARBOSA, Francisco. **PROGRAMA BRASILEIRO PARA A CONSERVAÇÃO E MANEJO DE ÁGUAS INTERIORES: Síntese das Discussões**. Belo Horizonte, 199?. 12 p.
- BACCARELLI, Andrea *et al.* Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study. **Chemosphere**, Itália, v. 60, n. , p.898-906, 2 mar. 2005.
- BRANCO, C. *et al.* Impact of climate on the vertical water column structure of Lajes Reservoir (Brazil): A tropical reservoir case. **Lakes & Reservoirs: Research and Management**, v.14, p.175-191, 2009
- BU, H. *et al.* Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. **Ecotoxicology and Environmental Safety**, v.73, p.907-913, 2010.
- CARRIJO, B. Cartografia geomorfológica com base em níveis de dissecação do relevo no médio curso do Rio Araguari-MG. **Caminhos de Geografia**, Uberlândia, MG, v.4, n.10, p.41-59, set/2003
- CARLETON, Christian; DAHLGREN, Randy; TATE, Kenneth. A relational database for the monitoring and analysis of watershed hydrologic functions: I. Database design and pertinent queries. **Computers & Geosciences**, Califórnia, v. 31, n. , p.393-402, 2005.
- CHAPMAN, Arthur D.. **Principles of Data Quality**. Copenhagen: Global Biodiversity Information Facility, 2005.
- CHRISMAN, Nicholas R.. The Role of Quality Information in the Long-term Functioning of a Geographic Information System. **S.I.**, Wisconsin, EUA, n. , p.80-87, 1986.
- COMPANHIA ENERGÉTICA DE MINAS GERAIS. **Relatório Anual e de Sustentabilidade**. Belo Horizonte, 2010. 150 p.
- COMPANHIA ENERGÉTICA DE MINAS GERAIS. **Temporal de IQA**. Disponível em: <<http://www2.cemig.com.br/sag/Relatorios/IQATemporal.aspx>>. Acesso em: 16 nov. 2011.

- COMPANHIA ENERGÉTICA DE MINAS GERAIS (Belo Horizonte). Sistema de Informação da Qualidade da Água dos Reservatórios da Cemig: Siságua. Disponível em: <www2.cemig.com.br/sag>. Acesso em: 13 jul. 2011.
- COMPANHIA DE TECNOLOGIA DE SANEAMENTO AMBIENTAL (São Paulo). **Qualidade da Água e suas Variáveis**. Disponível em: <http://www.quimlab.com.br/publish/pub/literatura.htm>. Acesso em: 17 nov. 2011.
- CONSELHO NACIONAL DO MEIO AMBIENTE - CONAMA. 2005. **Resolução Conama n^o 357**. Disponível em: <www.mma.conama.gov.br/conama> Acesso em 31/08/2011.
- CHRISTOFARO, Cristiano. **Avaliação Probabilística de Risco Ecológico de Metais nas Águas Superficiais da Bacia do Rio das Velhas - MG**. 2009. 311 f. Tese (Doutorado) - Departamento de Desa, Ufmg, Belo Horizonte, 2009.
- ENGLISH, Larry P.. **Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits**. New York: John Wiley & Sons, Inc., 1999. 518 p.
- FELTRAN, R. *et al.* Prevalência, abundância, intensidade e amplitude de infecção de nematóides intestinais em *Leporinus friderici* (Bloch, 1794) e *L. obtusidens* (Valenciennes, 1836) (Pisces, Anostomidae), na represa de Nova Ponte (Perdizes, MG). **Revista de Zootecnia**, v.6, n.2, p.169-179, 2004.
- GOMES E SOUZA, M.; VON SPERLING, E. Uso do zooplâncton como indicador de qualidade da água- estudo de caso da bacia do rio Araguari-MG. In: **23^o Congresso Brasileiro de Engenharia Sanitária e Ambiental**. 2005, Campo Grande /MS.
- HELENA, B. *et al.* Temporal evolution of groundwater composition in an alluvial (Pisuerga river, Spain) by principal component analysis. **Water Research**, v.34, p.807–816, 2000.
- HELSEL, Dennis R.. Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it. **Chemosphere**, Lakewood, v. 65, n. , p.2434-2439, 5 jun. 2006.
- HELSEL, Dennis R.; GANSER, Gary H.. A Comparison of Several Methods for Analyzing Censored Data. **Occupational Hygiene**, West Virginia, v. 51, n. 7, p.611-632, 17 ago. 2007.
- HEWETT, Paul; GANSER, Gary H.. A Comparison of Several Methods for Analyzing Censored Data. **Occupational Hygiene**, West Virginia, v. 51, n. 7, p.611-632, 2007.
- INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS. **Monitoramento da Qualidade das Águas Superficiais no Estado de Minas Gerais**: Relatório trimestral. Belo Horizonte, 2011. 109 p.
- JAYAKUMAR, R. e SIRAZ, L. Factor analysis in hydro geochemistry of coastal aquifers – a preliminary study. **Environmental Geochemistry**, v.31, p.174–177, 1997.
- Jørgensen, S.E., Bendricchio, G., **Fundamentals of ecological modelling**. Elsevier, Amsterdam, 530 p. 2001.

- LAMPARELLI, Marta Condé. **Grau de Trofia em Corpos d'água do Estado de São Paulo: Avaliação dos métodos de monitoramento.** 2004. 238 f. Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2004.
- LEMOS, Wictor Edney Dajtenko. **MONITORAMENTO E GESTÃO DA QUALIDADE DA ÁGUA EM RESERVATÓRIOS INCORPORANDO PROCESSOS HIDRODINÂMICOS E CLIMÁTICOS DE REGIÕES TROPICAIS SEMIÁRIDAS.** 2011. 164 f. Dissertação (Mestrado) - Universidade Federal do Ceará, Fortaleza, 2011.
- LORANTOU, A.; THOMÉ, J.P. e GOFFART, A. Water quality assessment of a recently refilled reservoir: The case of Bütgenbach Reservoir, Belgium. **Lakes & Reservoirs: Research and Management**, v.12, p.261-274, 2007.
- MICROSOFT. Office Excel. Versão 12.0: Microsoft Corporation, 2007.
- MWAURA, F. The spatio-temporal characteristics of water transparency and temperature in shallow reservoirs in Kenya. **Lakes & Reservoirs: Research and Management**, v.8, p.259-268, 2003.
- NATALE, O.; LEIS, M. Biotic Ligand Model estimation of copper bioavailability and toxicity in the Yacyretá Reservoir on the Paraná River. **Lakes & Reservoirs: Research and Management**, v.13, p. 231–244, 2008
- NADIM, F. *et al.* Water quality characteristics of two reservoir lakes in eastern Connecticut, USA. **Lakes & Reservoirs: Research and Management**, v.12, p.187-202, 2007
- OLIVEIRA, João M. et al. **A Qualidade das Base de Dados como Fator Crucial em Estudos Ambientais: Condições de Referência e Tipologia com Base Piscícola para Rios Portugueses.** In: CONGRESSO DA ÁGUA - MARCAS D'ÁGUA, 10^o., 2010, Algarve. Artigo. : Associação Portuguesa Dos Recursos Hídricos, 2010. p. 1 - 13.
- PARKA, Su-young et al. Design of a water quality monitoring network in a large river system using the genetic algorithm. **Ecological Modelling**, Coréia, v. 199, n. , p.289-297, 20 jul. 2006.
- RANGEL-PERAZA, J. *et al.* Statistical assessment of water quality seasonality in large tropical reservoirs. **Lakes & Reservoirs: Research and Management**, v.14, p. 315-323, 2009.
- REDMAN, Thomas. **Data quality for the information age.**Michigan: Artech House, 1996. 303 p.
- REDMAN, Thomas. **Data Quality: The Field Guide.** Boston: Digital Press, 2001. 256 p.
- RÊGO, A. *et al.* Relação peso-comprimento para *Prochilodus lineatus* (Valenciennes, 1836) e *Leoporinus friderici* (Bloch, 1794) (Characiformes) no reservatório de Nova Ponte – EPDA de Galheiro, rio Araguari, MG. **Revista Brasileira de Zootecias**, v.10, n.1, p.13-21, 2008.
- SIMEONOVA, V., e EINAX, J.. Environmental modeling and interpretation of river water monitoring data. **Anal. Bioanal. Chem.** v.374, p.898–905, 2002
- SINGH, K. *et al.* Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. **Water Research**, v.38, p.3980–3992, 2004.

- STATSOFT, Inc. 2007. Statistica: data analysis software system. Versão 8.0
- SUNDARAY, S. *et al.* Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of the Mahanadi river–estuarine system (India)—a case study. *Environmental Geochemistry Health*, v.28, p.317–330, 2006
- VEGA, M. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, v.32, p.3581–3592, 1998.
- VICINI, L. **Análise multivariada da teoria à prática**. Santa Maria, RS, 2005. (Trabalho de conclusão de curso) Universidade Federal de Santa Maria. Apud: REGAZZI, A. J. **Análise multivariada**. Viçosa: Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas. Departamento de Informática, 2001. 166p. (Apostila de disciplina).
- Vigilância e Controle da Qualidade da água para consumo humano, **Ministério da Saúde**, Brasília, DF. Disponível em: <<http://bvsmms.saude.gov.br/bvs/publicacoes>>. Acesso em: 17 nov 2011.
- VON SPERLING, Marcos. **Princípios do tratamento biológico de águas residuárias**. 3 ed. v.1 Belo Horizonte: Editora Ufmg, 2005. 452 p
- VON SPERLING, Marcos. **Estudos e modelagem da qualidade da água de rios**: v. 7. Belo Horizonte: Editora Ufmg, 2007. 588 p.

Condutividade Elétrica *in loco*

Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	34	2.9	0.0	8.8	31.7	43.2	19.7	23.3	15.1	35.9
NP021	32	3.0	0.0	6.3	21.6	19.7	16.5	18.1	12.6	24.0
NP025	29	0.0	0.0	6.9	64.4	36.6	53.2	54.8	35.8	110.3
NP110	42	2.3	0.0	0.0	23.6	5.3	23.4	23.0	18.0	28.0
NP101	4	0.0	0.0	25.0	44.4	37.2	27.2	36.0	23.3	86.2
NP140PT	28	0.0	0.0	0.0	36.8	16.8	32.0	33.5	20.1	57.0
NP140ZF	63	0.0	0.0	3.2	24.3	9.3	22.8	23.4	18.0	30.8
NP140S	4	0.0	0.0	25.0	98.9	127.8	40.8	57.8	26.8	241.5
NP140F	55	0.0	0.0	1.8	38.6	20.8	31.5	34.1	19.5	66.7
NP150F	11	0.0	0.0	18.2	31.8	17.5	23.0	28.4	18.9	62.2
NP150ZF	16	0.0	0.0	0.0	22.7	2.8	22.5	22.5	18.6	26.1
NP170F	25	0.0	0.0	8.0	36.0	19.4	28.9	32.3	20.0	68.1
NP170S	4	20.0	0.0	0.0	89.9	87.6	63.8	59.7	23.7	177.0
NP170ZF	27	0.0	0.0	14.8	25.7	7.5	24.0	24.8	19.0	35.5
NP180F	21	0.0	0.0	0.0	37.1	18.4	28.0	33.2	21.1	59.6
NP180F	0	100.0	0.0	0.0						
NP180F	25	0.0	0.0	8.0	23.9	4.6	23.0	23.5	19.4	30.0
NP180F	23	0.0	0.0	4.3	29.1	15.8	25.0	26.3	16.1	39.9
NP180F	5	16.7	0.0	0.0	67.6	69.4	26.5	44.3	19.6	153.0
NP180F	23	0.0	0.0	4.3	23.5	5.1	22.3	23.0	19.0	30.0
471										

Turbidez

Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	35	0.0	0.0	11.4	46.3	67.8	22.0	19.7	4.0	86.4
NP021	33	0.0	0.0	12.1	93.4	152.3	20.5	34.0	8.4	267.6
NP025	29	0.0	0.0	6.9	61.7	166.6	26.0	25.6	9.9	78.4
NP110	43	0.0	0.0	2.3	6.3	4.8	4.8	4.8	1.9	13.9
NP101	4	0.0	0.0	0.0	4.0	2.7	3.6	3.3	1.7	6.8
NP140PT	28	0.0	0.0	0.0	7.5	6.8	6.3	4.3	0.9	16.8
NP140ZF	62	1.6	0.0	6.5	1.6	1.2	1.2		0.5	2.5
NP140S	4	0.0	0.0	0.0	4.2	3.6	3.2	3.2	1.5	4.1
NP140F	55	0.0	0.0	12.7	21.9	47.1	9.0		2.1	36.4
NP150F	11	0.0	0.0	0.0	13.7	10.0	15.0	8.3	1.0	25.5
NP150ZF	16	0.0	0.0	6.3	1.8	1.7	1.1	1.3	0.6	3.0
NP170F	25	0.0	0.0	0.0	6.0	4.0	5.0	4.7	1.9	12.0
NP170S	5	0.0	0.0	40.0	6.8	11.1	2.0	3.1	1.6	19.4
NP170ZF	26	3.7	0.0	0.0	3.4	3.3	2.0		0.6	9.2
NP180F	21	0.0	0.0	4.8	9.8	7.0	9.2	7.4	3.0	16.2
NP180S	1	0.0	0.0	0.0	2.0		2.0	2.0	2.0	
NP180ZF	25	0.0	0.0	20.0	2.4	2.4	1.3		0.5	6.7
NP200F	23	0.0	0.0	17.4	85.7	267.6	17.0	17.9	2.2	123.1
NP200S	6	0.0	0.0	0.0	24.1	27.3	11.9	11.0	2.0	44.1
NP200ZF	23	0.0	0.0	13.0	10.0	19.4	2.8	3.5	1.0	35.2
475										

Sólidos suspensos totais - SST

Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	30	14.3	0.0	10.0	92.7	195.5	30.8	26.6	3.2	216.7
NP021	27	18.2	0.0	14.8	102.3	149.3	31.0	39.5	9.5	323.9
NP025	25	13.8	0.0	8.0	64.9	70.5	48.2	39.4	12.1	107.8
NP110	27	37.2	0.0	3.7	11.8	20.6	5.0	6.0	1.5	20.8
NP101	0	100.0								
NP140PT	18	35.7	0.0	11.1	15.8	20.3	7.5		3.0	33.1
NP140ZF	36	42.9	0.0	16.7	7.0	23.5	1.4		0.3	10.6
NP140S	0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NP140F	29	47.3	0.0	17.2	37.3	80.3	12.0	11.0	1.7	112.4
NP150F	11	0.0	0.0	0.0	9.5	8.5	7.0		1.0	18.3
NP150ZF	16	0.0	0.0	6.3	2.2	2.4	1.5		0.8	3.0
NP170F	21	16.0	0.0	9.5	14.7	23.0	6.0	6.3	1.3	27.4
NP170S	1	80.0	0.0	0.0	2.5		2.5	2.5	2.5	
NP170ZF	25	7.4	0.0	16.0	4.4	9.0	1.5		0.5	9.4
NP180F	21	0.0	0.0	0.0	11.9	10.4	7.6	7.5	1.5	28.5
NP180S	1	0.0	0.0	0.0	2.1		2.1	2.1	2.1	
NP180ZF	25	0.0	0.0	4.0	3.3	4.4	1.6		0.8	6.0
NP200F	19	17.4	0.0	0.0	29.4	28.9	19.0	16.0	3.6	76.1
NP200S	2	66.7	0.0	0.0	12.0	5.2	12.0	11.5	9.1	8.3
NP200ZF	23	0.0	0.0	13.0	10.7	16.0	2.4	4.2	1.1	39.2
357										

Fósforo total										
Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	33	5.7	0.0	18.2	0.1	0.4	0.0		0.0	0.3
NP021	32	3.0	0.0	6.3	0.1	0.1	0.0	0.0	0.0	0.1
NP025	29	0.0	0.0	13.8	1.4	6.1	0.1	0.1	0.0	0.6
NP110	26	39.5	30.8	7.7	0.9	4.5	0.0		0.0	0.0
NP101	4	0.0	100.0	0.0	0.1	0.0	0.1	0.1	0.1	0.1
NP140PT	18	35.7	0.0	11.1	0.0	0.0	0.0		0.0	0.0
NP140ZF	45	28.6	26.7	11.1	0.3	1.5	0.0		0.0	0.1
NP140S	3	25.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1
NP140F	45	18.2	0.0	15.6	0.1	0.3	0.0	0.0	0.0	0.4
NP150F	11	0.0	0.0	9.1	1.1	3.6	0.0	0.1	0.0	1.3
NP150ZF	16	0.0	0.0	12.5	1.7	4.6	0.0	0.0	0.0	7.2
NP170F	25	0.0	32.0	8.0	0.0	0.0	0.0		0.0	0.1
NP170S	5	0.0	80.0	20.0	0.0	0.0	0.1	0.0	0.0	0.1
NP170ZF	25	7.4	0.0	8.0	0.6	2.8	0.0		0.0	0.0
NP180F	20	4.8	0.0	10.0	0.0	0.0	0.0		0.0	0.0
NP180S	1	0.0	0.0	0.0	0.0		0.0	0.0	0.0	
NP180ZF	24	4.0	0.0	12.5	0.4	2.0	0.0		0.0	0.0
NP200F	22	4.3	0.0	4.5	0.1	0.1	0.0		0.0	0.2
NP200S	6	0.0	66.7	0.0	0.0	0.0	0.1	0.0	0.0	0.1
NP200ZF	20	13.0	0.0	5.0	0.0	0.0	0.0		0.0	0.0
410										

Mn Solúvel										
Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	4	88.6	0.0	0.0	0.04	0.02	0.05	0.03	0.01	0.05
NP021	4	87.9	0.0	0.0	0.04	0.04	0.04	0.03	0.01	0.08
NP025	4	86.2	0.0	0.0	0.10	0.11	0.06	0.06	0.02	0.21
NP110	1	97.7	0.0	0.0	0.1		0.1	0.1	0.1	0.1
NP101	4	0.0	0.0	0.0	0.02	0.01	0.02	0.02	0.01	0.04
NP140PT	11	60.7	0.0	9.1	0.53	0.23	0.56	0.41	0.35	0.80
NP140ZF	25	60.3	28.0	0.0	0.04	0.03	0.05		0.00	0.06
NP140S	4	0.0	0.0	0.0	0.04	0.04	0.03	0.02	0.01	0.07
NP140F	27	50.9	0.0	0.0	0.43	0.23	0.44	0.31	0.12	0.72
NP150F	9	18.2	0.0	0.0	0.49	0.36	0.60	0.29	0.05	0.81
NP150ZF	14	12.5	0.0	0.0	0.03	0.02	0.05		0.00	0.05
NP170F	25	0.0	0.0	4.0	0.44	0.42	0.38	0.33	0.13	0.59
NP170S	4	20.0	25.0	0.0	0.03	0.02	0.03	0.02	0.01	0.04
NP170ZF	23	14.8	0.0	30.4	0.08	0.18	0.05		0.01	0.09
NP180F	17	19.0	0.0	0.0	0.37	0.22	0.48	0.28	0.07	0.62
NP180S	0	100.0								
NP180ZF	19	24.0	0.0	5.3	0.05	0.03	0.05	0.04	0.02	0.05
NP200F	18	21.7	0.0	22.2	0.12	0.16	0.05	0.06	0.02	0.36
NP200S	5	16.7	40.0	20.0	0.02	0.02	0.01	0.01	0.01	0.04
NP200ZF	17	26.1	0.0	35.3	0.08	0.15	0.05		0.00	0.07
235										

Fe dissolvido										
Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	7	80.0	0.0	0.0	0.3	0.2	0.3	0.2	0.1	0.5
NP021	6	81.8	0.0	16.7	0.6	0.7	0.4	0.3	0.0	1.4
NP025	5	82.8	0.0	20.0	1.6	1.7	0.9	1.0	0.4	3.3
NP110	1	97.7	100.0	0.0	2.0		2.0	2.0	2.0	2.0
NP101	4	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.2
NP140PT	17	39.3	0.0	11.8	1.5	1.9	0.9		0.0	4.3
NP140ZF	42	33.3	26.2	4.8	0.2	0.4	0.1		0.0	0.5
NP140S	4	0.0	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.1
NP140F	44	20.0	0.0	6.8	1.8	2.1	0.9		0.1	5.3
NP150F	9	18.2	0.0	22.2	117.7	351.1	0.4		0.0	317.9
NP150ZF	14	12.5	35.7	14.3	0.5	0.9	0.1		0.0	1.7
NP170F	24	4.0	0.0	12.5	1.0	1.5	0.3		0.0	3.0
NP170S	4	20.0	25.0	0.0	0.1	0.0	0.1	0.1	0.1	0.1
NP170ZF	24	11.1	0.0	8.3	0.2	0.3	0.0		0.0	0.4
NP180F	19	9.5	31.6	15.8	0.9	1.1	0.5		0.0	2.7
NP180S	0	100.0								
NP180ZF	24	4.0	0.0	4.2	0.2	0.2	0.0		0.0	0.4
NP200F	20	13.0	0.0	0.0	0.5	0.5	0.3		0.0	1.2
NP200S	5	16.7	0.0	0.0	0.2	0.2	0.1	0.1	0.1	0.4
NP200ZF	21	8.7	0.0	4.8	0.2	0.3	0.1		0.0	0.5
294										

 Oxigênio dissolvido - OD

Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	35	0.0	0.0	0.0	7.9	1.1	8.0	7.8	6.2	9.0
NP021	33	0.0	0.0	6.1	7.6	0.9	7.5	7.5	6.8	8.4
NP025	29	0.0	0.0	6.9	7.6	1.2	7.7	7.5	6.4	8.7
NP110	42	2.3		2.4	6.1	1.7	6.4	5.8	4.4	7.9
NP101	4	0.0	0.0	0.0	7.4	2.8	8.0	6.9	4.7	9.4
NP140PT	28	0.0	0.0	0.0	3.1	2.1	2.2	2.4	0.9	5.6
NP140ZF	63	0.0	0.0	3.2	7.1	1.2	7.3	7.0	5.4	8.3
NP140S	4	0.0	0.0	25.0	6.0	1.7	6.3	5.7	4.4	7.4
NP140F	55	0.0	0.0	0.0	3.3	2.1	3.0	2.5	0.8	6.3
NP150F	11	0.0	0.0	0.0	4.5	3.1	5.7	3.2	1.2	8.0
NP150ZF	16	0.0	0.0	0.0	7.3	0.9	7.5	7.2	6.3	8.4
NP170F	25	0.0	0.0	0.0	3.2	2.2	2.6	2.4	0.9	6.3
NP170S	5	0.0	0.0	0.0	6.6	1.0	6.2	6.5	5.8	6.6
NP170ZF	27	0.0	0.0	3.7	7.1	1.2	7.3	6.9	6.0	8.1
NP180F	21	0.0	0.0	0.0	3.2	2.3	2.4	2.3	0.8	6.0
NP180S	1	0.0	0.0	0.0	8.5		8.5	8.5	8.5	
NP180ZF	25	0.0	0.0	0.0	7.0	1.3	7.3	6.8	5.6	8.2
NP200F	23	0.0	0.0	0.0	5.4	2.3	5.3	4.8	2.3	7.9
NP200S	6	0.0	0.0	0.0	7.2	2.4	6.6	6.9	5.1	8.7
NP200ZF	23	0.0	0.0	4.3	7.5	1.0	7.6	7.4	6.2	8.5
476										

DQO

Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	27	22.9	29.6	3.7	13.6	18.9	9.4	8.2	3.3	25.2
NP021	26	21.2	34.6	11.5	15.6	26.1	8.9	8.6	3.3	28.0
NP025	26	10.3	26.9	15.4	25.2	43.6	10.7	10.3	3.3	30.0
NP110	21	51.2	0.0	4.8	10.1	9.6	6.3	6.7	2.0	22.0
NP101	4	0.0	75.0	25.0	4.4	2.2	3.3	4.1	3.3	6.8
NP140PT	17	39.3	0.0	0.0	8.0	6.6	6.5	5.7	2.0	19.3
NP140ZF	43	31.7	0.0	11.6	9.6	17.7	5.0	5.0	1.7	19.4
NP140S	4	0.0	0.0	25.0	3.9	1.2	3.3	3.8	3.3	3.3
NP140F	44	20.0	0.0	2.3	9.7	8.4	6.6	7.0	2.7	17.6
NP150F	9	18.2	0.0	33.3	13.4	17.3	8.0	8.4	3.8	28.0
NP150ZF	12	25.0	0.0	16.7	10.1	6.7	8.0	8.8	6.0	18.4
NP170F	25	0.0	28.0	12.0	7.6	6.3	5.1		3.0	15.7
NP170S	4	20.0	100.0	0.0	3.3	0.0	3.3	3.3	3.3	3.3
NP170ZF	25	7.4	0.0	8.0	21.4	63.7	6.3	6.8	2.2	19.2
NP180F	21	0.0	0.0	9.5	9.6	12.6	7.2		2.1	13.3
NP180S	0	100.0								
NP180ZF	25	0.0	0.0	8.0	13.4	31.2	6.7	6.3	2.2	15.6
NP200F	22	4.3	0.0	0.0	5.4	3.8	5.0		1.6	9.0
NP200S	5	16.7	60.0	20.0	2.8	1.0	3.3	2.6	1.9	3.3
NP200ZF	23	0.0	0.0	4.3	8.1	6.1	6.4	6.2	2.1	12.7
383										

Nitrogênio amoniacal total

Estação de monitoramento	Nº dados	% Faltantes	% Censurados	% Outiers	Média	Desvio padrão	Mediana	Med. Geométrica	Perc. 10%	Perc. 90%
NP006	34	2.9	29.4	14.7	2.4	13.2	0.0	0.0	0.0	0.2
NP021	32	3.0	0.0	15.6	0.4	1.4	0.0	0.1	0.0	0.4
NP025	29	0.0	0.0	17.2	2.6	13.5	0.0	0.1	0.0	0.4
NP110	28	34.9	0.0	14.3	4.0	20.6	0.0	0.1	0.0	0.4
NP101	4	0.0	75.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NP140PT	18	35.7	0.0	5.6	0.4	0.3	0.3	0.2	0.0	0.7
NP140ZF	47	25.4	0.0	4.3	0.8	3.2	0.1		0.0	0.4
NP140S	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NP140F	45	18.2	0.0	2.2	0.4	0.3	0.4	0.2	0.0	0.9
NP150F	11	0.0	0.0	9.1	1.7	4.7	0.4	0.2	0.0	2.2
NP150ZF	16	0.0	0.0	12.5	1.0	4.0	0.0		0.0	0.2
NP170F	25	0.0	0.0	4.0	0.2	0.2	0.1	0.1	0.0	0.5
NP170S	4	20.0	75.0	25.0	0.0	0.0	0.0	0.0	0.0	0.0
NP170ZF	25	7.4	0.0	12.0	4.1	16.8	0.1	0.1	0.0	0.5
NP180F	21	0.0	0.0	4.8	0.4	0.5	0.3	0.2	0.0	0.6
NP180S	0	100.0								
NP180ZF	24	4.0	0.0	12.5	1.2	3.5	0.0	0.1	0.0	1.2
NP200F	22	4.3	27.3	9.1	0.1	0.3	0.1	0.1	0.0	0.3
NP200S	5	16.7	60.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0
NP200ZF	23	0.0	0.0	8.7	0.6	2.5	0.0	0.1	0.0	0.2