

# REGRESSÃO E VALIDAÇÃO DO MODELO DE CORRELAÇÃO ENTRE CE E STD: UMA CONTRIBUIÇÃO AO ESTUDO DO AQUIFERO CABEÇAS. SUDESTE DA BACIA SEDIMENTAR DO PARNAÍBA – PI.

Adson Brito Monteiro<sup>1</sup>; João Alberto O. Diniz<sup>2</sup> & Francisco Lages Correia Filho<sup>3</sup>

## RESUMO

Este trabalho teve como objetivo principal o estabelecimento de um modelo matemático de regressão linear simples que descreva a relação entre a condutividade elétrica (CE) e os sólidos totais dissolvidos (STD) para as águas do aquífero Cabeças na região sudeste da Bacia Sedimentar do Parnaíba. A equação de regressão ajustada é dada por  $STD=0,69*CE-11,52$ . As Previsões para sólidos totais dissolvidos (STD) devem ser realizadas dentro dos limites da condutividade elétrica (CE) que é entre 31,4 e 989  $\mu\text{S}/\text{cm}$ . O Coeficiente de Correlação de Pearson (R) indica que ocorre uma correlação fortíssima e positiva entre as variáveis e o Coeficiente de Determinação ( $R^2$ ) mostra que aproximadamente 98 % da variação dos sólidos totais dissolvidos (variável resposta) são explicados pela variação da condutividade elétrica (variável regressora). O Diagnóstico de Normalidade, Homocedasticidade, Independência dos Resíduos e de Pontos Influentes validaram de forma satisfatória o modelo de regressão ajustado. Os resultados mostraram que o modelo é aceitável no nível de confiança de 95%.

**Palavras Chave** - Regressão Linear Simples, Aquífero Cabeças, Bacia Sedimentar do Parnaíba

## ABSTRACT

This work aimed to establish a mathematical model of linear regression describing the relationship between electrical conductivity (EC) and total dissolved solids (TDS) to the waters of the aquifer heads in the southeastern part of the Sedimentary Basin Parnaíba. The fitted regression equation is given by  $TDS = 0,69*EC-11, 52$ . The forecast for total dissolved solids (TDS) should be performed within the limits of the electrical conductivity (EC) which is between 31.4 and 989  $\mu\text{S}/\text{cm}$ . The Pearson correlation coefficient (R) indicates that there is a very strong and positive correlation between the variables and coefficient of determination ( $R^2$ ) indicates that approximately 98% of the variation of total dissolved solids (response variable) are explained by the variation of electrical conductivity (regression variable). Diagnosis of Normality, homoscedasticity, Independence Residuals and Influential Points satisfactorily validated the regression model adjusted. The results showed that the model is acceptable level of 95%.

<sup>1</sup>Geólogo/Mestre em Geociências – Companhia de Pesquisa de Recursos Minerais (CPRM) – Av. Sul, 2291 – Afogados, Recife – PE. Fone (81) 33161469. [adson.monteiro@cprm.gov.br](mailto:adson.monteiro@cprm.gov.br). <sup>2</sup>Geólogo/Mestre em Geociências – Companhia de Pesquisa de Recursos Minerais (CPRM) – Av. Sul, 2291 – Afogados, Recife – PE. Fone (81) 33161400. [joao.diniz@cprm.gov.br](mailto:joao.diniz@cprm.gov.br). <sup>3</sup>Geólogo/Especialista em Recursos Hídricos – Companhia de Pesquisa de Recursos Minerais (CPRM) - Rua Goiás, 312 – Frei Serafim – Teresina – PI – Fone: (86) 32224153. [francisco.lages@cprm.gov.br](mailto:francisco.lages@cprm.gov.br).

## 1. INTRODUÇÃO

A condutividade elétrica de uma água é a medida de sua capacidade de conduzir corrente elétrica sendo dependente do número e do tipo de espécies iônicas dispersas, função da concentração de íons dissolvidos presentes nas águas.

A relação empírica entre a *Condutividade Elétrica* (*C.E.*,  $\mu S/cm$ ) e a concentração de *Sólidos Totais Dissolvidos* (*STD*,  $mg/L$ ), mostrada na equação (1) tem sido proposta (Metcalf & Eddy, 1991; APHA, 1992) para, na prática, estimar um parâmetro em função do outro.

$$STD = 0,64 * CE \quad (1)$$

O estabelecimento desta relação, bem como a facilidade de medição dos valores da condutividade elétrica por meio de instrumentos portáteis utilizados no campo, tem promovido uma generalizada utilização dos mesmos para se estimar valores de sólidos totais dissolvidos, ainda que a equação acima nem sempre possa ser corretamente utilizada, uma vez que a condutividade elétrica pode ter outras origens além da salinidade. O funcionamento dos medidores portáteis de condutividade elétrica consiste na aplicação de energia (tensão), entre dois ou mais eletrodos. Íons carregados positivamente (por exemplo, sódio,  $Na^+$ ; cálcio,  $Ca^{++}$ ; magnésio,  $Mg^{++}$ ; hidrogênio,  $H^+$ , etc.) irão se mover em direção ao eletrodo de carga negativa, e íons negativamente carregados (por exemplo, cloreto,  $Cl^-$ ; sulfato,  $SO_4^{-2}$ ; bicarbonato,  $HCO_3^-$ ; etc.) se moverão em direção ao eletrodo de carga positiva (Figura 1). Pelo fato desses íons estarem carregados e em movimento, eles são constituídos de corrente elétrica. O medidor monitora a quantidade de corrente que passa entre os eletrodos medindo quantos íons há na solução.

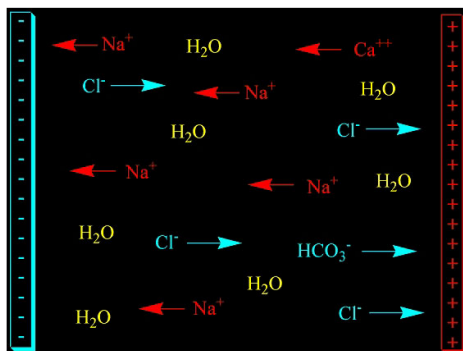


Figura 1. Dispositivo de Medição da Condutividade Elétrica.  
Fonte: AquaBase (2010)

Como pode ser visto na figura 1, os medidores de condutividade elétrica só detectam íons carregados e em movimento. Eles não irão detectar qualquer composto neutro (sem carga). Estes compostos incluem açúcar, álcool, produtos orgânicos e muitos outros (incluindo muitos pesticidas e seus resíduos), além de diversas formas de sílica, amônia e dióxido de carbono. Esses medidores também não detectam partículas macroscópicas, pois essas, devido ao seu grande tamanho não se movem em um campo elétrico. Assim, se existirem, por exemplo, partículas de óxido de ferro, as mesmas não serão medidas. Bactérias e vírus, também não serão detectados. Portanto, a

generalização da utilização de valores de condutividade elétrica transformados em sólidos totais dissolvidos através da equação (1) pode se constituir em um grande equívoco.

O que este trabalho propõe, é que, a partir do tratamento estatístico dos dados de condutividade elétrica e sólidos totais dissolvidos, coletados em águas de um mesmo aquífero, seja estabelecido um modelo de regressão linear simples que descreva com confiabilidade a relação entre estas variáveis, minimizando, desta forma os erros decorrentes da aplicação da equação 1.

## 2. LOCALIZAÇÃO E VIAS DE ACESSO

A área de estudo situa-se na região sudeste do Estado do Piauí e ocupa uma superfície de 6.063 km<sup>2</sup> (Figura 2). A cidade de Canto do Buriti está situada em sua porção noroeste e é um dos principais centros populacionais da microrregião. O acesso, a partir de Teresina, se faz pelas rodovias BR-316/BR-343 até a cidade de Floriano. Daí, pela rodovia estadual PI-140, pavimentada, até Canto do Buriti, num percurso total de aproximadamente 400 km, já dentro da área de estudo. Outro acesso, com pavimentação asfáltica, a partir da BR-343, em Regeneração, é através da rodovia PI-236, até Oeiras, seguindo adiante pela rodovia PI-143, até a cidade de Simplicio Mendes, alcançando São João do Piauí, no sudeste da área, através da BR-020, num trecho de 485 km, com início na capital, Teresina.

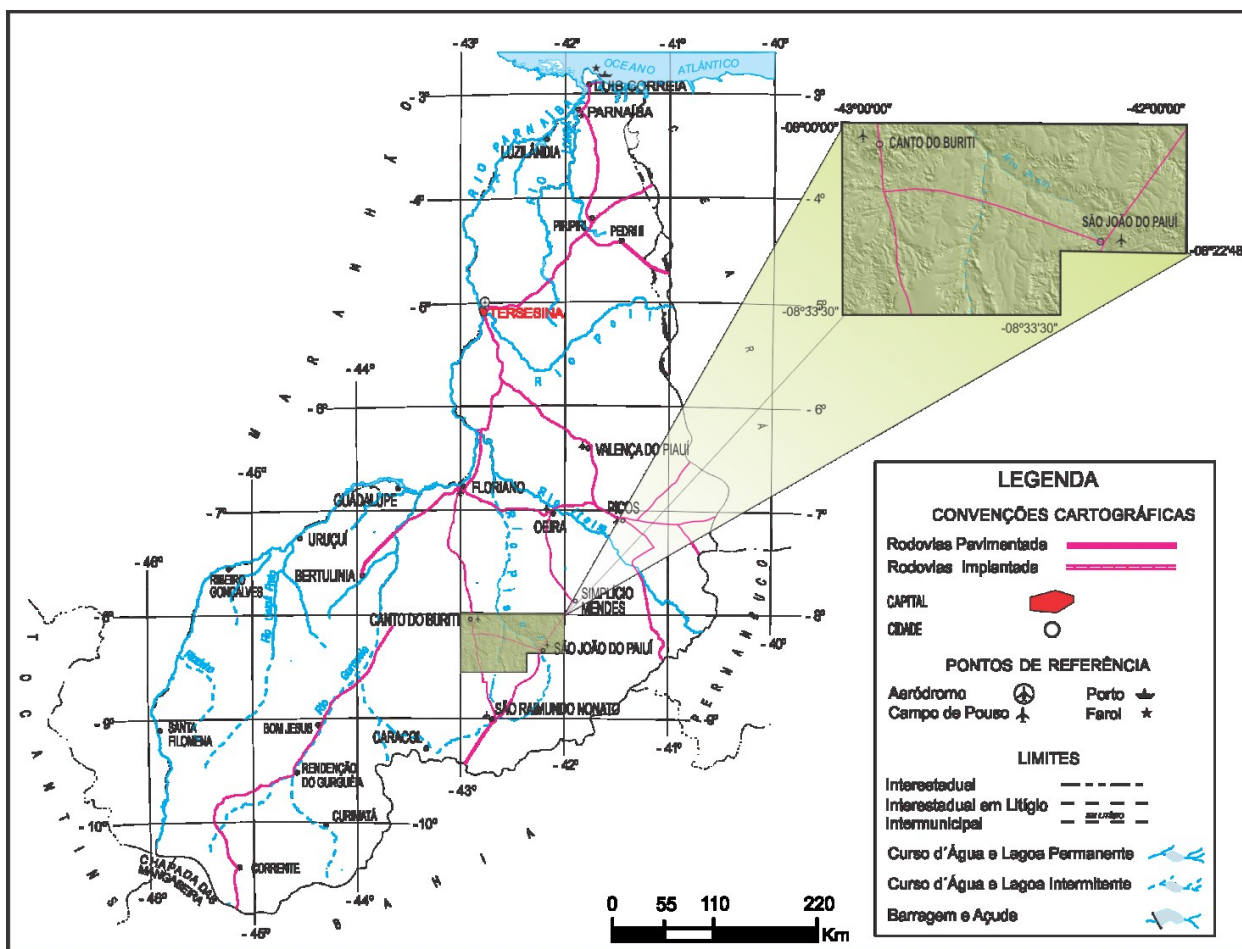


Figura 2. Localização e Vias de Acesso da Área.

### 3. OBJETIVOS

- Estabelecer o grau de relacionamento entre a condutividade elétrica (CE) e os sólidos totais dissolvidos (STD) para as águas do aquífero Cabeças na área em estudo através da determinação de uma equação matemática linear que descreva essa relação.
- Estimar valores de uma variável, com base em valores conhecidos da outra, dentro do intervalo estudado.
- Extrapolar valores projetados de uma variável de acordo com a tendência resultante da reta de regressão linear simples.
- Dotar a comunidade científica e os órgãos de cadastramento de poços de um instrumento de transformação da condutividade elétrica em sólidos totais dissolvidos no âmbito dos municípios abrangidos pelo estudo.

### 4. HIDROGEOLOGIA

A tabela 1 resume as principais características do aquífero Cabeças, para a área de estudo. A figura 3 mostra o mapa hidrogeológico. Não foram consideradas no mapa as coberturas detrítico-lateríticas por serem de pequena expressão hidrogeológica.

Tabela 1. Características do Aquífero Cabeças para Área de Estudo.

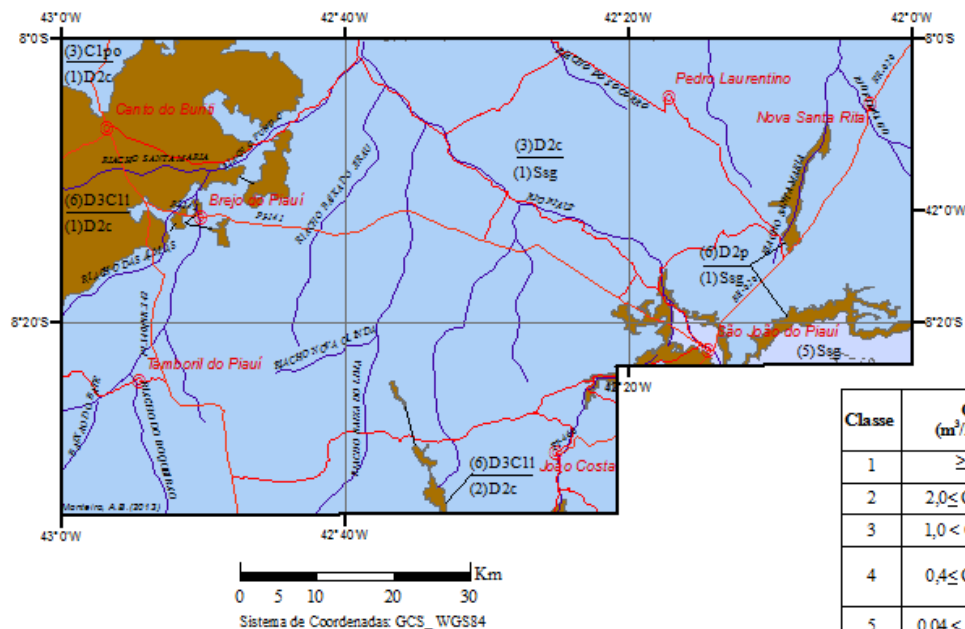
Litologia	Arenitos rosados, esbranquiçados e amarronzados, finos a médios, com frequente estratificação cruzada, micáceos, níveis de siltitos e folhelhos arroxeados, e presença de níveis conglomeráticos na porção média e inferior do pacote.
Comportamento Hidrodinâmico	Predominantemente livre, excetuando-se, numa reduzida área, a noroeste, onde está confinado pela Formação Longá.
Alimentação	Infiltração direta das águas das chuvas e, por filtração vertical ascendente, a partir do Aquífero Serra Grande.
Direção de Fluxo	Sul para Norte, com inflexão para Nordeste no lado sudoeste do vale do rio Piauí.
Porosidade Primária	Alta
Permeabilidade	Alta
Produtividade	Moderada (Classe 3) a Alta (Classe 2)
Espessura	120 metros
Capacidade Espécifica	1,77 m <sup>3</sup> /h/m
Transmissividade	6,39x10 <sup>-3</sup> m <sup>2</sup> /s
Condutividade Hidráulica	5,32x10 <sup>-5</sup> m/s
STD	161,8 mg/L

Fonte: Correia Filho & Monteiro (2010)

### 5. METODOLOGIA

#### 5.1 – Dados Utilizados e Base Estatística

Os valores de condutividade elétrica e de sólidos totais dissolvidos foram provenientes de 41 análises físico-químicas para o aquífero Cabeças, realizadas pelo Projeto Borda Sudeste da Bacia Sedimentar do Parnaíba (Correia Filho & Monteiro, 2010). A tabela 2 mostra os dados utilizados. A estatística teórica foi baseada no Portal ACTION, além dos autores citados no texto.



LEGENDA

Biblioteca das Unidades Hidrogeológicas

Nome	Símbolo	Litologia Predominante
Poti	Clpo	Arenito fino a médio
Longá	D3C11	Folhelho e Siltito
Cabeças	D2c	Arenito fino a médio
Pimenteiras	D2p	Folhelho e siltito
Serra Grande	Ssg	Arenito fino a grosseiro e conglomerados

Definição da Classe e Estimativa da Produtividade e dos Aquíferos (Struckmeir & Margat, 1995 - Modificado em Diniz, 2012).

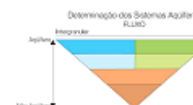
Classe	Q's (m³/h/m)†	T (m²/s)	K (m/s)	Vazão (m³/h)	Produtividade**
1	≥ 4,0	>10 <sup>-02</sup>	>10 <sup>-04</sup>	>100	Muito Alta
2	2,0 ≤ Q's < 4,0	10 <sup>-03</sup> ≤ T < 10 <sup>-02</sup>	10 <sup>-05</sup> ≤ K < 10 <sup>-04</sup>	50 ≤ Q < 100	Alta
3	1,0 < Q's < 2,0	10 <sup>-04</sup> < T < 10 <sup>-03</sup>	10 <sup>-06</sup> < K < 10 <sup>-05</sup>	25 < Q < 50	Moderada
4	0,4 ≤ Q's < 1,0	10 <sup>-05</sup> ≤ T < 10 <sup>-04</sup>	10 <sup>-07</sup> ≤ K < 10 <sup>-06</sup>	10 ≤ Q < 25	Geralmente baixa, porém localmente moderada
5	0,04 < Q's < 0,4	10 <sup>-06</sup> < T < 10 <sup>-05</sup>	10 <sup>-08</sup> < K < 10 <sup>-07</sup>	1 < Q < 10	Geralmente muito baixa, porém localmente baixa
6	< 0,04	< 10 <sup>-06</sup>	< 10 <sup>-08</sup>	< 1,0	Pouco Produtiva ou Não Aquifera

† Tempo de 6 horas e rebaixamento máximo de 25 metros para aquíferos granulares.  
 \*\* Na definição da produtividade para aquíferos Fraturados e Cársticos só deve ser considerado a vazão.

Produtividade das Unidades Hidrogeológicas Aflorantes

- Aquíferos Granulares**
- Moderada
  - Geralmente muito baixa, porém localmente baixa
  - Não Aquíferos
  - Pouco Produtiva ou Insignificante

Representação em Mapa das Classes dos Aquíferos  
 (Classe) Unidade Hidrogeológica Aflorante = (6)D3C11  
 (Classe) Aquífero mais produtivo e em menor profundidade (1)D2c



Convenções Cartográficas

⊙ Cidades — Estradas — Rios e Riachos

Figura 3. Mapa Hidrogeológico da Área de Estudo.

Tabela 2: Dados de CE e STD do Aquífero Cabeças.

Poço	Localidade	Município	Logitude	Latitude	CE	STD
CE026	Cotia	Pedro Laurentino	42° 13' 24.00"	8° 7' 32.00"	322,00	261,20
CE028	Estação	São João do Piauí	42° 10' 22.00"	8° 15' 9.00"	363,00	258,90
CE033	Angical	Brejo do Piauí	42° 40' 49.96"	8° 28' 6.10"	31,40	23,10
CE034	Alegre	João Costa	42° 35' 56.95"	8° 28' 33.09"	520,00	308,60
CE035	Recanto da serra	João Costa	42° 34' 18.98"	8° 28' 17.10"	166,00	86,00
CE103	Morros	Brejo do Piauí	42° 45' 12.95"	8° 10' 43.09"	86,00	45,00
CE107	Morro do império	Brejo do Piauí	42° 39' 43.98"	8° 13' 44.09"	65,00	34,00
CE221	Patos piripiri	Brejo do Piauí	42° 46' 6.96"	8° 7' 0.10"	58,00	45,80
CE223	Cansação	Brejo do Piauí	42° 41' 40.99"	8° 4' 34.08"	191,00	99,00
CE230	Trindade	Canto do Buriti	42° 52' 28.98"	8° 6' 44.09"	92,10	78,90
CE245	Tanque novo	Nova Santa Rita	42° 0' 41.97"	8° 10' 31.09"	381,00	313,10
CE247	Benguê	Nova Santa Rita	42° 2' 21.18"	8° 12' 36.09"	206,00	108,00
CE279	Umbuzeiro	Nova Santa Rita	42° 6' 35.98"	8° 8' 11.11"	989,00	663,60
CE280	Sobradinho	Nova Santa Rita	42° 6' 42.97"	8° 6' 53.11"	816,00	560,70
CE306	Maciel	Pedro Laurentino	42° 23' 5.97"	8° 7' 45.10"	86,00	45,00
CE329	Catirina	Pedro Laurentino	42° 16' 12.96"	8° 2' 10.11"	715,00	582,60
CE331	Marcos	Pedro Laurentino	42° 18' 36.94"	8° 3' 5.09"	185,20	102,90
CE340	Felipe	João Costa	42° 25' 0.97"	8° 31' 56.09"	832,00	592,10
CE342	Sede	João Costa	42° 25' 8.97"	8° 29' 15.12"	243,00	126,00
CE407	Saquarema	Nova Santa Rita	42° 6' 12.95"	8° 3' 8.09"	291,00	210,80
CE412	Palmeira	Nova Santa Rita	42° 4' 44.95"	8° 0' 41.11"	65,20	45,00
CE436	Comprida	Nova Santa Rita	42° 3' 39.97"	8° 4' 43.11"	570,00	296,00
CE445	Baixa da esquerda	Nova Santa Rita	42° 3' 37.96"	8° 8' 13.09"	594,00	412,80
CE655	Estreito i	São João do Piauí	42° 18' 58.01"	8° 15' 32.00"	391,00	274,60
CF093	Barreiro	São João do Piauí	42° 36' 23.96"	8° 17' 44.10"	118,00	61,00
GU036	Santa virgem	Pajeú do Piauí	42° 47' 47.98"	8° 4' 56.09"	66,00	34,00
GU052	Baixa grande	Pajeú do Piauí	42° 50' 20.95"	8° 4' 8.09"	56,50	29,50
GU148	São José	Canto do Buriti	42° 56' 45.98"	8° 10' 15.12"	130,40	103,90
GU172	Morro dos cavalos	Canto do Buriti	42° 55' 50.97"	8° 8' 10.10"	110,50	75,10
GU176	Várzea grande	Canto do Buriti	42° 53' 47.98"	8° 9' 6.09"	241,00	125,00
GU209	Assent. Três de abril	Canto do Buriti	42° 55' 58.99"	8° 3' 50.12"	49,30	42,80
GU220	Canto do angico	Pajeú do Piauí	42° 53' 13.98"	8° 0' 26.11"	176,00	92,00
GU293	Hotel rimo - sede	Canto do Buriti	42° 56' 53.96"	8° 5' 50.09"	84,00	44,00
GU314	Aroeira do policarpo	Canto do Buriti	42° 59' 35.96"	8° 4' 45.09"	76,30	65,10
GU543	Patos	Ribeira do Piauí	42° 29' 9.97"	8° 10' 8.11"	300,00	156,00
GU605	Caraibas	Tamboril do Piauí	42° 57' 3.98"	8° 25' 11.09"	112,80	67,00
GU616	Mandacaru	Tamboril do Piauí	42° 59' 53.02"	8° 28' 39.01"	456,00	287,50
GU617	Corrente	Tamboril do Piauí	42° 59' 38.97"	8° 23' 2.11"	261,00	183,60
GU618	Ponto d'água	Tamboril do Piauí	42° 53' 29.01"	8° 21' 35.99"	101,00	52,00
GU623	Lagoinha	Tamboril do Piauí	42° 56' 57.98"	8° 16' 40.10"	116,10	80,60
GU627	Morro vermelho	Tamboril do Piauí	42° 58' 2.97"	8° 21' 31.11"	43,80	37,10
GU629	Serra do alecrim	Tamboril do Piauí	42° 57' 50.99"	8° 18' 37.99"	457,00	395,10
GU633	Lagoa da entrada	Tamboril do Piauí	42° 54' 0.98"	8° 27' 45.09"	139,00	73,00
GU729	Caicara	Ribeira do Piauí	42° 31' 14.65"	8° 9' 2.12"	316,00	165,00
GU736	Várzea grande	Ribeira do Piauí	42° 37' 3.98"	8° 2' 8.10"	70,00	36,00
GU737	Povoado arvoredo	Ribeira do Piauí	42° 39' 5.96"	8° 1' 24.10"	95,60	36,20

## 5.2 – Base Cartográfica e Programas Utilizados

- Folha São João do Piauí (SC.23-X-B) – Escala 1:250.000;
- Imagem SRTM (Shuttle Radar Topography Mission) com processamento digital e recorte do mosaico estadual realizado pela EMBRAPA – Monitoramento por Satélite, compatível com a carta IBGE, SC.23-X-B, Escala 1:250.000;
- STATISTICA Trial Version 9 (Statsoft, Inc. 1984-2009);
- ACTION 2.3 – Software livre de estatística desenvolvido sob plataforma R. – 2012;

- ArcGis 10.0 –ESRI GIS and Mapping Software – 1999-2010.

### 5.3 – Regressão Linear Simples

#### 5.3.1 – Conceito

A regressão linear simples constitui uma tentativa de estabelecer uma equação matemática linear que descreva o relacionamento entre duas variáveis, sendo uma variável de entrada “x”, independente e uma variável de saída “y”, dependente. Devido em muitas situações, essa relação não ser perfeita, se faz necessário uma análise gráfica através do diagrama de dispersão. Esse gráfico dá indícios sobre o tipo de relação existente entre as variáveis, a variabilidade associada e pontos atípicos ou discrepantes, podendo-se calcular a equação de uma curva de tendência que melhor se ajuste à distribuição. Postulada que existe uma relação linear entre as variáveis x e y, podemos formular o seguinte modelo de regressão linear simples representando o conjunto de pontos pela equação da reta:

$$\hat{y} = a + bx_i + \varepsilon \quad (2)$$

onde  $\hat{y}$  - variável de resposta,  $x_i$  - variável regressora ou explicativa,  $a$  – coeficiente linear ou intercepção,  $b$  – coeficiente angular ou inclinação da linha de tendência e  $\varepsilon$  - toda variabilidade de  $\hat{y}$  não explicada por x.

Operacionalmente, neste caso, encontra-se a equação da reta para previsão de  $\hat{y}$ , considerando que os valores da variável aleatória  $\varepsilon$  são mínimos, tem média igual a zero e variância ( $\sigma^2$ ) constante, isto é, dado um valor de x, a variância de  $\varepsilon$  é sempre a mesma. Landim (2000) afirma que não é raro o termo  $\varepsilon$  ser numericamente mais importante do que a explicação motivada pela variável x, significando que outras variáveis devem ser incorporadas ao modelo, a fim de explicar o comportamento de y. Quando isto acontece, o método a ser utilizado é o da análise da regressão múltipla linear.

A reta de regressão é usada para resumir a estimativa linear entre duas variáveis aleatórias (Lapponi, 1997), e a análise simples desta regressão determina apenas a forma numérica de associação entre x e y, sem estabelecer nenhuma relação de causa.

#### 5.3.2 – Suposições do Modelo

- A relação matemática entre x e y é linear;
- Os valores de x são fixos;
- A média do erro é nula;
- Para um dado valor de x, a variância de  $\varepsilon$  é sempre constante;
- Os erros são não correlacionáveis e têm distribuição normal.

### 5.3.3 – Estimativa dos Parâmetros do Modelo

O primeiro passo na análise da regressão é estimar os parâmetros “a” e “b” do modelo. Os valores dessa estimativa são obtidas a partir de uma amostra de  $n$  pares de valores  $(x_i, y_i)$  que correspondam a  $n$  pontos no diagrama de dispersão. A técnica mais usada para ajustar uma linha reta a um conjunto de pontos é *Método de Mínimos Quadrados*. Esta metodologia consiste em adotar como estimativa dos parâmetros os valores que minimizem a soma dos quadrados dos desvios. Os valores de “a” e “b” da reta de regressão determinados são:

$$b = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \quad (3) \quad a = \bar{y} - b\bar{x} \quad (4)$$

Onde  $S_{xx}$ - somas dos quadrados dos desvios das médias;  $S_{xy}$  – soma dos produtos cruzados dos desvios de x e y;  $\bar{x}$  e  $\bar{y}$  – médias amostrais.

### 5.3.4 – Teste de Significância da Regressão

A dispersão da variação aleatória “y” pode ser medida através da soma dos quadrados dos desvios em relação a sua média  $\bar{y}$  (Lopes, 2003). Essa soma de quadrados é denominada *Soma dos Quadrados Total (SQT)* e pode ser dividida em duas partes: *Soma dos Quadrados da Regressão (SQR)* que é explicada pela regressão e *Soma dos Quadrados dos Erros (SQE)* que é a variação residual não explicada pela regressão. A Soma de Quadrados da Regressão (SQR) possui 1 grau de liberdade, enquanto a Soma dos Quadrados dos Erros ou Resíduos (SQE) possui  $n-2$  graus de liberdade. A divisão da soma de quadrados pelos respectivos graus de liberdade é o quadrado médio. Desta maneira, o quociente entre o *Quadrado Médio da Regressão (QMR)* e o *Quadrado Médio dos Erros (QME)* segue uma distribuição F de Snedecor com 1 e  $n-2$  graus de liberdade. Assim sendo, podemos avaliar a significância da regressão empregando o Teste F, através da análise de variância - ANOVA (Tabela 3).

Tabela 3. Análise de Significância usando ANOVA.

	<b>GL</b>	<b>SQ</b>	<b>QM</b>	<b>F<sub>calculado</sub></b>
<i>Regressão</i>	1	$SQR = bS_{xy}$	$QMR = \frac{SQR}{1}$	$F_0 = \frac{QMR}{QME}$
<i>Resíduo</i>	$n-2$	$SQE = S_{yy} - bS_{xy}$	$QME = \frac{SQE}{(n-2)}$	
<i>Total</i>	$n-1$	$S_{yy}$		

$$\text{Onde } S_{xy} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \quad (5) \quad e \quad S_{yy} = \sum_{i=1}^n y_i^2 - n\bar{y}^2 \quad (6)$$

Para testar a significância da regressão, ou seja, se realmente existe uma relação linear entre x e y, formula-se as seguintes hipóteses:  $H_0: \beta=0$  e  $H_1: \beta \neq 0$ , onde  $\beta$  representa o coeficiente de regressão paramétrico “b”. Se o valor de F, calculado através da tabela ANOVA, superar o valor teórico de F com 1 e  $(n-2)$  graus de liberdade, para o nível de significância  $\alpha$ , rejeita-se  $H_0$  e conclui-se que a regressão é significativa.



## 5.4 – Coeficiente de Correlação Linear

Conhecido como Coeficiente de Correlação de Pearson (R), determina o grau e o tipo de relacionamento entre as variáveis. O valor de R pode ser calculado por:

$$R = \frac{n(\sum_{i=1}^n xy) - (\sum_{i=1}^n x)(\sum_{i=1}^n y)}{\sqrt{n(\sum_{i=1}^n x^2) - (\sum_{i=1}^n x)^2} \sqrt{n(\sum_{i=1}^n y^2) - (\sum_{i=1}^n y)^2}} \quad (7)$$

A partir dos valores de R que varia entre -1 e 1; podemos verificar o tipo de correlação existente entre as variáveis estudadas de acordo com a tabela abaixo.

Tabela 4. Coeficiente de Correlação de Pearson.

<i>R(negativo)</i>	<i>Correlação</i>	<i>R(positivo)</i>
-1,0	Perfeita	1,0
-1 < R ≤ -0,75	Fortíssima	1 < R ≤ 0,75
-0,75 < R ≤ -0,50	Forte	0,75 < R ≤ 0,50
-0,50 < R ≤ -0,25	Média	0,50 < R ≤ 0,25
-0,25 < R < 0,0	Fraca	0,25 < R < 0,0

Quando R=0, a correlação é nula.

## 5.5 – Coeficiente de Determinação

O Coeficiente de Determinação ( $R^2$ ) indica a proporção da variável y explicada pela variável x através da função ajustada. O valor de  $R^2$  varia de 0 a 1. Quanto mais próximo de 1 significa que a variação explicada corresponde a uma grande percentagem da variação total. É calculado pela seguinte relação:

$$R^2 = \frac{(S_{xy})^2}{S_{xx}S_{yy}} \quad (8)$$

Onde,  $S_{xx}$  e  $S_{yy}$  – Soma dos quadrados dos desvios das médias;

$S_{xy}$  - Soma dos produtos cruzados dos desvios de x e y.

## 5.6 - Testes de Hipótese e Intervalos de Confiança para os Parâmetros do Modelo

### 5.6.1 – Intercepto (a)

Nesse modelo são testadas as hipóteses  $H_0: \beta_0=0$  e  $H_1: \beta_0 \neq 0$ , onde  $\beta_0$  representa o intercepto “a”. Para isso, temos que:

$$T = \frac{a}{\sqrt{QME(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})}} \quad (9)$$

Onde  $T$  – t de Student calculado;  $QME$  – Quadrado médio dos erros;  $n$  – número de amostras,  $\bar{x}$  é a média amostral e “a” - Intercepto.

Rejeita  $H_0$ , se o valor de |T| calculado pela equação (9) for maior que t de Student com  $(1 - \frac{\alpha}{2}; n-2)$ .

Então, o intervalo de confiança (95%) para o coeficiente linear ou intercepto é dado por:

$$[a - t_{(1-\frac{\alpha}{2})} \sqrt{QME(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})}; a + t_{(1-\frac{\alpha}{2})} \sqrt{QME(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})}] \quad (10)$$

### 5.6.2 - Coeficiente de Regressão (b)

Nesse modelo são testadas as hipóteses  $H_0:\beta=0$  e  $H_1:\beta\neq 0$ , onde  $\beta$  representa o coeficiente angular “b”. Para isso, temos que:

$$T = \frac{b}{\sqrt{\frac{QME}{S_{xx}}}} \quad (11)$$

Onde  $T - t$  de Student calculado;  $QME -$  Quadrado Médio dos Erros;  $n -$  número de amostras e  $S_{xx} -$  Soma dos quadrados dos desvios da média.

Rejeita  $H_0$ , se o valor de  $|T|$  calculado pela equação (11) for maior que  $t$  de Student com  $(1-\frac{\alpha}{2}; n-2)$ .

Então, o intervalo de confiança (95%) para o coeficiente angular é dado por:

$$\left[ b - t_{(1-\frac{\alpha}{2})} \sqrt{\frac{QME}{S_{xx}}}; \left[ b + t_{(1-\frac{\alpha}{2})} \sqrt{\frac{QME}{S_{xx}}} \right] \right] \quad (12)$$

### 5.6.3 – Intervalo de Confiança para Predição

O intervalo de confiança para a reta de regressão é dado por:

$$\left[ \hat{y} - t_{(1-\frac{\alpha}{2}, n-2)} \sqrt{QME \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}; \left[ \hat{y} + t_{(1-\frac{\alpha}{2}, n-2)} \sqrt{QME \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \right] \right] \quad (13)$$

## 6. DISGNÓSTICO DO MODELO

As suposições do modelo (item 5.3.2) precisam ser validadas para que os resultados sejam confiáveis. Desta forma, a análise dos resíduos é necessária. A equação (14) define os resíduos.

$$\hat{e} = y_i - \hat{y} \quad (14)$$

Onde  $\hat{e} =$  resíduo,  $y_i -$  valor da variável resposta e  $\hat{y} -$  valor ajustado.

As técnicas usadas para analisar os resíduos e validar o modelo foram as seguintes:

- **Papel de probabilidade normal** – Verificação da normalidade dos dados. É realizada através do gráfico “Resíduos Ordenados ( $\hat{e}_{(1)} \leq \hat{e}_{(2)} \leq \dots \leq \hat{e}_{(n)}$ ) versus Quantis da Distribuição Normal Padrão, calculado no ponto  $d_{(i)}$ ”. Se os pontos tendem a estar próximo da reta, é um indício que os erros são normalmente distribuídos.
- **Teste de Goldfeld Quandt** – Verificação da homocedasticidade do modelo, ou seja, variância constante dos erros em observações diferentes. Consiste em ordenar as observações de acordo com a variável explicativa. Depois, divide-se em três grupos ( $n_1, n_2, n_3$ ) de forma que  $n_2$  tenha 20% dos dados e  $n_1, n_3$  tenham quantidades iguais de dados. A estatística do teste é dada por:

$$F_{GQ} = \frac{\frac{SQE_{grupo3}}{n_3 - (p+1)}}{\frac{SQE_{grupo1}}{n_1 - (p+1)}} \quad (15)$$

Onde,  $SQE -$  Soma dos Quadrados dos Resíduos para o grupo 1 e o grupo 3;  $n -$  número de observações para os grupos 1 e 3;  $P -$  número de variáveis explicativas. Se  $F_{GQ} > P$ -valor do teste para o nível de significância  $\alpha=5\%$ , se aceita a hipótese de homocedasticidade dos dados.

- **Teste de Durbin-Watson** – Avaliação da hipótese de Independência de Dados. Consiste em mostrar que os resíduos de uma análise de regressão são independentes, entre si. Sendo  $e_i$  o resíduo associado à  $i$ -ésima observação, temos que a estatística do teste de Durbin-Watson é dada por :

$$d_w = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=2}^n e_i^2} \text{ em que } 0 \leq d_w < 4. \quad (16)$$

Toma-se a decisão sobre a independência dos resíduos comparando o valor de  $d_w$  com os valores críticos  $d_L$  e  $d_U$  da Tabela de Durbin-Watson. Assim, se:

- $0 \leq d_w < d_L \Rightarrow$  rejeitamos  $H_0$  (dependência);
- $d_L \leq d_w \leq d_U \Rightarrow$  o teste é inconclusivo;
- $d_U < d_w < 4 - d_U \Rightarrow$  não rejeitamos  $H_0$  (independência);
- $4 - d_U \leq d_w \leq 4 - d_L \Rightarrow$  o teste é inconclusivo;
- $4 - d_L < d_w \leq 4 \Rightarrow$  rejeitamos  $H_0$  (dependência).

- **Pontos Influentes** – Um ponto é influente se sua exclusão do ajuste da regressão causa uma mudança nos valores ajustados. Várias técnicas foram desenvolvidas para identificar observações influentes, entre estas, o *DFFITs*. O *DFFITs* mede a influência que a observação  $i$  tem sobre seu próprio valor ajustado, isto é, mede a diferença dos valores preditos de  $y_i$  com e sem a observação  $i$ , expressa em unidades de desvios padrões dos valores preditos de  $y_i$ . Se o valor do *DFFITs* for maior que 1, em módulo, existem pontos influentes. É definido por:

$$DFFITs = e_i \left( \sqrt{\frac{(N-p-1)}{SQE(1-h_{ii})-e_i^2}} \right) \left( \sqrt{\frac{h_{ii}}{(1-h_{ii})}} \right) \quad (17)$$

Onde  $e_i$  – resíduos;  $N$  –  $n^\circ$  de amostras;  $p$  – parâmetros ajustados; *SQE* – Soma do Quadrado dos Resíduos;  $h_{ii}$  – é o  $i$ -ésimo elemento da diagonal principal da matriz  $H$ , chamado de leverage.

## RESULTADOS E DISCUSSÃO

### 7.1 - Estimação dos Parâmetros do Modelo

As estimativas dos parâmetros  $a$  (*intercepto*) e  $b$  (*inclinação*) serão determinadas através do *Método dos Mínimos Quadrados*, aplicado ao conjunto de pontos considerados. A ideia é que, minimizando a soma dos quadrados dos resíduos, encontraremos “ $a$ ” e “ $b$ ” que trarão a menor diferença entre a previsão de “ $\hat{y}$ ” e o “ $y_i$ ”, realmente observado. Aplicando as fórmulas (3) e (4) aos pontos amostrados (Tabela 2) teremos os seguintes resultados:

$$b = 0,69 \text{ e } a = -11,52$$

Portanto a equação de regressão resultante é

$$\hat{y} = -11,52 + 0,69x, \text{ ou seja, } STD = 0,69 * CE - 11,52 \quad (18)$$

Graficamente mostrada na figura 4.

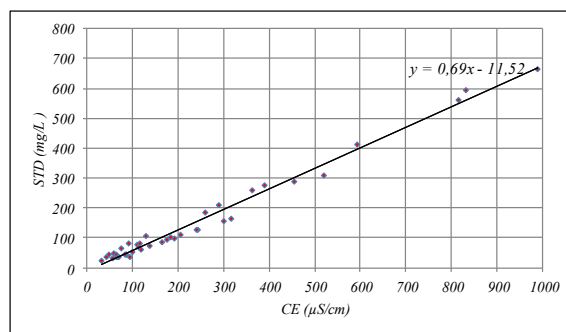


Figura 4. Equação e Reta de Regressão relacionando CE e STD.

Observa-se, graficamente que a correlação é alta e positiva, ou seja, os pontos se concentram em pequenos desvios com relação à reta que passa entre eles e os valores crescentes de CE (variável independente) estão associados a valores crescentes de STD (variável dependente).

Deve-se ter muito cuidado ao extrapolar valores fora do âmbito dos dados amostrais, uma vez que a mesma relação pode não ser verificada.

## 7.2 - Teste de Significância da Regressão Linear

A avaliação da significância da regressão pode ser feita empregando o Teste F de Snedecor através da análise de variância - ANOVA (Tabela 5) utilizando as fórmulas que constam na tabela 03 e que calculam a *Soma dos Quadrados da Regressão (SQR)*, *Soma dos Quadrados dos Erros (SQE)*, *Quadrado Médio da Regressão(QMR)*, *Quadrado Médio dos Erros(QME)* e *F de Snedecor calculado*.

Tabela 5. Análise de Significância da Regressão usando ANOVA.

	<b>GL</b>	<b>SQ</b>	<b>QM</b>	<b>F<sub>calculado</sub></b>
<i>Regressão</i>	1	<i>SQR=990760,30</i>	<i>QMR=990760,30</i>	2538,24
<i>Resíduo</i>	39	<i>SQE=15222,96</i>	<i>QME=390,33</i>	
<i>Total</i>	40	<i>SQT=1005983,26</i>		

Considerando o modelo de Regressão Linear Simples, a análise de regressão estabelece um teste para avaliar o parâmetro “ $\beta$ ”, isto é, testar as hipóteses  $H_0: \beta = 0$  e  $H_1: \beta \neq 0$ , onde  $\beta$  representa o coeficiente paramétrico “b”. Se  $F_{calculado} > F_{teórico}$  para o nível de significância  $\alpha$ , com 1 e (n-2) graus de liberdade, rejeita-se  $H_0$  e conclui-se que a regressão é significativa. Então, para o nível de significância  $\alpha=5\%$ , tem-se  $F_{calculado}=2538,24$  maior que  $F_{tabelado}=F_{(0,05;1;39)}= 4,091$ . Portanto, rejeita-se  $H_0$  com o nível de confiança de 95% e conclui-se que a variável explicativa tem correlação com a variável resposta.

## 7.3 - Coeficiente de Correlação de Pearson (R)

Aplicando a equação (7) aos dados da tabela 2, chega-se ao valor  $R \approx 0,99$  que, de acordo com a tabela 4, indica uma correlação fortíssima e positiva, entre as variáveis, condutividade elétrica (CE) e sólidos totais dissolvidos (STD).

#### 7.4 – Coeficiente de Determinação ( $R^2$ )

Objetivando definir se a reta de regressão se ajusta de maneira satisfatória ao conjunto de dados foi calculado o coeficiente de determinação  $R^2$ . Este coeficiente representa a proporção da variabilidade de  $y$ , explicada por  $x$ . Utilizando a equação (8) chega-se ao valor  $R^2 \approx 0,98$ . Este resultado indica que, aproximadamente 98 % da variação dos sólidos totais dissolvidos (variável resposta) são explicados pela variação da condutividade elétrica (variável regressora).

#### 7.5 – Valores Ajustados da Amostra

Valores ajustados dos dados da amostra são obtidos substituindo-se os correspondentes valores da variável preditora  $CE$  na função de regressão – Equação (18). A tabela 6 mostra os resultados para a variável resposta com 95% de confiança.

Tabela 6 – Valores Ajustados para STD.

CÓD.	CE	STD	STD Ajustado	CÓD.	CE	STD	STD Ajustado
CE028	363,00	258,90	238,95	GU036	66,00	34,00	34,02
CE033	31,40	23,10	10,15	GU052	56,50	29,50	27,47
CE034	520,00	308,60	347,28	GU148	130,40	103,90	78,46
CE035	166,00	86,00	103,02	GU172	110,50	75,10	64,73
CE103	86,00	45,00	47,82	GU176	241,00	125,00	154,77
CE107	65,00	34,00	33,33	GU209	49,30	42,80	22,50
CE221	58,00	45,80	28,50	GU220	176,00	92,00	109,92
CE223	191,00	99,00	120,27	GU293	84,00	44,00	46,44
CE230	92,10	78,90	52,03	GU314	76,30	65,10	41,13
CE247	206,00	108,00	130,62	GU543	300,00	156,00	195,48
CE279	989,00	663,60	670,89	GU605	112,80	67,00	66,31
CE280	816,00	560,70	551,52	GU616	456,00	287,50	303,12
CE306	86,00	45,00	47,82	GU617	261,00	183,60	168,57
CE331	185,20	102,90	116,27	GU618	101,00	52,00	58,17
CE340	832,00	592,10	562,56	GU623	116,10	80,60	68,59
CE342	243,00	126,00	156,15	GU627	43,80	37,10	18,70
CE407	291,00	210,80	189,27	GU633	139,00	73,00	84,39
CE412	65,20	45,00	33,47	GU729	316,00	165,00	206,52
CE445	594,00	412,80	398,34	GU736	70,00	36,00	36,78
CE655	391,00	274,60	258,27	GU737	95,60	36,20	54,44
CF093	118,00	61,00	69,90				

#### 7.6 – Testes de Hipóteses e Intervalo de Confiança para os Parâmetros do Modelo

- Para “a”, temos o seguinte teste de hipótese:  $H_0: a=0$  e  $H_1 \neq 0$ . A estatística do teste é dada pela equação (9). Então, considerando o nível de significância  $\alpha=5\%$ ,  $T$  calculado é igual a  $-2,63$  e  $t_{(0,975;39)}=2,02$ . Como o  $|T|_{calculado} > t_{tabelado}$ , rejeita-se  $H_0$  com o nível de confiança de 95%. O intervalo de confiança (95%) de “a”, utilizando-se as equações em (10) será  $[-20,39 ; -2,65]$ .
- Para “b”, temos o seguinte teste de hipótese:  $H_0: b=0$  e  $H_1 \neq 0$ . A estatística do teste é dada pela equação (11). Então, considerando o nível de significância  $\alpha=5\%$ ,  $T$  calculado é igual a  $50,69$  e

$t_{(0,975;39)}=2,02$ . Como o  $|T|_{calculado} > t_{tabelado}$ , rejeita-se  $H_0$  com o nível de confiança de 95%. O intervalo de confiança (95%) de “b”, utilizando-se as equações em (12) será  $[0,66 ; 0,72]$ .

### 7.7 - Intervalo de Confiança de Predição para a Equação de Regressão

Na tabela 7 são apresentados os valores preditos (valores ajustados) e os respectivos intervalos de confiança e desvios padrão para cada observação e a figura 5 mostra, graficamente.

### 7.8 – Análise dos Resíduos

A análise de resíduos consiste em um conjunto de técnicas para investigar a adequabilidade do modelo com base nos resíduos. Nesse sentido foi realizado diagnóstico de normalidade através do “*papel de probabilidade normal*”, diagnóstico de homocedasticidade pelo “*Teste de Goldfeld-Quandt*”, diagnóstico de independência com o “*Teste de Durbin-Watson*” e detecção de pontos atípicos através da identificação de observações influentes. Os resíduos são os erros ao se ajustar o modelo. A tabela 8 mostra o cálculo dos resíduos utilizando a equação (14) e a tabela 9, o cálculo dos quantis usando os procedimentos relatados no item 6.

Tabela 7 – Intervalo de Confiança para a Predição.

CE	STD Ajustado	Intervalo de confiança (95%)		Desvio Padrão	CE	STD Ajustado	Intervalo de Confiança (95%)		Desvio Padrão
		Limite Inferior	Limite Superior				Limite Inferior	Limite Superior	
363,00	238,95	231,70	246,20	3,58	66,00	34,02	26,34	41,70	3,80
31,40	10,15	1,87	18,42	4,09	56,50	27,47	19,62	35,31	3,88
520,00	347,28	337,13	357,43	5,02	130,40	78,46	71,65	85,26	3,36
166,00	103,02	96,54	109,50	3,20	110,50	64,73	57,68	71,77	3,48
86,00	47,82	40,44	55,20	3,65	241,00	154,77	148,52	161,02	3,09
65,00	33,33	25,63	41,03	3,81	49,30	22,50	14,54	30,46	3,94
58,00	28,50	20,68	36,32	3,87	176,00	109,92	103,51	116,33	3,17
191,00	120,27	113,94	126,60	3,13	84,00	46,44	39,03	53,85	3,66
92,10	52,03	44,74	59,32	3,60	76,30	41,13	33,60	48,65	3,72
206,00	130,62	124,35	136,89	3,10	300,00	195,48	188,94	202,02	3,23
989,00	670,89	649,06	692,72	10,79	112,80	66,31	59,30	73,32	3,47
816,00	551,52	534,20	568,84	8,56	456,00	303,12	294,29	311,95	4,37
86,00	47,82	40,44	55,20	3,65	261,00	168,57	162,27	174,87	3,12
185,20	116,27	109,91	122,62	3,14	101,00	58,17	51,00	65,34	3,54
832,00	562,56	544,83	580,29	8,77	116,10	68,59	61,62	75,56	3,45
243,00	156,15	149,90	162,40	3,09	43,80	18,70	10,65	26,76	3,98
291,00	189,27	182,80	195,74	3,20	139,00	84,39	77,68	91,10	3,32
65,20	33,47	25,77	41,17	3,81	316,00	206,52	199,84	213,20	3,30
594,00	398,34	386,52	410,16	5,85	70,00	36,78	29,16	44,40	3,77
391,00	258,27	250,60	265,94	3,79	95,60	54,44	47,20	61,68	3,58
118,00	69,90	62,95	76,85	3,44					

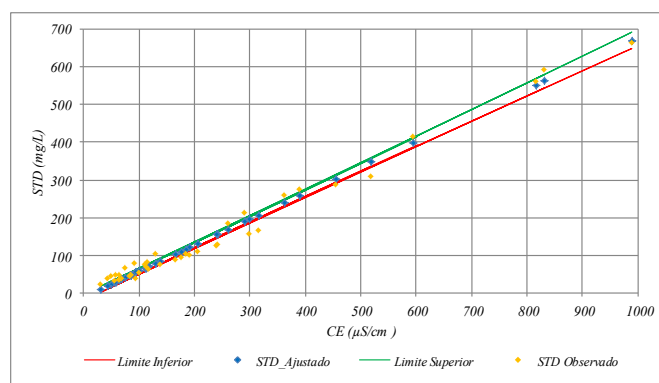


Figura 5 – Intervalo de Confiança para a Predição.

Tabela 8 – Cálculo dos Resíduos.

CÓD.	STD (y <sub>i</sub> )	STD Ajustado	Resíduo (ê <sub>i</sub> )	CÓD.	STD (y <sub>i</sub> )	STD Ajustado	Resíduo (ê <sub>i</sub> )
CE028	258,90	238,95	19,95	GU036	34,00	34,02	-0,02
CE033	23,10	10,15	12,95	GU052	29,50	27,47	2,04
CE034	308,60	347,28	-38,68	GU148	103,90	78,46	25,44
CE035	86,00	103,02	-17,02	GU172	75,10	64,73	10,38
CE103	45,00	47,82	-2,82	GU176	125,00	154,77	-29,77
CE107	34,00	33,33	0,67	GU209	42,80	22,50	20,30
CE221	45,80	28,50	17,30	GU220	92,00	109,92	-17,92
CE223	99,00	120,27	-21,27	GU293	44,00	46,44	-2,44
CE230	78,90	52,03	26,87	GU314	65,10	41,13	23,97
CE247	108,00	130,62	-22,62	GU543	156,00	195,48	-39,48
CE279	663,60	670,89	-7,29	GU605	67,00	66,31	0,69
CE280	560,70	551,52	9,18	GU616	287,50	303,12	-15,62
CE306	45,00	47,82	-2,82	GU617	183,60	168,57	15,03
CE331	102,90	116,27	-13,37	GU618	52,00	58,17	-6,17
CE340	592,10	562,56	29,54	GU623	80,60	68,59	12,01
CE342	126,00	156,15	-30,15	GU627	37,10	18,70	18,40
CE407	210,80	189,27	21,53	GU633	73,00	84,39	-11,39
CE412	45,00	33,47	11,53	GU729	165,00	206,52	-41,52
CE445	412,80	398,34	14,46	GU736	36,00	36,78	-0,78
CE655	274,60	258,27	16,33	GU737	36,20	54,44	-18,24
CF093	61,00	69,90	-8,90				

Tabela 9 – Cálculo dos Quantis.

Resíduo (ê)	Quantil $\Phi^{-1}(d_i)$	Resíduo (ê <sub>i</sub> )	Quantil $\Phi^{-1}(d_i)$	Resíduo (ê <sub>i</sub> )	Quantil $\Phi^{-1}(d_i)$	Resíduo (ê <sub>i</sub> )	Quantil $\Phi^{-1}(d_i)$
-41,52	-2,12	-13,37	-0,57	0,67	0,06	16,33	0,72
-39,48	-1,74	-11,39	-0,50	0,69	0,12	17,30	0,80
-38,68	-1,51	-8,9	-0,44	2,04	0,18	18,40	0,89
-30,15	-1,34	-7,29	-0,37	9,18	0,24	19,95	0,99
-29,77	-1,21	-6,17	-0,31	10,38	0,31	20,30	1,09
-22,62	-1,09	-2,82	-0,24	11,53	0,37	21,53	1,21
-21,27	-0,99	-2,82	-0,18	12,01	0,44	23,97	1,34
-18,24	-0,89	-2,44	-0,12	12,95	0,50	25,44	1,51
-17,92	-0,80	-0,78	-0,06	14,46	0,57	26,87	1,74
-17,02	-0,72	-0,02	0,00	15,03	0,65	29,54	2,12
-15,62	-0,65						

- Diagnóstico de Normalidade: Através do gráfico de probabilidade normal para os resíduos do modelo ajustado (Figura 6), observa-se que os pontos tendem a seguir o comportamento da reta, indicando que os erros são normalmente distribuídos.

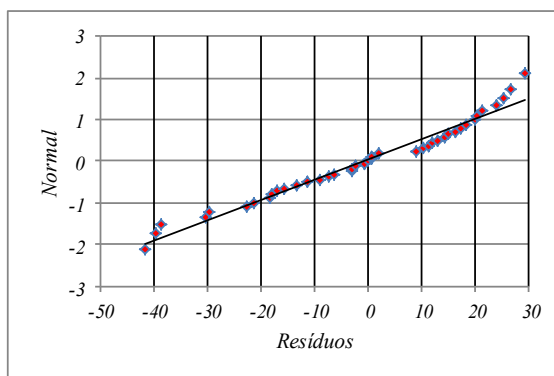


Figura 6 – Gráfico de Papel de Probabilidade para os Resíduos.

- Diagnóstico de Homocedasticidade: Utilizando o teste de Goldfeld-Quandt, de acordo com os procedimentos relatados no item 6 tem-se que  $F_{GD}$  é igual a 4,065. O P-valor do teste,  $P(F > 4,065)$  é 0,0061. Como  $F_{GD} > P$ -valor, considerando um nível de confiança de 95%, não rejeita-se a hipótese de homocedasticidade.
- Diagnóstico de Independência dos Dados: Com base nos dados da tabela 10 e aplicando o teste de Durbin-Watson (item 6) tem-se que se  $\sum_{i=2}^n (e_i - e_{i-1})^2 = 38315,16$  e  $\sum_{i=1}^n e_i^2 = 15299,03$  então substituindo na fórmula (16) temos que  $d_w = 2,5$ . Para encontrar os valores críticos  $d_L$  e  $d_U$  na tabela de Durbin-Watson deve-se considerar  $\alpha = \alpha^*/2$ , em que  $\alpha^*$  é o nível de significância do teste. Desta forma, para  $\alpha^* = 0,05$  tem-se  $\alpha = 0,025$ ,  $n = 41$  e  $k = 1$ , logo  $d_L = 1,358$  e  $d_U = 1,456$ . Como  $d_U = 1,456 < 2,504 < 4 - d_U = 2,544$ , não se rejeita  $H_0$  e, portanto pode-se afirmar que para um nível de confiança de 95%, os resíduos são independentes.

Tabela 10 – Medidas para o cálculo da estatística de Durbin-Watson.

CÓD.	$\hat{e}_i$	$\hat{e}_i - \hat{e}_{i-1}$	$(\hat{e}_i - \hat{e}_{i-1})^2$	$\hat{e}_i^2$	CÓD.	$(\hat{e}_i)$	$\hat{e}_i - \hat{e}_{i-1}$	$(\hat{e}_i - \hat{e}_{i-1})^2$	$\hat{e}_i^2$
CE028	19,95			398,00	GU036	-0,02	8,88	78,85	0,00
CE033	12,95	-7,00	48,94	167,81	GU052	2,04	2,06	4,22	4,14
CE034	-38,68	-51,63	2666,07	1496,14	GU148	25,44	23,41	547,98	647,40
CE035	-17,02	21,66	469,16	289,68	GU172	10,38	-15,07	227,07	107,64
CE103	-2,82	14,20	201,64	7,95	GU176	-29,77	-40,15	1611,62	886,25
CE107	0,67	3,49	12,18	0,45	GU209	20,30	50,07	2507,31	412,21
CE221	17,30	16,63	276,56	299,29	GU220	-17,92	-38,22	1461,00	321,13
CE223	-21,27	-38,57	1487,64	452,41	GU293	-2,44	15,48	239,63	5,95
CE230	26,87	48,14	2317,56	722,05	GU314	23,97	26,41	697,65	574,70
CE247	-22,62	-49,49	2449,36	511,66	GU543	-39,48	-63,45	4026,28	1558,67
CE279	-7,29	15,33	235,01	53,14	GU605	0,69	40,17	1613,47	0,47
CE280	9,18	16,47	271,26	84,27	GU616	-15,62	-16,31	265,95	243,98
CE306	-2,82	-12,00	144,00	7,95	GU617	15,03	30,65	939,42	225,90
CE331	-13,37	-10,55	111,26	178,70	GU618	-6,17	-21,20	449,44	38,07
CE340	29,54	42,91	1841,10	872,61	GU623	12,01	18,18	330,55	144,26
CE342	-30,15	-59,69	3562,90	909,02	GU627	18,40	6,39	40,79	338,49
CE407	21,53	51,68	2670,82	463,54	GU633	-11,39	-29,79	887,32	129,73
CE412	11,53	-10,00	99,96	132,99	GU729	-41,52	-30,13	907,82	1723,91
CE445	14,46	2,93	8,57	209,09	GU736	-0,78	40,74	1659,75	0,61
CE655	16,33	1,87	3,50	266,67	GU737	-18,24	-17,46	304,99	332,84
CF093	-8,90	-25,23	636,55	79,21					



- Pontos Influentes: O gráfico (Figura 7) dos resultados obtidos para o cálculo do DFFITS (Equação 17) mostra que nenhum valor é, em módulo, maior que 1. Portanto, não existe observação que represente um ponto influente.

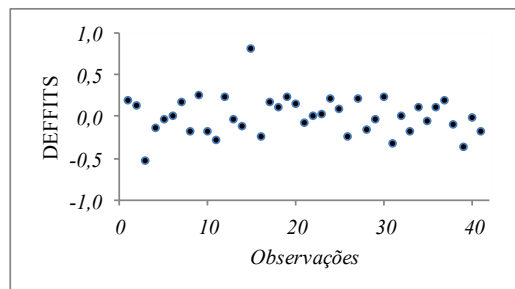


Figura 7 – Pontos Influentes.

## 8. CONCLUSÕES

- A área localiza-se na região sudeste do Piauí e abrange os municípios de Canto do Buriti, Tamboril do Piauí, Pajeú do Piauí, Brejo do Piauí, Socorro do Piauí, Ribeira do Piauí, João Costa, Pedro Laurentino, São João do Piauí e Nova Santa Rita;
- O aquífero Cabeças é a unidade hidrogeológica que ocupa a maior parte da área. Apresenta-se predominantemente livre, excetuando-se, numa reduzida área, a noroeste; onde está confinada pela formação Longá. Possui espessura média de 150 metros, capacidade específica  $1,77 \text{ m}^3/\text{h}/\text{m}$ , transmissividade de  $6,39 \times 10^{-3} \text{ m}^2/\text{s}$  e condutividade hidráulica de  $5,32 \times 10^{-5} \text{ m/s}$ ;
- O Aquífero Cabeças apresenta produtividade moderada, classe 3, variando a alta, classe 2, de acordo com a classificação de Struckmeir & Margat(1995) modificado por Diniz et al. (2012);
- A *Correlação* mede a força ou o grau de relacionamento entre duas variáveis e a *Análise de Regressão* expressa esse relacionamento através de um modelo matemático. Desta forma, o objetivo principal deste trabalho foi estabelecer uma equação matemática, linear, que descreva a relação entre a condutividade elétrica (CE) e os sólidos totais dissolvidos, para o aquífero Cabeças, na área em estudo;
- O *Diagrama de Dispersão* indicou uma tendência linear entre as variáveis, com os pontos se concentrando em pequenos desvios, em relação a uma reta imaginária, que passa entre eles. Observa-se que valores crescentes da variável independente (CE) estão associados a valores crescentes da variável resposta (STD);
- Na estimação dos parâmetros dos parâmetros “a” e “b” do modelo foi utilizado o *Método dos Mínimos Quadrados*. Os valores encontrados foram os seguintes:  $a=-11,52$  e  $b=0,69$ , representando o intercepto e o coeficiente de regressão, respectivamente;
- O *Modelo Ajustado* é dado pela equação de regressão  $STD=0,69*CE-11,52$ ;

- As *Previsões* para sólidos totais dissolvidos (STD) devem ser realizadas dentro dos limites da condutividade elétrica (CE) que é entre 31,4 e 989  $\mu\text{S}/\text{cm}$ . Deve-se ter muito cuidado ao extrapolar valores, fora do âmbito dos dados amostrais, uma vez que a mesma relação pode não ser verificada;
- A avaliação da *Significância da Regressão* foi feita empregando o *Teste F de Snedecor*. O resultado mostrou que para o nível de significância  $\alpha=5\%$ , tem-se  $F_{\text{calculado}}$  maior que  $F_{\text{tabelado}}$ . Portanto, rejeita-se  $H_0$ , com o nível de confiança de 95%, e conclui-se que a variável explicativa tem correlação com a variável resposta;
- O *Coefficiente de Correlação de Pearson (R)* indica que ocorre uma correlação fortíssima e positiva, entre as variáveis;
- O *Coefficiente de Determinação (R<sup>2</sup>)* mostra que aproximadamente 98 % da variação dos sólidos totais dissolvidos (variável resposta) é explicado pela variação da condutividade elétrica (variável regressora);
- O diagnóstico de *Normalidade, Homocedasticidade, Independência dos Resíduos e Pontos Influentes* validou de forma satisfatória o modelo de regressão ajustado;
- Os resultados mostraram que o modelo matemático que expressa a correlação entre a Condutividade Elétrica e os Sólidos Totais Dissolvidos são confiáveis a nível de 95% e podem ser utilizados pelo Serviço Geológico do Brasil (CPRM), Órgãos Gestores de Recursos Hídricos e pela comunidade científica em geral, especificamente para a área estudada.

## 9. BIBLIOGRAFIA

AQUABASE, 2010. O que é um TDS? Disponível em: < <http://aquabahiasergipe.com.br>>. Acesso em: 10/10/2013.

APHA, 1992. Standard Methods for the Examination of Water and Wasterwater. 18 th edition. American Public Health Association. Washington, DC.

DINIZ, J. A. O.; MONTEIRO, A. B.; FEITOSA, F. A. C.; FREITAS, M. A. e PEIXINHO, F. C.; 2012. Metodologia para Elaboração de Mapas Hidrogeológicos. XVII Congresso Brasileiro de Águas Subterrâneas; XVIII Encontro Nacional de Perfuradores de Poços. 23-26 out. 2012, Bonito, MS. Resumos Expandidos... Bonito, MS: CBAS, out. 2012. p. 1-4.

CORREIA FILHO, F. L. e MONTEIRO, A. B., 2010. Projeto Borda Sudeste da Bacia Sedimentar do Parnaíba: relatório final. Teresina: CPRM, 153 p.

LANDIM, Paulo, M. Barbosa, 2000. Análise Estatística de Dados Geológicos Multivariados. Departamento de Geologia Aplicada. Laboratório de Geomatématica. UNESP/Campus de Rio Claro. Texto Didático 03. Rio Claro, SP.

LAPPONI, J.C.; 1997. Estatística usando Excel 5 e 7. São Paulo: Lapponi Treinamentos e Editora.

METCALF & EDDY. 1991. Wastewater engineering: treatment, disposal and reuse. 3ed. New York: McGraw-Hill, 1334 p.

MIRANDA, E. E. de; (Coord.). Brasil em Relevo. Campinas: Embrapa Monitoramento por Satélite, 2005. Disponível em: <<http://www.relevobr.cnpm.embrapa.br>>. Acesso em: 9 ago. 2011.

PORTAL ACTION, 2012. Software de Estatística. Conteúdo Estatístico e Manual Action. [www.portalaction.com.br](http://www.portalaction.com.br).

STRUCKMEIR, W.F. & MARGAT, J.,1995. Hydrogeological maps: a guide and a standard legend. IAH International Contributions to Hydrogeology. Vol 17/1995. International Association of Hydrogeologists (IAH).