



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DESENVOLVIMENTO DE BANCO DE DADOS PARA ARMAZENAMENTO E DISTRIBUIÇÃO DE DADOS ECO-HIDROLÓGICOS

Maria Clara Oliveira Domingos Ruas

Relatório de Iniciação Científica do
programa PIBIC, orientada pela Dra.

Laura de Simone Borma

INPE

São José dos Campos

2022



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DESENVOLVIMENTO DE BANCO DE DADOS PARA ARMAZENAMENTO E DISTRIBUIÇÃO DE DADOS ECO-HIDROLÓGICOS

Maria Clara Oliveira Domingos Ruas

Relatório de Iniciação Científica do
programa PIBIC, orientada pela Dra.

Laura de Simone Borma

INPE

São José dos Campos

2022

RESUMO

Os dados eco-hidrológicos são coletados periodicamente, acumulando-se e exigindo muitas horas de trabalho no tratamento manual dos dados coletados. Assim, o principal objetivo do projeto é tornar mais eficiente o tratamento e exibição dos conteúdos coletados. Para isso, primeiramente foi feito um levantamento de todos os dados, como eles são coletados e armazenados, deixando explícito a maneira exata que ocorre a captação, o que cada dado significa no documento gerado (.xlsx), também sobre o que se espera de cada campo e o que seria considerado dado inválido e por fim quais são as saídas esperadas. No segundo passo, com a padronização já em mente, é realizada uma transformação dos dados e a importação, em que as informações são colocadas dentro do banco de dados em um formato de tabelas, contendo diferentes chaves de acesso definidas no passo anterior. A seguir, ainda no banco de dados, é realizado um tratamento sobre dados inválidos que com as orientações do levantamento, tomam as medidas necessárias, seja o arredondamento de valores ou exclusão. É bom ressaltar que a quantidade e o tipo de dado inválido são informações que posteriormente serão uteis em relatórios de eficiência e manutenção. Por fim, é exibido ao usuário ou cliente final aquilo que foi requisitado, no caso, os dados tratados em forma de gráficos e tabelas. Como resultado, podem ser citados a economia de tempo ao não se precisar tratar os dados manualmente, a segurança dos dados, já que o sistema evita qualquer perda ou irregularidade em meio aos processos, e por fim, a fácil acessibilidade para consulta em pesquisas posteriores.

Palavras-chave: Automatização de Processos. Volume de Dados. Segurança. Eficiência. Process Automation. Data Volume. Security. Efficiency.

LISTA DE FIGURAS

	<u>Pág.</u>
1.3.1 Imagem do momento de coleta de informação de manutenção dos Loggers Fonte: Laura Borma.....	6
1.3.2 Sensores de Umidade do Solo, Potencial Matricial e Temperatura. Fonte: Laura Borma.....	6
1.3.3 Estação Meteorológica. Fonte: Laura Borma	7
1.3.4 Sensores de Umidade do Solo e Sensores de Infiltração. Fonte: Laura Borma.....	7
1.3.5 Sensor de Fluxo de Seiva. Fonte: Laura Borma.....	8
2.1.1 Esquema com etapas da solução. Fonte: produção da própria autora.....	9
2.2.1 Distribuição de dados pela região. Fonte: produção da própria autora.....	10
2.3.1 Diagrama da Arquitetura do Banco. Fonte: produção da própria autora	13
2.3.2 Fluxo de todo o processo. Fonte: produção da própria autora.....	14
2.4.1 Gráficos gerados a partir dos dados tratados. Fonte: Laura Borma.....	15

Sumário

1	Introdução	5
	1.1. Contexto Geral	5
	1.2. Objetivos	5
	1.3. Situação Trabalhada.....	5
2	Desenvolvimento	9
	2.1. Visão Geral	9
	2.2 Detalhamento sobre o Armazenamento	10
	2.3 Processo de Tratamento	11
	2.4 Resultados	14
3	Conclusão.....	16
4	Referências Bibliográficas.....	17

1 Introdução

1.1 Contexto Geral

Com os avanços da tecnologia, hoje em dia é possível se coletar muita informação e possuir diversas formas acessíveis de acumulá-la. O armazenamento de memória deixou de ser um problema com melhorias dos aparelhos eletrônicos e o barateamento de opções como HD's externos, servidores locais ou até mesmo o aluguel de servidores na nuvem. Juntamente com tal facilidade de acúmulo, também é preciso pensar não apenas em guardar os dados, mas também em como são armazenados e organizados.

A grande quantidade carrega consigo também um sobrecarregamento de informações, e quando estão sem uma estrutura definida, é fácil se perder ou não conseguir visualizar o que realmente está acontecendo. Por isso, a criação de uma metodologia personalizada para o problema para o tratamento e organização dos dados é de grande importância para os pesquisadores em questão.

1.2 Objetivos

O trabalho desenvolvido nesse projeto se trata da criação de um banco de dados relacional com a finalidade de organizar os dados coletados de pesquisas de recursos hídricos, meteorológicos e ambientais de forma automatizada e eficaz. Além da vantagem da organização estruturada, o desenvolvimento do banco também traz consigo uma agilidade no processo do tratamento. Dessa forma os pesquisadores podem concentrar seu esforço em atividades que exigem maior senso crítico e conhecimento, como análise e interpretação dos dados tratados, desenvolvimento de conclusões e até mesmo de novas hipóteses.

1.3 Situação Trabalhada

Como já citado anteriormente, são realizadas diversas coletas de dados desde o ano de 2019 em regiões próximas de São José dos Campos. Nelas, existem vários sensores de diversos tipos monitorados pelos profissionais responsáveis pela pesquisa.

Figura 1.3.1 – Coleta de informação e manutenção dos Loggers



Fonte: Laura Borma

Figura 1.3.2 - Sensores de Umidade do Solo, Potencial e Temperatura



Fonte: Laura Borma

Existem no escopo do projeto três sensores a serem estudados, modelados e tratados: coletores de dados do solo, da atmosfera e das árvores. Cada um dos tipos possui peculiaridades distintas e um processo de tratamento diferente. Para se construir um banco de dados funcional e útil, é preciso estudar bem sobre o caminho e realizar o julgamento de qual é a melhor forma de armazená-los e exibi-los.

Figura 1.3.3 - Estação Meteorológica



Fonte: Laura Borma

Figura 1.3.4 - Sensores de umidade do solo e sensores de infiltração



Fonte: Laura Borma

Figura 1.3.5 - Sensor de fluxo de seiva



Fonte: Laura Borma

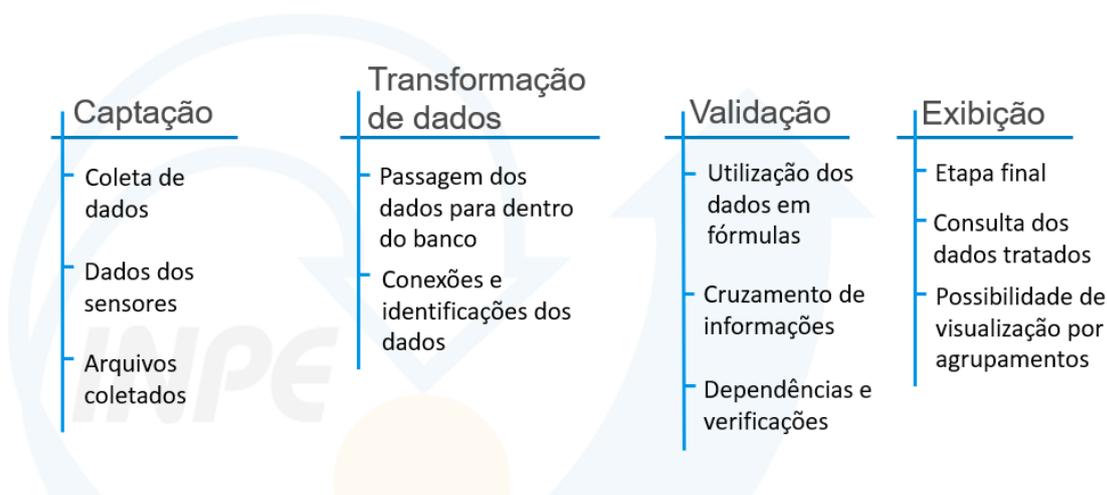
É importante ressaltar que para a pesquisa e desenvolvimento do banco não são necessários conhecimentos específicos sobre as informações que estão sendo coletadas, e sim sobre o caminho dos dados. Por isso, o auxílio dos pesquisadores responsáveis para a explicação do processo realizado é de grande importância. Neste trabalho será explicado melhor justamente sobre o processo de compreensão do caminho dos dados e a modelagem da estrutura e arquitetura do banco.

2 Desenvolvimento

2.1 Visão Geral

Tendo uma visão geral da solução, o processo dos dados tem a princípio 4 etapas: captação, transformação de dados, validação e exibição. A captação é referente a coleta dos dados, envolvendo dados de sensores, localização e os arquivos coletados. Na transformação de dados, é preciso de um detalhamento sobre como os dados são armazenados. Esse detalhamento envolve especificações de como cada dado é armazenado, como localiza-los, quais são suas unidades de medida, sensores e qual é a porta associada a ele. Feito as devidas especificações, é possível realizar a importação dos dados corretamente e partir para a validação. Essa etapa se caracteriza por todas as contas e verificações necessárias para a entrega de resultados. Por fim, com todos os cálculos feitos, a exibição é feita conforme a necessidade da pesquisa.

Figura 2.1.1 - Esquema com todas as etapas da solução

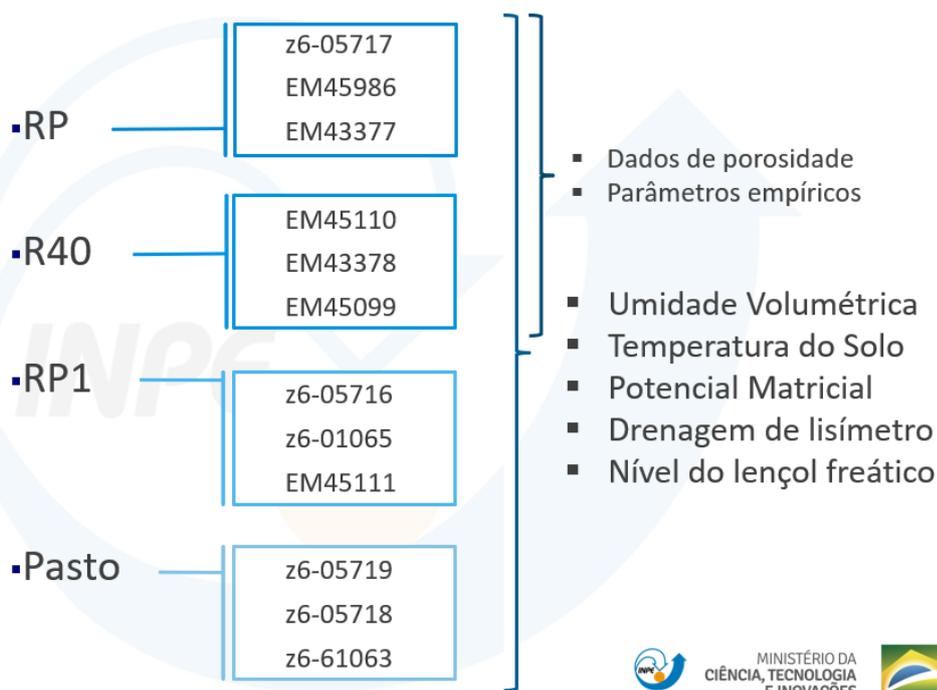


Fonte: produção da própria autora

2.2 Detalhamento sobre o Armazenamento

Em 4 regiões (RP, R40, RP1 e Pasto), estão distribuídos o total de 12 sensores. Os mesmos coletam dados de umidade volumétrica, potencial matricial e temperatura do solo em 7 diferentes profundidades (2.0, 1.5, 1.0, 0.6 ou 0.5 -dependendo da região-, 0.3, 0.2 e 0.1). Além disso, também é medido dados de drenagem do lisímetro e nível do lençol freático. Os dados são coletados a partir da data de 20 de fevereiro de 2019, coletando dados periodicamente, com intervalos de 15 minutos, posteriormente aumentados para 1 hora em 1 hora. O armazenamento nas colunas de cada dado segue conforme a tabela. Adicionalmente aos dados coletados, as regiões R40 e RP possuem variáveis próprias referentes a porosidade e parâmetros empíricos (variam de acordo com a profundidade) que serão utilizadas em cálculos no processo de tratamento.

Figura 2.2.1 – Distribuição de dados pela região



Fonte: produção da própria autora

Os dados são armazenados nas planilhas e identificados de acordo com sua posição na coluna. Dentro do arquivo, é possível identificar qual é o dado sendo coletado de acordo com seu arquivo de origem, a data ao qual está associado e em qual coluna se localiza. Para melhor organização, foi realizado uma tabela para localização rápida.

Tabela 2.2.1 – Disposição dos dados nos arquivos

Logger	Coluna 1	Coluna 2	Coluna 3	Coluna 4	Coluna 5	Coluna 6	Coluna 7	Coluna 8	Coluna 9	Coluna 10	Coluna 11	Coluna 12	Coluna 13	
z6-05716	Data e hora	Umidade Volumétrica 2,0 m	Umidade Volumétrica 1,5 m	Umidade Volumétrica 1,0 m	Umidade Volumétrica 0,5 m	Umidade Volumétrica 0,3 m	Umidade Volumétrica 0,1 m							
z6-01065	Data e hora	Potencial Matricial 2,0 m	Temperatura do Solo 2,0 m	Potencial Matricial 1,5 m	Temperatura do Solo 1,5 m	Potencial Matricial 1,0 m	Temperatura do Solo 1,0 m	Potencial Matricial 0,5 m	Temperatura do Solo 0,5 m	Potencial Matricial 0,3 m	Temperatura do Solo 0,3 m	Potencial Matricial 0,1 m	Temperatura do Solo 0,1 m	
EM45111	Data e hora	Potencial Matricial 0,2 m	Temperatura do Solo 0,2 m	Umidade Volumétrica 0,2 m										
z6-05717	Data e hora	Umidade Volumétrica 2,0 m	Umidade Volumétrica 1,5 m	Umidade Volumétrica 1,0 m	Umidade Volumétrica 0,6 m	Umidade Volumétrica 0,3 m	Lençol							
EM45986	Data e hora	Potencial Matricial 2,0 m	Temperatura do Solo 2,0 m	Potencial Matricial 1,5 m	Temperatura do Solo 1,5 m	Potencial Matricial 1,0 m	Temperatura do Solo 1,0 m	Potencial Matricial 0,6 m	Temperatura do Solo 0,6 m	Umidade Volumétrica 0,2 m				
EM43377	Data e hora	Potencial Matricial 0,3 m	Temperatura do Solo 0,3 m	Potencial Matricial 0,2 m	Temperatura do Solo 0,2 m	Potencial Matricial 0,1 m	Temperatura do Solo 0,1 m	Potencial Matricial 0,1 m	Lisímetro					
EM45110	Data e hora	Umidade Volumétrica 2,0 m	Umidade Volumétrica 1,5 m	Umidade Volumétrica 1,0 m	Umidade Volumétrica 0,5 m	Umidade Volumétrica 0,3 m								
EM43378	Data e hora	Potencial Matricial 2,0 m	Temperatura do Solo 2,0 m	Potencial Matricial 1,5 m	Temperatura do Solo 1,5 m	Potencial Matricial 1,0 m	Temperatura do Solo 1,0 m	Potencial Matricial 0,5 m	Temperatura do Solo 0,5 m	Umidade Volumétrica 0,1 m				
EM45099	Data e hora	Potencial Matricial 0,3 m	Temperatura do Solo 0,3 m	Potencial Matricial 0,2 m	Temperatura do Solo 0,2 m	Potencial Matricial 0,1 m	Temperatura do Solo 0,1 m	Umidade Volumétrica 0,2 m	Lisímetro					
z6-05719	Data e hora	Umidade Volumétrica 0,1 m	Umidade Volumétrica 0,3 m	Umidade Volumétrica 0,5 m	Umidade Volumétrica 1,0 m	Umidade Volumétrica 1,5 m	Umidade Volumétrica 2,0 m							
z6-05718	Data e hora	Potencial Matricial 0,1 m	Temperatura do Solo 0,1 m	Potencial Matricial 0,3 m	Temperatura do Solo 0,3 m	Potencial Matricial 0,5 m	Temperatura do Solo 0,5 m	Potencial Matricial 1,0 m	Temperatura do Solo 1,0 m	Potencial Matricial 1,5 m	Temperatura do Solo 1,5 m	Potencial Matricial 2,0 m	Temperatura do Solo 2,0 m	
z6-01063	Data e hora	Lisímetro					Potencial Matricial 0,2 m	Temperatura do Solo 0,2 m	Umidade Volumétrica 0,2 m					

Fonte: produção da própria autora

2.3 Processo de Tratamento

O tratamento de todos os dados envolve a exclusão de dados inválidos - armazenados na forma de string, dados ausentes constam na tabela como #N/D, #N/A ou #VALUE!-, e primeiramente, a realização de médias diárias dos dados de umidade volumétrica, potencial matricial e temperatura do solo para todas as regiões e profundidades. Isto é, a soma de todos os valores coletados em um único dia dividido pela quantidade de dados coletados.

Para as regiões RP e R40, existem mais contas a serem feitas. Primeiro, contas referentes à umidade volumétrica medida, contas de grau de saturação e capacidade de campo. Com a finalidade de comparação de resultados, também é feito um cálculo estimado para o ajuste da curva de retenção, obtendo-se um valor teórico para a umidade volumétrica, o qual também é feito as contas de grau de saturação e capacidade de campo.

$$GS = \frac{SM}{Poros}$$

Fórmula de Grau de Saturação, onde SM é a umidade do solo e Poros a porosidade na profundidade estudada.

$$FC = \frac{SM}{kpa}$$

Fórmula de Capacidade de Campo, onde SM é a umidade do solo e kpa é uma constante do campo estudado.

$$SMc = UR + \frac{Poros - UR}{(1 + (AC.PM.(-10))^n)^m}$$

Fórmula para Umidade do Solo calculada, onde UR é umidade residual, Poros a porosidade na profundidade estudada, AC parâmetro empírico de ajuste da curva, PM o potencial matricial e n e m são parâmetros empíricos da função.

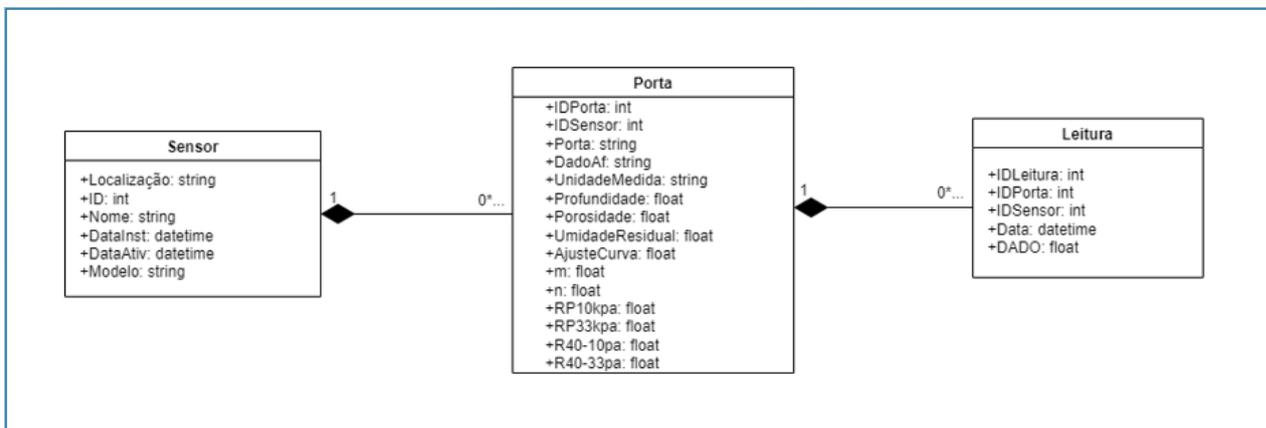
Com todo o processo e caminho dos dados explicitados, agora é preciso a construção da arquitetura do banco de dados. Para sua construção, são criadas 3 entidades/objetos: sensores, portas e leituras. Cada sensor pode estar ligado a várias portas e cada porta possui vários dados aferidos. Indo pelo sentido inverso, uma leitura só pode estar associada a uma única porta, e cada porta só pode estar associada a um único sensor. Cada classe possui seus atributos, e suas chaves (identificadores de sua porta ou sensor).

Tabela 2.3.1 – Dicionário de dados

Nome da Tabela	Coluna	Descrição	Tipo de Dado	Tamanho
Sensores	IDSensor		Int	
	Nome	Identificação do Sensor	varchar	50
	Areas	Localização Física	varchar	50
	DTinst	Data de instalação	datetime	
	Dtatv	Data de ativação	datetime	
	Modelo	Modelo do sensor	varchar	50
	Tiposensor		varchar	50
Portas	IDPorta		Int	
	IDSensor		Int	
	NumPorta		varchar	50
	Profundidade		float	
	Porosidade		float	
	DadoAf		varchar	100
	UnidadeMedida		varchar	10
Leituras	IDLeitura		Int	
	IDSensor		Int	
	IDPorta		Int	
	Data		datetime	
	DadoValido		float	
	DadoInvalido		varchar	10

Fonte: produção da própria autora

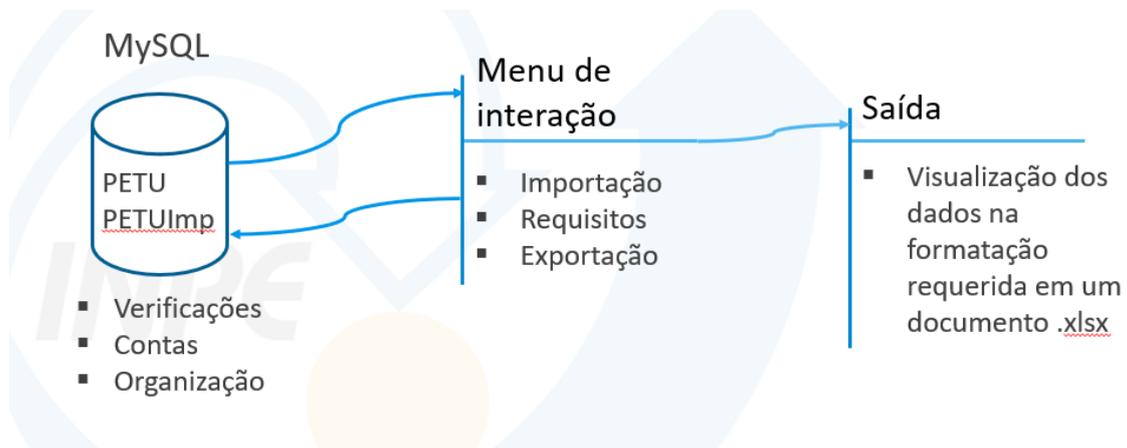
Figura 2.3.1 – Diagrama da Arquitetura do Banco



Fonte: produção da própria autora

Assim, para criação de cada objeto, é necessário fornecer em qual entidade ele está ligado. Após a definição de cada variável, é possível realizar a importação e organização de cada dado dependendo de sua posição dentro do arquivo e do próprio arquivo em si. A importação para o banco de dados é feita. O banco de dados utilizado foi o MySQL.

Figura 2.3.2 – Fluxo de todo o processo

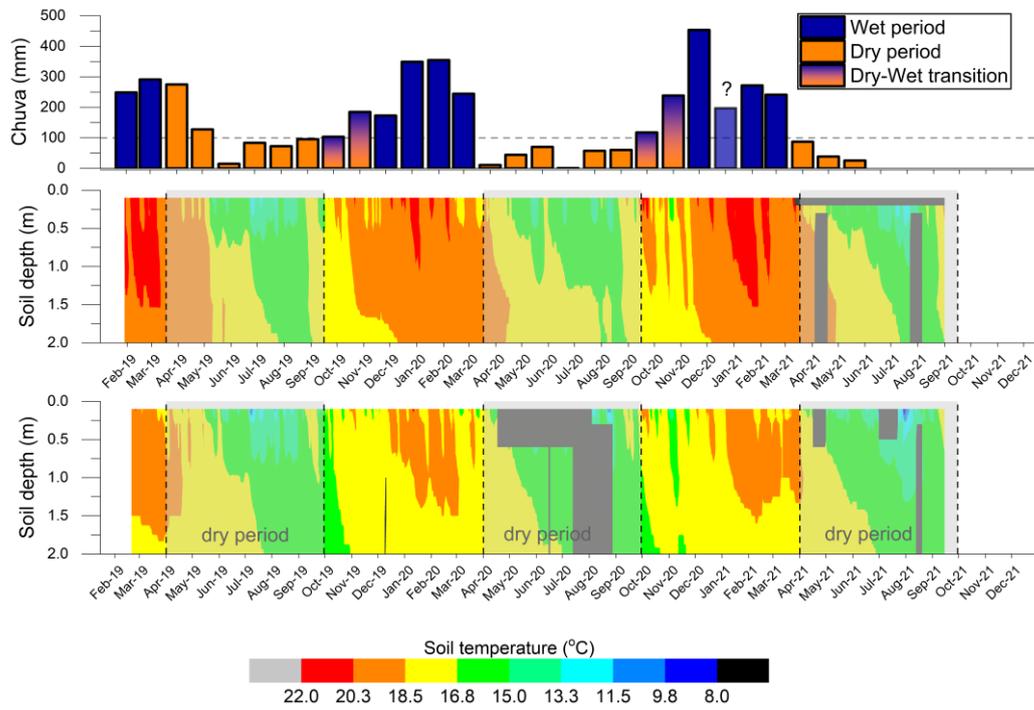


Fonte: produção da própria autora

2.4 Resultados

Com todos os dados devidamente rotulados, é possível realizar todas as contas de maneira rápida e eficiente. Os resultados organizados em tabelas temporárias, dentro do próprio banco garantem a possibilidade de realizar a exposição da maneira desejada. Com os dados organizados, basta copiá-los e transferi-los para uma tabela Excel e entregar aos responsáveis pela realização de gráficos e análises. O Excel foi escolhido devido a familiaridade dos pesquisadores com o software, junto com a praticidade e eficiência.

Figura 2.4.1 – Gráficos gerados a partir dos dados tratados



O primeiro gráfico é referente às chuvas ao longo do tempo, o segundo às temperaturas do solo na região RP e o último às temperaturas do solo na região R40.

Fonte: Laura Borma.

3 Conclusão

Com o que foi feito no último ano, já houve um grande avanço em sentidos de tratamento de dados e otimização. O processo que antes levava de 5 a 7 dias agora pode ser realizado em menos de 5 minutos. Assim, podemos dizer como o objetivo inicial do projeto está próximo de ser completado.

Entretanto, ainda existem pontas soltas para serem conectadas. A seguir, o principal objetivo seria o polimento do processo, tornando-o mais automatizado, além de sua manutenção, acompanhando seu funcionamento a longo prazo. Após a construção de um ambiente mais automático, também é importante ensinar aos usuários como utilizar o programa e oferecer a autonomia para os mesmos, e também continuar o mesmo processo realizado no solo para as árvores e para a atmosfera, estas que estão na etapa do levantamento dos dados.

4 Referências Bibliográficas

TEIXEIRA, Paulo Costa; DONAGEMMA, Giula Keyla; FONTANA, Andre; TEIXEIRA, Wagner Gomes. **Manual de métodos de análise de solo**. 3ª. ed., 2017.

Plataforma Moretti. RFDP | Jocelito Buch Castro da Cruz - Ajuste da Curva de Retenção da Água do Solo com o Solve Excel.. YouTube, 16/10/2018. Disponível em < https://www.youtube.com/watch?v=GHM_44IPRAc >. Acesso em: 15/11/2021.