

# **Aplicação de NoSQL em dados de redes de sensores sem fio: um estudo de caso utilizando MongoDB**

---

Danilo de Souza Miguel  
Roberto Fray da Silva

PCS5031 - Introdução à Ciência dos Dados

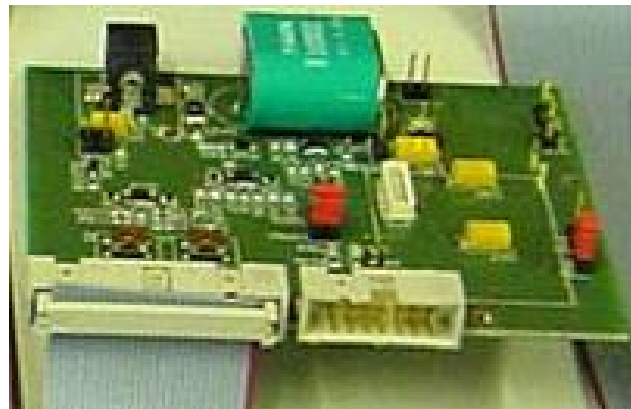
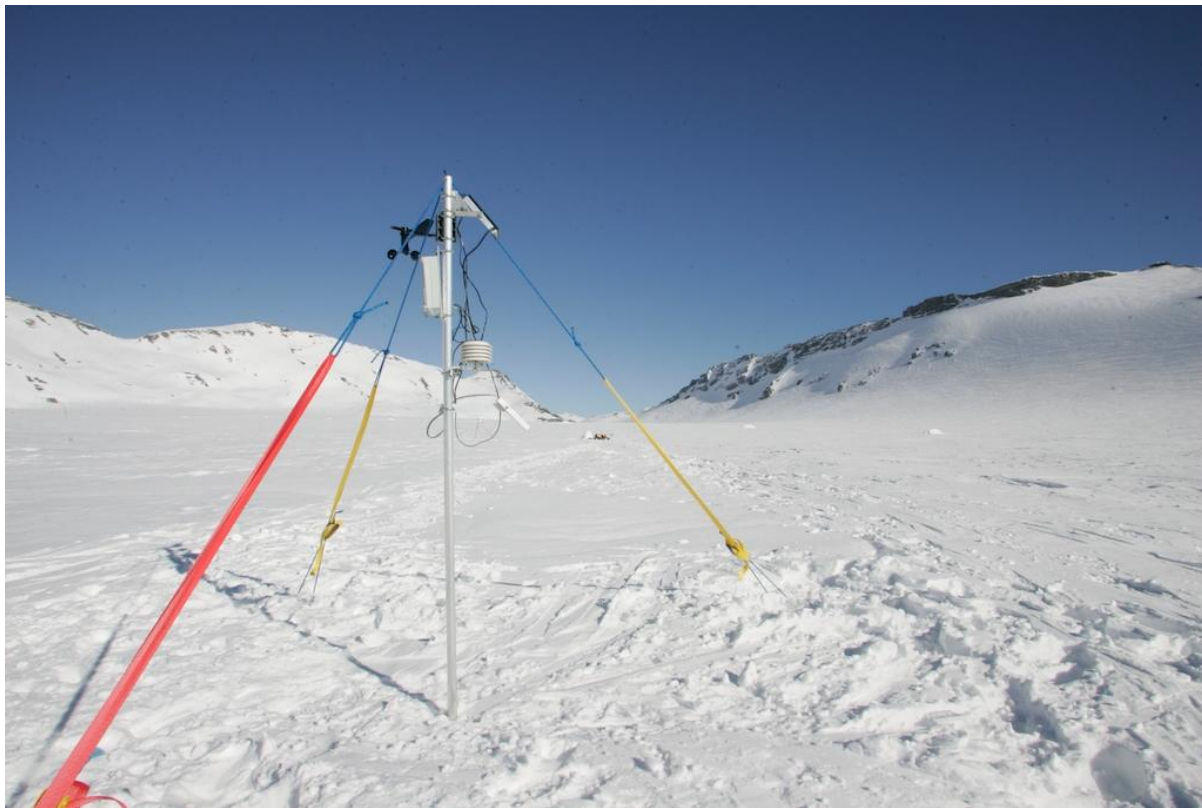
# Introdução

- **RSSF** → sensores espalhados em um ambiente → observar e **monitorar um fenômeno**
- **Grande quantidade de dados**, que podem ter características distintas

# Descrição do dataset

- Dataset → projeto **Sensorscope** → **RSSF** de baixo custo para monitoramento ambiental em montanhas na Suíça
- Set e Out 2008, 23 estações, **588.524** medições

# Projeto Sensorscope



# Descrição do dataset

- **Sensores** → temperatura (ar e solo), radiação solar, umidade (relativa e solo), velocidade e direção do vento
- **Partes:** localização geográfica das estações (2) e dados coletados (24)

# Objetivo

- Analisar a **aplicação de NoSQL** e técnicas de **análises estatísticas** para dados de **monitoramento ambiental** do projeto Sensorscope

# Metodologia

1. **Descrição** do dataset
2. Estudo de **ciclo de vida** dos dados (DMPTool)
3. Elaboração de **questões de pesquisa**
4. Escolha dos **software** (arm., proc., vis.)
5. Realização das **análises**
6. **Avaliação** dos resultados
7. **Relatório final e compartilhamento** dos dados e resultados via GitHub

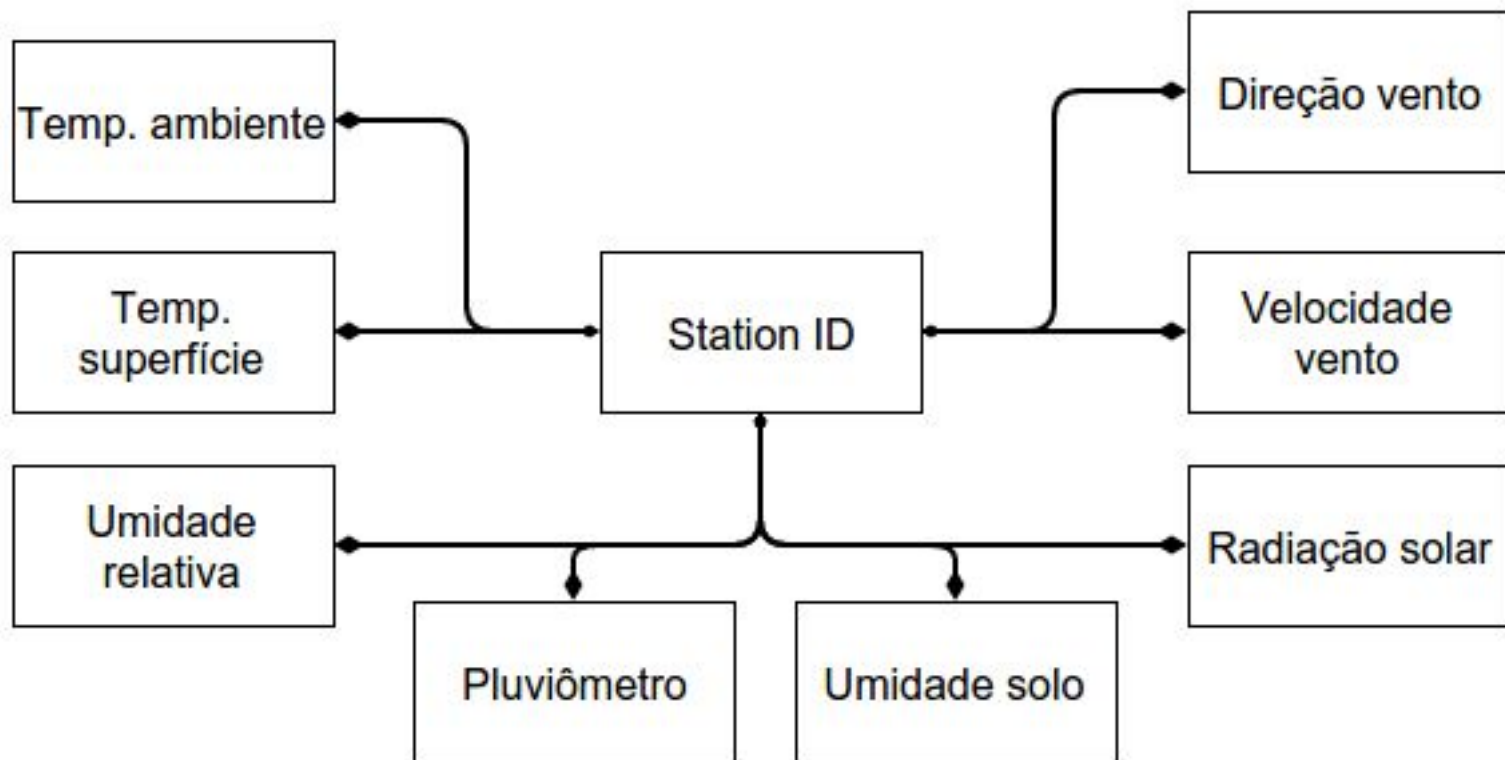
# Metodologia

- Coleta → **Google Chrome**
- Armazenamento de dados → **MongoDB**
- Pré-processamento → **Rstudio**
- Processamento e análise de dados → **Rstudio e LibreOffice Calc**
- Visualização dos dados → **R Shiny**
- Publicação dos resultados → **GitHub, Datacite, Artigo evento**



# Coleta e armazenamento

- **Site** oficial do projeto
- Análise da **estrutura**
- Grande **quantidade de medições** + **velocidade de inserção** no banco + **acesso** → **NoSQL**
- Redis, Neo4j, Cassandra, **MongoDB**



# Pergunta 1

- Quais as médias, modas e desvios padrão para as variáveis?

	<b>Média</b>	<b>Moda</b>	<b>DesvPad</b>
Temperatura ambiente(c)	3	-1	5
Umidade relativa(%)	65	2	18
Umidade do solo(%)	21	20	8
Radiação solar(w/m2)	122	0	206
Temperatura do solo(c)	3	1	7
Marca d'água(kPa)	-11	-1	4
Direção do vento(o)	150	0	80
Velocidade do vento(m/s)	2	1	2

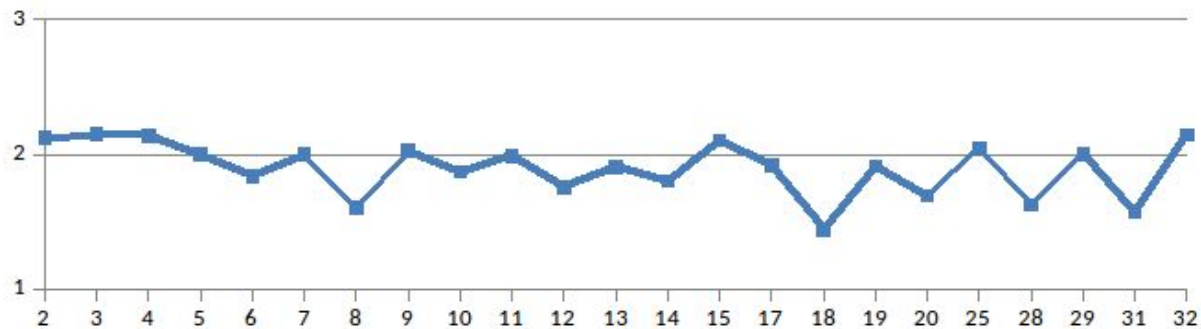
## Pergunta 2

- Qual é a distribuição da direção do vento (graus) ao longo da amostra?

**Direção vento (graus)**



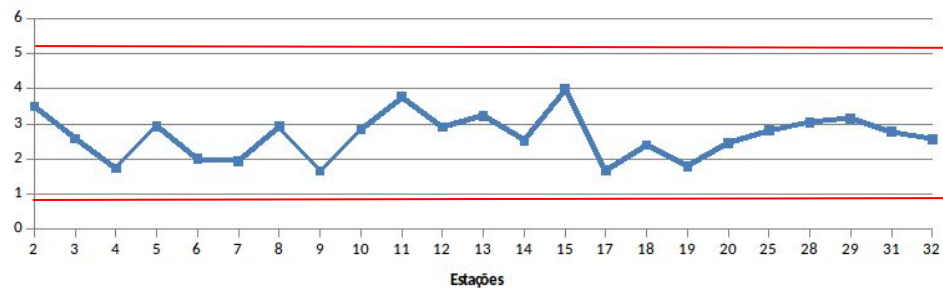
**Velocidade vento (m/s)**



# Pergunta 3

- Existem diferenças significativas entre as observações das diferentes estações?
  - Sim, nas **temperaturas**

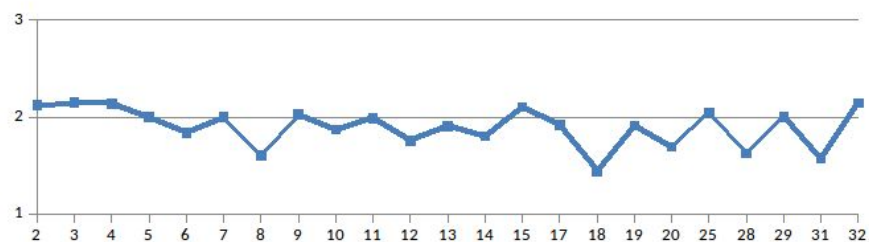
**Temperatura ambiente (C)**



**Temperatura solo (C)**



**Velocidade vento (m/s)**



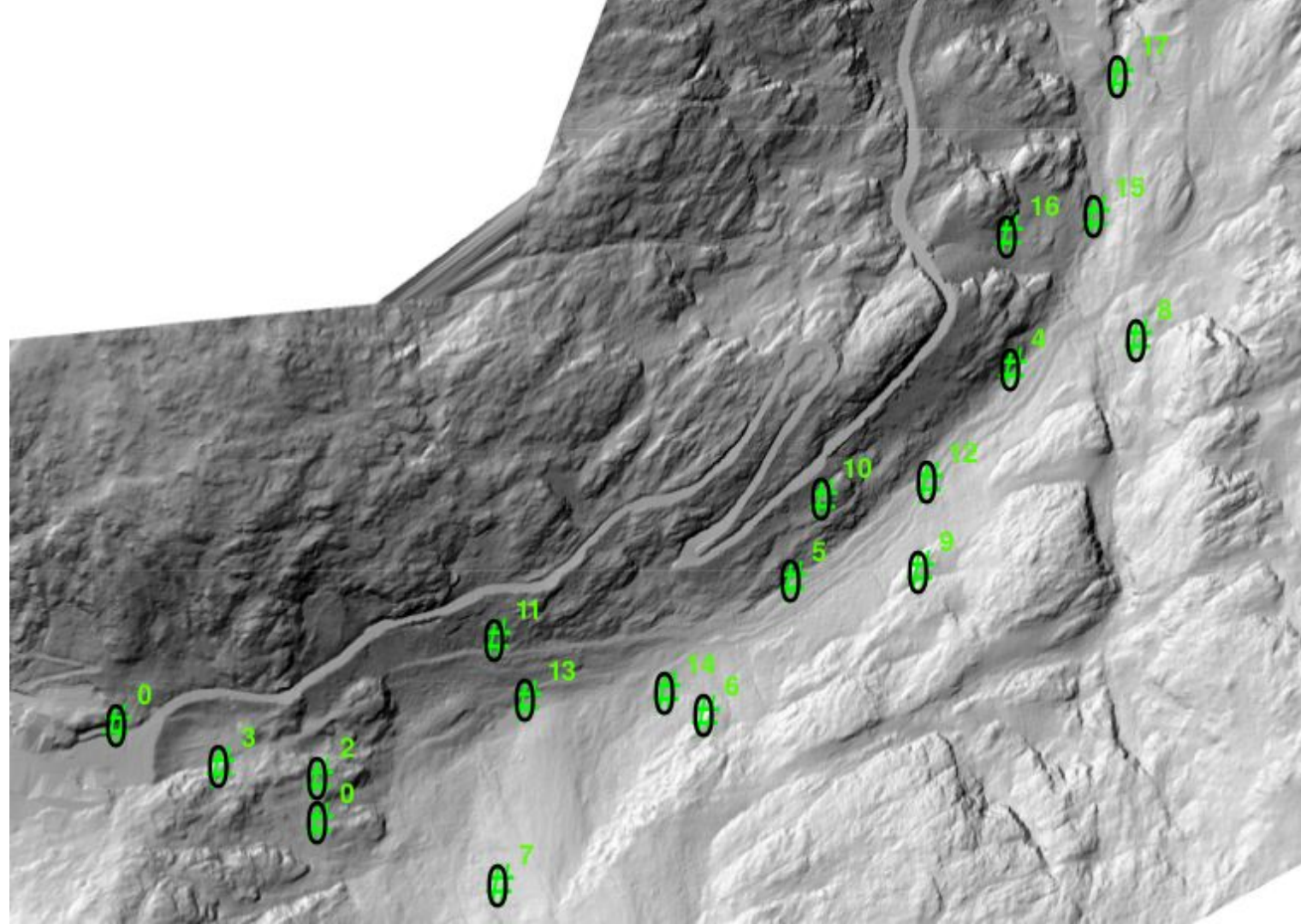
**Radiação solar (w/m2)**





# Pergunta 4

- As diferenças observadas podem ser relacionadas à localização das estações?
  - **Acreditamos que sim**, pois estas estão dispostas em **áreas com características distintas** (exposição solar, vento, altitude)



# Pergunta 5

- Quais os aspectos limitadores considerados no desenvolvimento do projeto Sensorscope?
  - **Quantidade de dados, custo** de leituras e de equipamentos, **acesso** em tempo real, **feedback** estado do sistema

# Pergunta 6

- Como foi a coleta, disponibilização e análise dos dados?
  - **Teste em laboratório da RSSF, implementação em campo**
  - **Servidores próprios e site do projeto**
  - **Análise estatística básica**

# Pergunta 7

- O uso de bancos de dados **NoSQL** apresenta vantagens neste projeto?
- **Sim** → **volume** de dados, **acesso** em tempo real + fácil **distribuição do BD** + facilidade de **criação de novos campos**
- **MapReduce** em projetos futuros

# Conclusões

- **Entendimento e aplicação do ciclo de vida dos dados**
- **Conhecimento no uso das ferramentas**
- **Entendimento do problema** e resposta a problemas da pesquisa
- **NoSQL** se mostrou como uma **alternativa interessante para o dataset**