



# DESENVOLVIMENTO DE UM PROGRAMA EM LINGUAGEM PYTHON PARA O CÁLCULO E ANÁLISE DE COEFICIENTES DE PICO DE CONSUMO HÍDRICO

Palavras-Chave: Engenharia hidráulica; Python; Sistemas de abastecimento.

Autores/as:

Kauan Polli de Oliveira [FECFAU - UNICAMP]

Prof.<sup>a</sup> Dr.<sup>a</sup> André Luis Sotero Salustiano Martim (orientador/a) [FECFAU - UNICAMP]

## INTRODUÇÃO:

O dimensionamento de um sistema de abastecimento pressupõe conhecimento a respeito do consumo de água da região a ser atendida. É sabido que este consumo não é constante ao longo do tempo. Segundo Balacco et al. (2017), as variações de consumo podem ser representadas pelos coeficientes de pico ( $C_p$ ). No Brasil, definem-se  $K_1$  (coeficiente de pico diário) e  $K_2$  (coeficiente de pico horário) como sendo os coeficientes de pico. Em posse dos dados de consumo hídrico de uma determinada região, os coeficientes  $K_1$  e  $K_2$  são obtidos através da aplicação das Equações 01 e 02, em que a Equação 02 faz referência ao dia de maior consumo horário (PORTO, 2001):

$$K_1 = \frac{\text{Maior consumo diário no ano}}{\text{Consumo diário médio no ano}} \quad (01)$$

$$K_2 = \frac{\text{Maior consumo horário}}{\text{Consumo horário médio no mesmo dia}} \quad (02)$$

A NBR 12218/2017 recomenda a utilização de 1,2 para  $K_1$  e de 1,5 para  $K_2$ . Estes valores, entretanto, são pouco explorados e carecem de um acervo de pesquisa plural e assertivo para que possam ser criticados. Deste modo, é essencial novas investigações sejam realizadas, a fim de verificar a existência de divergências entre os valores reais de  $K_1$  e  $K_2$  e as recomendações normativas, de que maneira a e elencar possíveis impactos nas redes de distribuição de água.

Uma maneira de avaliar os coeficientes de pico é aplicando as Equações 01 e 02 tendo como base os dados de consumo hídrico de uma determinada região. Estes dados podem ser fornecidos pelas instituições responsáveis pelo abastecimento de água em planilhas eletrônicas. Porém, estas planilhas podem possuir um custo de processamento grande: um ano de estudo pode gerar uma planilha eletrônica com aproximadamente 20000 linhas.

Para amenizar este problema, podem ser aplicados ambientes de programação, com o desígnio de otimizar o processamento e a capacidade de tratamento de dados (MCKINNEY, 2018). Dentre as diversas linguagens existentes, a linguagem Python, segundo Massaron e Mueller (2015), é capaz de realizar o multiprocessamento em grandes conjuntos de dados, reduzindo o tempo necessário para analisá-los.

Dada a possibilidade de utilizar linguagens de programação em um processo de automatização de análise de dados e a necessidade de novos estudos que busquem verificar os coeficientes de pico a partir de consumo medidos ao longo do tempo, é identificada uma possibilidade de desenvolvimento de pesquisa, haja visto que a existência de um *script* em linguagem Python pode acarretar aos laboratórios de pesquisa um aumento na produtividade das análises de seus bancos de dados. Deste modo, o presente estudo teve como objetivo analisar o uso e aplicação de desenvolvimento de um algoritmo na linguagem de programação Python capaz de automatizar o cálculo dos coeficientes de pico  $K_1$  e  $K_2$ ,

## METODOLOGIA:

A Figura 01 exibe o fluxograma de funcionamento do *script* desenvolvido. Primeiro, importa-se as bibliotecas necessárias (Pandas e Numpy). Depois, um bloco é criado para realizar a leitura de uma planilha em XLSX., que deve ser previamente processada, de maneira a conter três colunas: uma com os dados de data da medição feita; outra com o valor do dado de consumo (l/s); e a última com referência sobre a qualidade da medição (“Good” para informação válida e “Bad” para não válida). As colunas devem estar necessariamente nesta ordem e as medições não devem exceder o período de um ano. A Figura 02 exibe um exemplo de planilha de entrada.

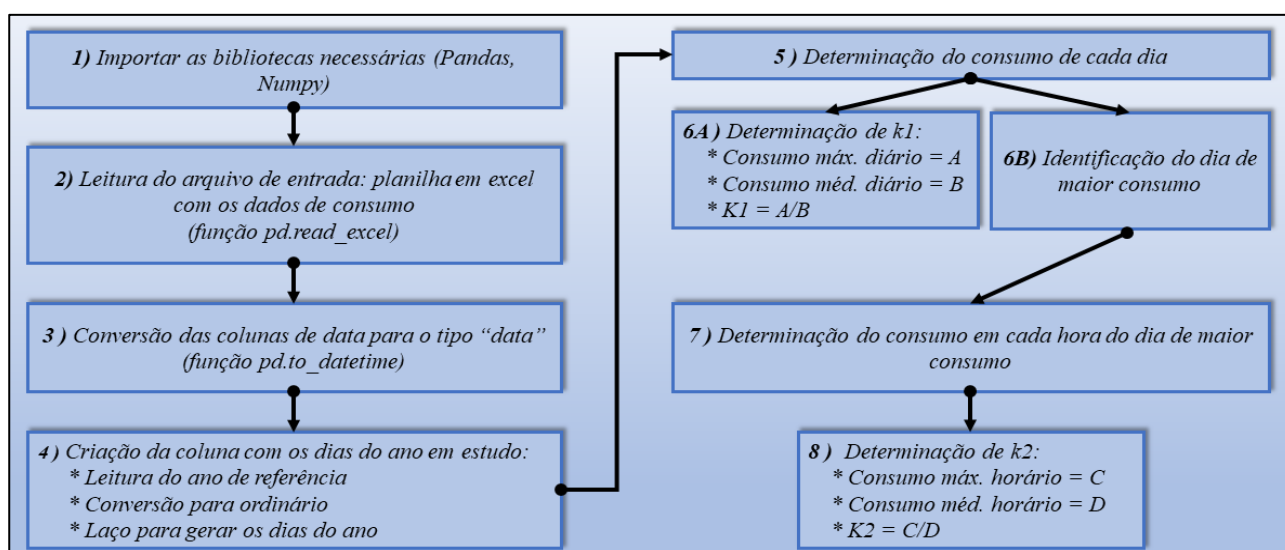


Figura 01 – Fluxograma de funcionamento do script desenvolvido em linguagem Python

	A	B	C
1	Data	Valor (l/s)	Qualidade
2	01-jan-20 00:21:00	3,49232626	Good
3	01-jan-20 00:24:00	4,703316689	Good
4	01-jan-20 00:54:00	3,635385513	Good
5	01-jan-20 01:24:00	4,216302872	Good
6	01-jan-20 01:54:00	3,635163784	Good

Figura 02 – Exemplo da disposição das colunas da planilha de entrada (a ser processada)

Realizada a leitura do arquivo, os dados da primeira coluna “Data” são convertidos “datetime”. O *script* então reconhece o ano em que foram feitas as medições. Em seguida, é criada uma coluna em seu *dataframe* com os dias e meses deste ano, relacionando cada dia ao seu respectivo consumo. Organizando-se o consumo em cada dia, é possível a determinação dos consumos máximos e médio diário. São aplicadas as Equações 01 e 02, obtendo-se os valores para K1 e K2. Por fim, o *script* gera uma “planilha de saída”, ilustrada na Figura 03, que possibilita verificar o consumo diário e o consumo horário do dia de maior consumo.

Data	Valor	Qualidade	Data_analise	Soma Inst	Medi coes	VazãoMe diaDiaria	Consumo DiarioTota	Hora	Somalnst Hora	Medicoes Hora	VazãoMedia Horaria	ConsumoH orarioTotal
2019-01-01 06:03:00	2,045187	Good	2019-01-11 00:00:00	310,5	63	4,928995	425,86516	10	6,98069	1	6,980690002	25,130484
2019-01-01 06:33:00	2,016404	Good	2019-01-12 00:00:00	242,5	52	4,663421	402,91961	11	14,00539	2	7,002695084	25,209702
2019-01-01 07:03:00	2,402766	Good	2019-01-13 00:00:00	263,9	58	4,549162	393,04763	12	13,09884	2	6,549419641	23,577911
2019-01-01 07:33:00	2,636101	Good	2019-01-14 00:00:00	234,7	53	4,427943	382,57426	13	12,83744	2	6,418719769	23,107391
2019-01-01 08:03:00	2,709604	Good	2019-01-15 00:00:00	221,4	52	4,257435	367,84234	14	28,19674	4	7,049183846	25,377062

Figura 03 – Exemplo da planilha de saída gerada pelo script

## RESULTADOS E DISCUSSÃO:

O principal resultado desta pesquisa é um código capaz de realizar o processamento de uma planilha eletrônica e retornar os valores correspondentes para K1 e K2. Este código foi utilizado para o processamento dos dados de consumo hídrico do bairro Jacaré (Cabreúva, SP) nos anos de 2019 e 2020. Para o primeiro ano, foram processadas 21030 linhas em 58 segundos. Para o ano de 2020, foram processadas 13521 linhas em 26 segundos. Os dados referentes ao tempo de processamento são relativos a um computador com memória RAM de 6GB e um processador de 2.3GHz. As Figuras 04 e 05 exibem as linhas de saída que o *script* retorna ao ser executado no ambiente de desenvolvimento IDLE.

```

Processando o valor de K1...
Valor de K1 processado.

Processando o valor de K2...
Valor de K2 processado.

Criando planilha de saida...

O valor de k1 é: 1.4256179990764284
O valor de k2 é: 1.2613318191079679
O dia de maior consumo é: 9 / 3 / 2019

Processo finalizado.
>>>

```

```

Processando o valor de K1...
Valor de K1 processado.

Processando o valor de K2...
Valor de K2 processado.

Criando planilha de saida...

O valor de k1 é: 1.5927901161826927
O valor de k2 é: 1.3095046967637258
O dia de maior consumo é: 27 / 3 / 2020

Processo finalizado.
>>>

```

Oliveira (2020) fez uso dos mesmos dados de entrada e calculou de maneira não automática os valores de coeficientes de pico para esta mesma região nestes mesmos anos. A Tabela 01 resume a igualdade entre resultados encontradas e apresenta características sobre o tempo de processamento.

		Oliveira (2020)	Script	Número de linhas processadas	Tempo de processamento (s)	Tempo de processamento a cada 1000 linhas
2019	K1	1,43	1,43	21030	58	2,76 segundos a cada 1000 linhas
	K2	1,26	1,26			
2020	K1	1,59	1,59	13521	26	1,92 segundos a cada 1000 linhas
	K2	1,31	1,31			

Tabela 01 – Valores de coeficientes de pico obtidos por Oliveira (2020) e pelo código desenvolvido e informações sobre tempo de processamento

É possível observar pela Tabela 01 que os resultados obtidos através do *script* são iguais aos obtidos por Oliveira (2020), atestando a funcionalidade do código para estes dois casos. Além disso, no ano de 2020, foi analisado um número aproximadamente 35% menor de linhas, o que levou a um tempo necessário para processar 1000 linhas cerca de 30% menor. Este fato indica que, para um ano em que se tenha um número consideravelmente maior de medições, o tempo de processamento pode aumentar de maneira não desprezível, apesar da assertividade do *script*.

## CONCLUSÕES:

Este trabalho teve como objetivo principal desenvolver um algoritmo na linguagem de programação Python capaz de automatizar o cálculo dos coeficientes de pico  $K_1$  e  $K_2$ . Os valores apresentados pelo programa foram comparados com aqueles apresentados por um estudo anterior realizado sobre o bairro Jacaré na cidade de Cabreúva, interior de São Paulo, obtidos através de cálculo não automatizado, e que utilizaram os mesmos dados de consumo hídrico como base. O *script* foi capaz de processar as planilhas e retornar valores consistentes para os coeficientes de pico, identificando o dia de maior consumo para cada ano.

O tempo de processamento apresentado pelo *script* é proporcional ao número de medições feitas. A causa principal é que o processamento utiliza blocos conhecidos como “laços”, que percorrem toda a planilha, linha por linha. Este fato indica a possibilidade de que futuras pesquisas se concentrem na otimização do código, por meio da busca de comandos específicos em outras bibliotecas da linguagem Python que eliminem a necessidade da utilização de laços.

Apesar do funcionamento verificado para estes dois casos, é importante que sejam feitos mais testes, verificando sobre a possibilidade de erros nos apontamentos feitos. Também é válida a produção de um manual do usuário, facilitando o uso do *script*.

Sendo notória a importância de estudar e entender melhor as características de consumo hídrico para melhor eficiência dos sistemas de abastecimento público, conclui-se que esta pesquisa contribui com uma ferramenta que, embora necessite de otimização, tem potencial de aumentar a produtividade de institutos e laboratórios de pesquisa, em se tratando da avaliação de dados de consumo hídrico para a aferição dos valores de coeficiente de pico.

## AGRADECIMENTOS

Agradeço ao CNPq pela bolsa de Iniciação Científica, à PRP-Unicamp pela oportunidade de trabalhar no projeto PIBIC, à SABESP e à SANASA pela disponibilização dos dados e à UNICAMP por fornecer toda a estrutura imprescindível para a realização desta pesquisa.

## BIBLIOGRAFIA

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. (2017). **NBR 12218: Projeto de rede de distribuição de água para abastecimento público - Procedimento**. Rio de Janeiro.

BALACCO, G.; CARBONARA, A.; GIOIA, A.; IACOBELLIS, V.; PICCINNI, A.F. *Evaluation of Peak Water Demand Factors in Puglia (Southern Italy)*, **Water**,9(2), 96. Italy, 2017.

MASSARON, L.; MUELLER, P. J. (2015). *Python for data science*. 1. ed. **New Jersey: John Wiley & Sons**.

MCKINNEY, W. (2018). **Python para análise de dados: Tratamento de dados com Pandas, NumPY e IPython**. 1. ed. São Paulo: Editora Novatec.

OLIVEIRA, K. P. (2020). **Avaliação dos coeficientes de consume de sistemas de abastecimento de água na região de Campinas**. Relatório final de iniciação científica. PIBIC quota 2019/2020. Campinas.

PORTO, R. M. (2001). **Hidráulica básica**. 2. ed. EESC-USP São Carlos- SP, 519 p.

TRIFUNOVIC, N. (2006). **Introduction to urban water distribution**. Unesco-IHE Lecture Note Series. CRC Press, 528 p.