

## **Trabalho de Final de Curso**

Apresentação do plano de pesquisa

*Localização ótima de estações de monitoramento de qualidade da água em redes de distribuição baseada em algoritmos genéticos*

**Aluno: Hugo Sussumo Nunes Hirose RA 061475**  
Supervisor. Dr. Edevar Luvizotto Junior

## Sumário

1	DEFINIÇÃO DO PLANO (Síntese)	4
2	Fundamentação Teórica	4
2.1	Introdução	4
2.2	Estações de monitoramento	6
2.3	Otimização empregando algoritmo genético	7
2.3.1	Aspectos gerais do algoritmo genético (AG)	7
2.4	Simulação hidráulica e cobertura da demanda	7
3	Método	9
3.1	Objeto	9
3.2	Etapas	9
3.2.1	Simulação hidráulica	9
3.2.2	Obtenção da matriz de fração de água	10
3.2.3	Obtenção da matriz de cobertura	11
3.2.4	Ajuste inicial	13
3.3	Função de ajuste (maximização)	13
3.3.1	Ampliação da escala de ajuste	15
3.4	Passos do AG	15
3.4.1	Geração aleatória da população inicial	15
3.4.2	Seleção dos indivíduos (pais) para a reprodução (crossover)	16
3.4.3	Reprodução (Crossover)	16
3.4.4	Mutação	18
3.5	Escolha das estações de monitoramento	18
3.5.1	Construção da matriz com as melhores soluções.	18
3.5.2	Escolha da solução ótima	20
4	Contribuição esperada	20
5	Referências bibliográficas	21

## Lista de Figuras

1	Fluxograma do método proposto . . . . .	8
2	Rede hipotética empreada para apresentação da proposta . . . . .	10

## Lista de Tabelas

1	Matriz de fração de água (senário 1) . . . . .	12
2	Matriz de cobertura . . . . .	14
3	Obtenção de Q . . . . .	14
4	População inicial gerada aleatoriamente . . . . .	16
5	Seleção de Pais por Torneio . . . . .	17
6	Seleção de Pais por Torneio . . . . .	18
7	Efeito da aplicação do operador mutação . . . . .	19
8	Vetores das melhores soluções para cenário 1 . . . . .	20

## 1 DEFINIÇÃO DO PLANO (Síntese)

O monitoramento da água potável transportada por uma rede de distribuição é essencial para garantir a salvaguarda da saúde humana e, para o cumprimento dos padrões de qualidade pelas normas locais e internacionais. A qualidade da água em uma rede de distribuição é avaliada através de amostras em locais que são representativas de toda a rede. Diferentes ferramentas baseadas em técnicas de otimização podem ser empregadas para a identificação das estações de monitoramento da qualidade da água. Este trabalho propõe uma forma baseada em algoritmo genético (AG). As etapas envolvidas no método a ser desenvolvido serão delineadas e, aplicações em redes hipotéticas serão desenvolvidas. A viabilidade do emprego da proposta será avaliada através de caso obtido na literatura

## 2 Fundamentação Teórica

### 2.1 Introdução

A qualidade da água pode sofrer deteriorização durante sua distribuição ao consumidor. Muitos fatores, que podem ser externos ou internos, causam a deterioração da qualidade da água entre o tratamento e o consumo. Algumas das causas principais são: a fonte da água, os processos de tratamento, a operação do sistema de transporte e a condição da rede de distribuição e dos reservatórios de armazenamento.

Para obtenção de um panorama geral sobre a qualidade da água em uma rede de distribuição, locais de amostragem devem ser definidos para monitorar os parâmetros de qualidade. A conveniência e representatividade espacial são os dois principais fatores ao se selecionar os locais de amostragem [1,2].

Após os locais de amostragem serem identificados, o acompanhamento regular da qualidade da água nesses locais, além do monitoramento da qualidade da fonte, é necessário. O monitoramento deve incluir parâmetros suficientes para indicar todos os problemas de qualidade e devem ser realizados em locais apropriados à partir das fontes de abastecimento. O programa de monitoramento deve incluir protocolos para frequência de amostragem e metodologia de análise e, deve ser projetado para estabelecer valores de referência para indicar as tendências de curto e longo prazo. Esse acompanhamento pode servir como um mecanismo para detectar a ocorrência de problemas de contaminação da água em seus primeiros estágios [3].

A prática atual de amostragem de água baseia-se em amostras obtidas em locais que são fáceis de alcançar. Os locais típicos utilizados para coleta de amostras de água incluem: hidrantes de incêndio, reservatórios de armazenamento, estações de bombeamento, edifícios comerciais, edifícios públicos e residências privadas, ou seja, não existem diretrizes sobre como localizar estações de amostragem.

Algumas metodologias foram desenvolvidas, recentemente, para localizar essas estações, com base em métodos científicos. Lee e Deininger [2] desenvolveram uma abordagem baseada no conceito de “cobertura da demanda”. O termo “cobertura da demanda” foi empregado para representar o percentual de demanda da rede que é controlado por uma estação de monitoramento específica. O objetivo desta metodologia foi alocar estações de monitoramento que permitisse obter o máximo de informação sobre a condição da qualidade da água em uma rede de distribuição. A solução sugerida por Lee e Deininger [2] foi baseada no aspecto geral de como os parâmetros de qualidade da água diminuem com o tempo e com a distância da fonte. Ou seja, se a qualidade da água em um nó da amostra é bom, então deve ser bom em um nó imediatamente a montante. O termo “nó coberto” foi utilizado para indicar que a qualidade da água em um determinado nó pode ser inferida pela água amostrada em alguns nós a jusante. Lee e Deininger [2] usaram as informações obtidas da análise hidráulica da rede, para identificar o modo como a qualidade da água da rede pode ser avaliada através da instalação de algumas estações de amostragem. As informações obtidas a partir destas vias, foram descritas em termos de uma matriz denominada “matriz de fração de água”, sendo convertido em um problema de programação inteira, sob um critério de limite de cobertura escolhido. Por esse método, o menor nível de conhecimento ocorrerá quando apenas uma fração muito pequena da água passar pelo nó, que foi chamado de “nó de fração mínima”. No entanto, para redes muito grandes, este método se tornou de difícil tratamento devido à dimensão do problema.

Kumar et al. [4] propuseram algumas melhorias no trabalho de Lee e Deininger [2] para resolver este problema de dimensão. Eles utilizaram a mesma *matriz de cobertura* desenvolvida por Lee e Deininger. Depois do cálculo da direção de escoamento em cada tubo da rede, os nós são numeradas em ordem crescente de vazões e, então, a estação de monitoramento, com máxima cobertura de nós a montante, é selecionada. Em seguida, a linha correspondente a estação selecionada é excluída da matriz de cobertura. As estações de monitorização subsequente são selecionadas repetindo o mesmo processo para o número de vezes que as estações de monitoramento forem requeridas, a partir da matriz de cobertura anterior. O método proposto por Kumar et al. [4] é mais simples do que a de Lee e Deininger [2], quanto à construção da matriz de cobertura, no entanto, é extenso o tempo de processamento computacional para otimizar as localizações de monitoramento, mesmo com os cálculos matemáticos simples, para redes de grande porte.

Baseado nos conceitos de Lee e Deininger [2], Kessler et al. [5] desenvolveram uma metodologia que é capaz de localizar estações de monitoramento ótimas para avaliar a qualidade da água em uma rede de distribuição, para uma situação de injeção acidental de contaminantes. Kessler et al. [5], definiram “nível de serviço”, como a quantidade máxima permitida de vazão de água através de um determinado nó antes da detecção dos contaminantes em relação ao tempo de detecção. Após a simulação hidráulica da rede utilizando simulação de período

extensivo, uma rede auxiliar é desenvolvida sob a forma de grafos de ligação, composto de nós e arcos direcionados, de tal forma que o comprimento do arco representa o tempo de viagem entre os nós. Todos os caminhos mais curtos foram calculadas utilizando a rede auxiliar e uma matriz de contaminação foi construída com coeficientes 0-1.

O estudo atual envolve uma extensão do modelo desenvolvido por Lee e Deininger [2] para ajudar na identificação de estações de monitoramento da qualidade de em uma rede de distribuição utilizando algoritmo genético, na forma sugerida por Al-Zarami e Moied[13], que se baseia nos conceitos anteriores aplicados na forma que se descreve à seguir.

## 2.2 Estações de monitoramento

O objetivo da investigação é determinar a localização adequada das estações de monitoramento de qualidade que são representativas de toda a rede. Assim, a qualidade da água analisada nessas estações irá representar a qualidade de toda a rede. Para uma rede de distribuição de água, a dimensão do domínio espacial, assim como o número de estações de monitoramento necessárias aumenta exponencialmente com um número crescente de nós. Assim, o número de possíveis soluções, assumindo que qualquer nó dentro da rede de distribuição é um candidato a estação de amostragem, é definido como:

$$NFS = N^S \quad (1)$$

sendo,  $NFS$  = Número de soluções Factiveis,  $N$  = é o número de nós da rede e  $S$  = é o número de estações de monitoramento necessárias.

Por exemplo, para localizar quatro estações de monitoramento em uma rede de 100 nós, o número de soluções factíveis é da ordem de  $100^4$ . No entanto, desses 100 milhões de soluções factíveis, só há uma ideal. Para localizar essa solução ótima, dentre um grande número de soluções factíveis, um poderoso algoritmo é necessário. Um algoritmo genético (AG) é um algoritmo que pode ser capaz de resolver este problema. Estes serão brevemente descritos adiante. Os AG's já foram empregados com sucesso em outras áreas que envolviam redes de abastecimento de água. Meier et al. [7], empregaram os AG's para otimizar os locais de medição de vazão necessários para a calibração do modelo de rede. Savic e Walters [11] ilustram o uso de AG's para o seleção de diâmetros de tubulação que forneciam o menor custo de implantação, respeitando as condições normativas.

## 2.3 Otimização empregando algoritmo genético

### 2.3.1 Aspectos gerais do algoritmo genético (AG)

Um rotina, baseada em AG, inicia com uma população de indivíduos representados por cadeias codificadas (seus cromossomos) gerados aleatoriamente e, que representam possíveis valores para as variáveis que apontam para uma localização no domínio da solução. As variáveis codificadas nesses cromossomos são chamadas de "genes" ou "alelos". A evolução natural ocorre nos cromossomos, que carregam informações hereditárias, sob a forma de genes [6]. Os valores das variáveis de decisão (genes) geralmente são expressos na forma de cadeias na maioria das vezes usando representação binária [7]. Neste estudo valores inteiros serão utilizados, para representar o número de nós na rede de distribuição, o que diminui o tamanho das cadeias das soluções factíveis (cromossomo).

Da população inicial, os indivíduos mais aptos (emprega-se o valor de sua função de aptidão ou ajuste) são selecionados para passar sua "informação genética" para a próxima geração. Esta operação é chamada de "seleção", que se assemelha a sobrevivência do mais apto no processo de seleção natural. Há muitas formas para a seleção, no entanto, todos compartilham o objetivo comum que é substituir os indivíduos menos aptos da população, para fazer avançar a operação de busca. Após a seleção, a população será, em média, mais apta do que era antes da seleção.

A seleção é então acompanhado por uma operação chamada de "crossover", que gera novas cadeias, que contêm propriedades que as distinguem das ancestrais das quais elas foram geradas. Em alguns casos os novos indivíduos terão menor adequação, em outros maior e, em uma certa porcentagem, irão se assemelhar a seus pais, com valores de aptidão iguais a estes. Como o crossover simplesmente recombina cadeias existentes em novas gerações poderá levar à características constantes na população. É possível que algumas sequências desejáveis que não foram incluídas na primeira geração (gerada aleatoriamente), ou que tenham sido perdidas por indivíduos que não possuíam as qualidades desejáveis e desapareceram da população após gerações. Uma operação denominada "mutação" é utilizada ocasionalmente para alterar uma cadeia (cromossomo), recuperando qualidades desejáveis ou criando novas qualidades nas cadeias geradas.

A Figura 1 mostra as etapas necessárias envolvidas, aplicando um algoritmo genético para localização ótima de estações de monitoramento de qualidade de água.

## 2.4 Simulação hidráulica e cobertura da demanda

Uma etapa prévia, antes da aplicação dos AG's é a simulação hidráulica da rede de distribuição de água. A simulação hidráulica é finalizada quando todas vazões e os sentidos de escoamento

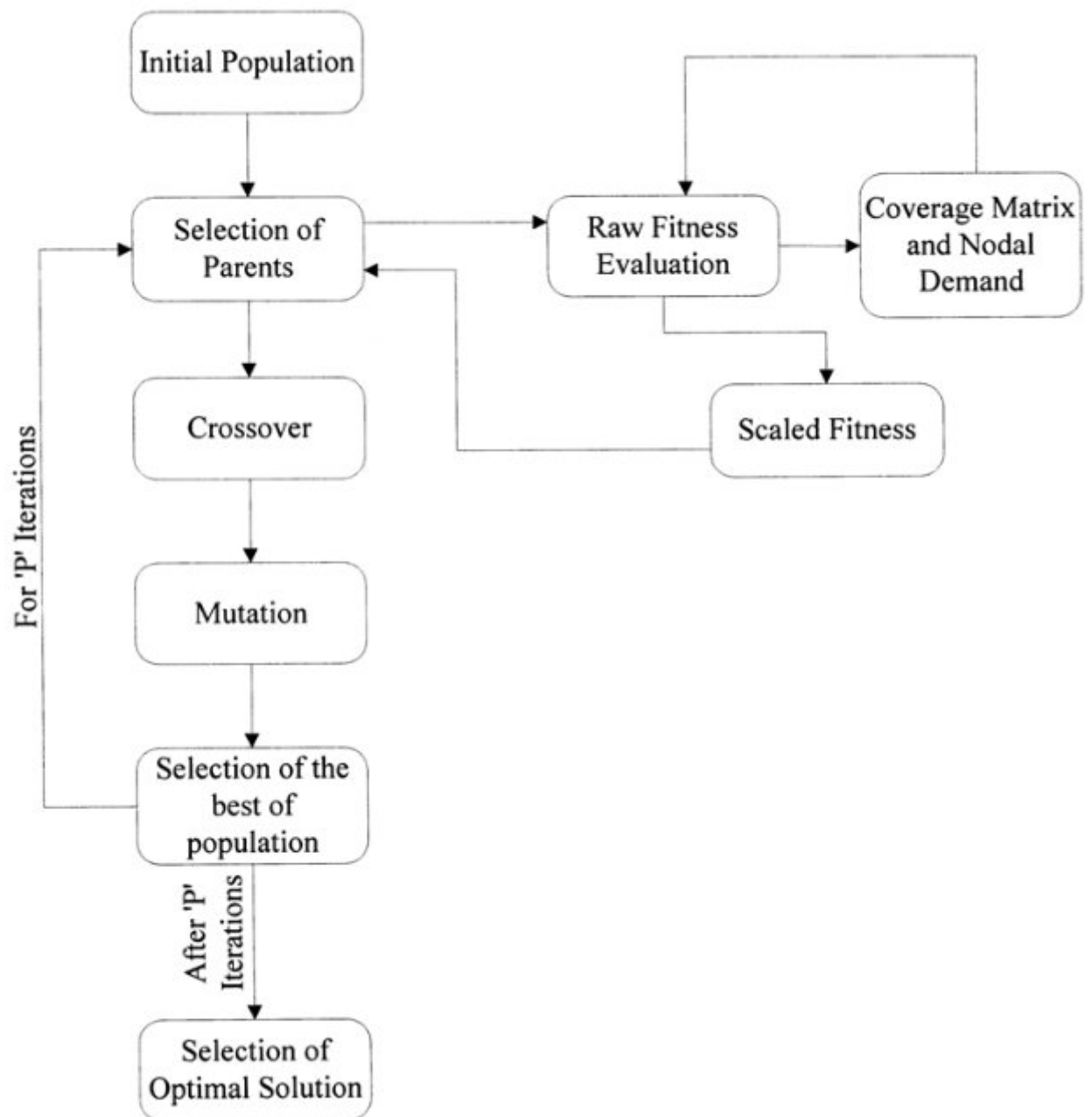


Fig. 1: Fluxograma do método proposto



nos elementos junto com as demandas em cada nó são conhecidas, de tal forma que todas as perdas de carga em cada circuito fechado somadas são zero, a fim de satisfazer a equação de balanço de massa e de energia.

Uma vez que as características hidráulicas da rede são conhecidas, a lógica de que a qualidade da água em um nó de montante é melhor do que a dos nós a jusante é aplicada. Ela vem do fato de que a qualidade da água se deteriora com o passar do tempo uma vez que esta estará fluindo para longe da fonte de abastecimento. Matrizes podem, ser escritas com o fim de analisar a cobertura máxima da rede de distribuição. Estas matrizes são chamadas de " Matriz de Cobertura de Demanda". Assim, a demanda, abrangida por cada nó a jusante é calculada considerando as demandas de todos os nós a montante na rota para esse nó. Como a direção do fluxo nos elementos se alteram com a mudança do padrão de demanda, é necessário construir uma matriz de cobertura de demanda para cada cenário de demanda separadamente. O conceito acima serão apresentados mais adiante, quando a metodologia desenvolvida é aplicado sobre uma rede hipotética de distribuição de água, justificando a sua adoção neste trabalho

## **3 Método**

### **3.1 Objeto**

A pesquisa, ora proposta, tem por objeto o desenvolvimento, de uma rotina computacional, baseada nos elementos descritos nos itens da fundamentação teórica. Para tanto será seguido o método descrito nos itens seguintes, que define a função do modelo hidráulico (baseado no EPANET 2.0 e EPANET Toolkit) e na rotina baseada em algoritmos genéticos conforme descrito na sequência.

### **3.2 Etapas**

#### **3.2.1 Simulação hidráulica**

Para ilustrar a aplicação do método, uma rede de distribuição de água hipotética constituída de 15 nós, 23 elementos, e 3 fontes de abastecimento (A, B e C) é proposta. Redes de distribuição de água são compostas por elementos interligados tais como; tubos, bombas e reservatórios para transmitir a partir de uma ou mais fonte, água tratada aos consumidores espalhados por uma área normalmente de grandes dimensões. Para esta rede hipotética, a simulação hidráulica para um cenário específico será determinada empregando o programa EPANET [9]. Assim, quantidades, direções de escoamento e demandas nodais serão determinadas. Essas quantidades e direções de escoamento nos elementos e as demandas em cada nó são mostrados

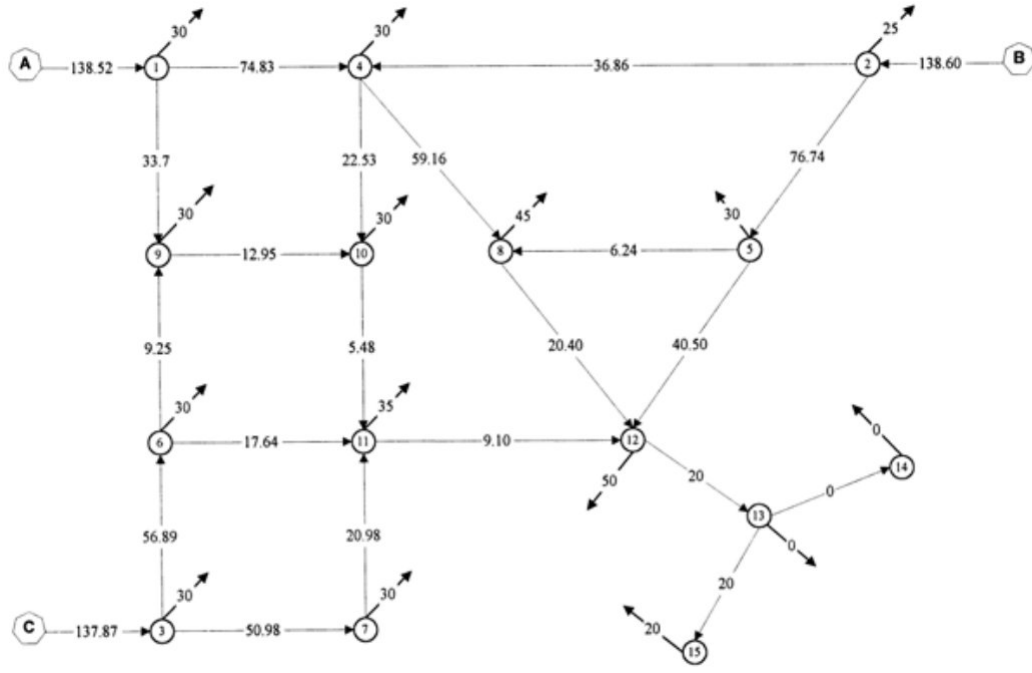


Fig. 2: Rede hipotética empregada para apresentação da proposta

na Figura 2 para um cenário hipotético. A oferta total para a rede a partir da fonte é definido por 415 unidades de vazão.

Os nós precisam ser identificados a fim de determinar o encaminhamento da água na rede de distribuição. Considerando um nó de jusante, a avaliação é feita no sentido montante para identificar os nós, que contribuem para o abastecimento de água para o nó considerado (a jusante). Para a rede hipotética, é bastante óbvio que o nó 9 recebe água do nó 1, o nó 5 recebe o abastecimento de água a partir do nó 2, e o nó 4 recebe água a partir dos nós 1 e 2. Da mesma forma, o nó 10 recebe água fornecida pelos nós 4 e 9. Assim, todos nós a jusante são avaliados para determinar o total de nós contribuintes da rede.

### 3.2.2 Obtenção da matriz de fração de água

Uma vez que os nós são identificados, uma matriz de fração de água é construída. Para isso, as frações do total de água recebidas em um determinado nó (a jusante) é obtida através dos nós que contribuem a montante. Por exemplo, se  $(w_{nk} + w_{nl} + w_{nm})$  é o total de água fornecida para um nó a jusante  $n$  dos nós  $k, l, m$  (a montante), as frações  $f_n$  são calculadas como [4]:

$$f_{nk} = \frac{w_{nk}}{w_{nk} + w_{nl} + w_{nm}} \quad (2)$$

$$f_{nl} = \frac{w_{nl}}{w_{nk} + w_{nl} + w_{nm}} \quad (3)$$

$$f_{nm} = \frac{w_{nm}}{w_{nk} + w_{nl} + w_{nm}} \quad (4)$$

sendo;

$f_{nk}$  a fração da água recebida pelo nó  $n$  do nó  $k$ ,  $f_{nl}$  a fração da água recebida pelo nó  $n$  do nó  $l$ , e  $f_{nm}$  a fração da água recebida pelo nó  $n$  do nó  $m$ .

Desde que os nós  $k$ ,  $l$  e  $m$  também estejam recebendo contribuições dos nós a montante 1,2,3,4 ..... , a fração de água de todos esses nós a montante do nó  $n$  (de monitoramento) será o vetor,  $f_n$  :

$$f_n = (f_{n1}, f_{n2}, f_{n3}, \dots, f_{nk}, f_{nl}, f_{nm}, \dots, f_{n(n-1)}, f_{nn}) \quad (5)$$

os elementos do vetor  $f_n$  podem ser calculados como;

$$f_n = [(f_{nk} * f_k) + (f_{nl} * f_l) + (f_{nm} * f_m)] \quad (6)$$

e a matriz de cobertura é representada por;

$$[f_n] = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1n} \\ & f_{22} & \dots & f_{2n} \\ & & \vdots & \vdots \\ & & & f_{nn} \end{bmatrix} \quad (7)$$

Uma vez que um nó sempre se cobre, todas as entradas são definidas  $f_{nn}$  igual a '1' na matriz de fração de água. A Tabela 1 representa a fração de água calculada para a rede hipotética do Cenário 1.

### 3.2.3 Obtenção da matriz de cobertura

Após a construção da matriz da fração de água, um critério de cobertura é estabelecido. Se ' $d$ ' é a demanda total de toda a rede e ' $d_i$ ' é a demanda de um determinado nó, então através do monitoramento do nó  $i$ , a fração  $\frac{d_i}{d}$  da rede pode ser coberta. Assim, para cobrir toda a rede, logicamente cada nó da rede deve ser monitorado, mas pode não ser possível devido a razões econômicas. Portanto, a fim de tornar mais fácil o processo de seleção, um critério de cobertura deve ser definido. No estudo a ser desenvolvido, um critério de cobertura de 50% será utilizado. De acordo com este critério de cobertura, os nós de montante que entregam valores superiores ou iguais a 50% da água a um nó a jusante são considerados como cobertos e marcados com "1" na matriz de cobertura de água. Caso contrário, eles são marcados com '0'. O critério de cobertura desempenha um papel no estabelecimento de um equilíbrio entre

Tab. 1: Matriz de fração de água (senário 1)

Fraacionamento de água através dos nós															
Nó	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	0	0,67	0	0	0	0,61	0,79	0,71	0,08	0,19	0,19	0	0,19
2	0	1	0	0,33	1	0	0	0,39	0	0,21	0,03	0,68	0,68	0	0,68
3	0	0	1	0	0	1	1	0	0,22	0,09	0,89	0,12	0,12	0	0,12
4	0	0	0	1	0	0	0	0,9	0	0,64	0,08	0,27	0,27	0	0,27
5	0	0	0	0	1	0	0	0,09	0	0	0	0,61	0,61	0	0,61
6	0	0	0	0	0	1	0	0	0,22	0	0,40	0,05	0,05	0	0,05
7	0	0	0	0	0	0	1	0	0	0	0,48	0,06	0,06	0	0,06
8	0	0	0	0	0	0	0	1	0	0	0	0,29	0,29	0	0,29
9	0	0	0	0	0	0	0	0	1	0,36	0,04	0,01	0,01	0	0,01
10	0	0	0	0	0	0	0	0		1	0,12	0,02	0,02	0	0,02
11	0	0	0	0	0	0	0	0	0	0	1	0,13	0,13	0	0,13
12	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

o número de estações de monitoramento e a cobertura da demanda da rede. Um pequeno valor adotado como critério de cobertura pode sugerir menor número de estações a serem monitoradas a fim de atingir um nível desejado. Por outro lado, uma grande valor adotado para este critério pode sugerir um grande número de estações de monitoramento para garantir que o mesmo nível de demanda coberta seja alcançado.

Segundo o critério de 50% de cobertura, a matriz de fração de água indicada na Tabela 1, é então convertida em uma matriz de cobertura de água, Tabela 2. Uma entrada igual a '1' indica que o nó foi coberto, enquanto uma entrada de '0' indica que o nó particular não é coberto por uma certa estação de monitoramento.

Observa-se, na Tabela 2, que o nó 1 na rede hipotética abrange apenas a si mesmo, o nó 5 abrange os nós 2 e 5, enquanto nó 12 abrange os nós 2, 5 e 12. Análises semelhantes foram feitas para todos os nós da rede. O vetor de demanda da rede é construída com  $d = d_i$ , onde  $d$  é a demanda do nó e  $i$  representa o número de nós na rede de distribuição. O vetor de demanda é o vetor de demandas nodais conhecidas. Valores hipotéticos foram assumidos para a rede hipotética. No entanto, para as redes reais, os valores para o vetor de demanda podem ser determinados a partir de memórias de cálculo ou das contas de água dos consumidores.

Para a rede hipotética, onde  $i=15$ , o vetor de demanda será expresso por;

$$\left( 30 \ 25 \ 30 \ 30 \ 30 \ 30 \ 30 \ 45 \ 30 \ 30 \ 35 \ 50 \ 0 \ 0 \ 20 \right)^T \quad (8)$$

### 3.2.4 Ajuste inicial

Ao contrário de outras técnicas de otimização, a função objetivo é formulada como uma função de ajuste em um algoritmo genético. Esta função de ajuste é obtida com base em uma maximização ou minimização do(s) objetivo(s). O valor de ajuste antes de uma solução possível é chamado de "valor de ajuste bruto", que é determinada pela Equação (9).

$$F_R = 100 - \left[ \frac{Z - Q}{Z} * 100 \right] \quad (9)$$

Sendo o  $F_R$  a função de aptidão bruta,  $Z$  a melhor aptidão definida como ideal, e  $Q$  é a cobertura que, uma solução particular pode alcançar.

Um passo inicial, antes do processo de otimização com AG é definir uma solução ideal ou "melhor" solução. Todas as soluções possíveis geradas durante a otimização são então comparados com esse valor "ideal", para determinar a sua adequação relativa. Quanto mais próximo o valor de ajuste de uma solução factível mais perto do "valor ideal" e, maior é a chance da sua seleção. Por exemplo, o "valor ideal" é definido como igual à oferta total de água na rede de distribuição. Esse valor também pode ser definido como igual ou superior ao total de água disponibilizada para a rede;

$$Z \geq W \quad (10)$$

sendo  $W$  a vazão total que abastece o sistema.

Para a rede hipotética e para o cenário hipotético empregado a oferta total de abastecimento é de 415 unidades de vazão, e o "valor ideal" será adotado 500.

## 3.3 Função de ajuste (maximização)

Com base na função de maximização ( $\max \sum_{i=1}^n d_i * y_i$ ), onde  $d$  é obtido através da Tabela 2 (matriz de cobertura de água). Adota-se em seguida um operador "Oring" as colunas da matriz de cobertura, que são indicados pela solução possível. Este processo de "Oring" é similar a tomar a união das colunas de uma dada linha.  $Q$  é avaliada através da adição do produto do vetor de cobertura resultante pela demanda correspondente ao vetor nodal. Por exemplo, se (3 8 11 15) é a primeira solução possível para a rede hipotética, então o vetor cobertura resultante é construído após "Oring" nas colunas 3, 8, 11 e 15 da matriz de cobertura mostrada na Tabela 2. A Tabela 3 mostra a avaliação da aptidão inicial de uma possível solução para o cenário definido.

Tab. 2: Matriz de cobertura

Fração de água através dos nós															
Nó	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	0	1	0	0	0	1	1	1	0	0	0	0	0
2	0	1	0	0	1	0	0	0	0	0	0	1	1	0	1
3	0	0	1	0	0	1	1	0	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Tab. 3: Obtenção de Q

Colunas obtidas da tabela 2				Veror de cobertura	Vetor de	Q (5)*(6)
Nó 3 (1)	Nó 8 (2)	No 11 (3)	Nó 15 (4)	Resultante após Oring (5)	demanda nodal (6)	
0	1	0	0	1	30	30
0	0	0	1	1	25	25
1	0	1	0	1	30	30
0	0	0	1	1	30	30
0	0	0	1	1	30	30
0	0	0	0	0	30	0
0	0	0	0	0	30	0
0	1	0	0	1	45	45
0	0	0	0	0	30	0
0	0	0	0	0	30	0
0	0	1	0	1	35	35
0	0	0	1	1	50	50
0	0	0	1	1	0	0
0	0	0	0	0	0	0
0	0	0	1	1	20	20
						$\sum Q = 295$

### 3.3.1 Ampliação da escala de ajuste

Embora, o objetivo do AG seja avaliar todo conjunto da população, por vezes é necessário o aumento do espaço de busca para que os membros verdadeiramente aptos possam ser identificados. Isto pode ser conseguido por meio da ampliação dos valores de aptidão, para evitar dominância prematura de um membro da população. Este “membro prematuro” também pode ser denominado como “membro insalubre ou fraco”, que não poderia deixar de ser identificado. Isso é extremamente importante nas populações de ordem superior em que os valores de aptidão média e máxima ficam muito próximos uns dos outros, assim o espaço de pesquisa precisa ser aumentado. Isso pode significar um pequeno atraso na convergência do algoritmo. Uma escala linear é proposta por Goldberg [10] e será adotado neste estudo, de modo que a aptidão média aplicada a escala é mantida igual à média da condição original. Assim, o fator de escala para a aptidão,  $F_s$ , de uma possível solução pode ser definido como:

$$F_s = F_r * C1 + C2 \quad (11)$$

Onde  $F_s$  é a função de aptidão ampliada e  $C1$  e  $C2$  são as constantes para a escala linear.

## 3.4 Passos do AG

### 3.4.1 Geração aleatória da população inicial

Depois de escrever a função de ajuste, o próximo passo é gerar uma população inicial de soluções possíveis. Aqui, a analogia com a natureza é estabelecida através da geração aleatória de um conjunto de soluções chamado de "população" [11]. Cada seqüência de uma solução chamada "Cromossomo", consiste de variáveis de decisão chamados de genes ou alelos. O número de cromossomos (indivíduos) que devem ser gerados depende da decisão sobre o tamanho da população. Como não existe um conjunto de regras para definir o tamanho exato da população; o julgamento subjetivo é empregado. O número acertado de indivíduos da população vem da experiência e conhecimento em relação a implementação do AG. Nessa pesquisa estudo, os números de cromossomos  $X$  (possíveis soluções) serão gerados aleatoriamente na população inicial. Estes cromossomos gerados aleatoriamente servirão como ancestrais (pais) para a seleção. O comprimento dos cromossomos (número de genes) é mantido igual ao número de estações de monitoramento desejado. O "gene" é ajustado para ter um valor inteiro entre 1 e  $n$ , onde  $n$  é o número total de nós na rede de distribuição. No entanto, durante a geração aleatória dos cromossomos, deve-se tomar cuidado para assegurar que todos os genes em um cromossomo tenham valores diferentes. A Tabela 4 mostra a população inicial gerada aleatoriamente com 10 cromossomos para a localização de "quatro" estações de monitoramento.

Tab. 4: População inicial gerada aleatoriamente

Índice do indivíduo (1)	População inicial (2)			
1	12	15	5	3
2	2	15	13	3
3	11	15	5	1
4	3	15	5	10
5	8	4	1	7
6	9	3	1	7
7	10	6	1	13
8	15	12	9	2
9	4	14	1	7
19	9	11	7	2

### 3.4.2 Seleção dos indivíduos (pais) para a reprodução (crossover)

Aleatoriamente os indivíduos da população são selecionados, com base na avaliação de sua aptidão, assim apenas os indivíduos selecionados podem reproduzir e gerar filhos para as gerações subsequentes. Neste estudo, será utilizada a seleção por torneio. Neste método de seleção, toda a população de cromossomos é dividida em subgrupos, de tal forma que cada subgrupo contém dois ou mais cromossomos. Em seguida, o cromossomo mais apto de cada subgrupo é selecionado. Para selecionar os membros restantes da população, os indivíduos da população e população de origem são embaralhados e o mesmo processo é repetido até que todos os membros da nova geração sejam selecionados, de modo que o tamanho da população deva ser constante ao longo da simulação. A Tabela 5 mostra a seleção de pais por torneio para a rede hipotética.

Neste método de seleção, um pai pode ter a chance de ser selecionado mais de uma vez, pelo seu valor de adequação elevado, enquanto o pais menos aptos também recebem, embora com uma menor probabilidade, chance de serem selecionados para a reprodução.

### 3.4.3 Reprodução (Crossover)

É o processo de evolução natural que descreve como indivíduos filhos são gerados à partir de indivíduos pais. Quando os pais “cruzam” uma porção de cada pai é fornecida para produzir os filhos. A concatenação de cromossomos ocorre num determinado ponto e os novos cromossomos (os filhos) são produzidos. Esses cromossomos são diferentes na estrutura de seus pais, mas herdam comportamento característico deles. Um conceito similar é empregado em AG, que é chamado de "operação de crossover".

Em AG, vários métodos estão disponíveis para realizar esta tarefa, gerar filhos de modo que eles possam manter características herdadas de seus pais. Cada usuário AG pode desenvolver sua própria técnica para produzir os filhos à partir dos pais, de acordo com condicionantes



Tab. 5: Seleção de Pais por Torneio

	População fonte	x - Perde √ - Vence	Resultado
1	12 15 5 3 2 15 13 3	x √	2 15 13 3
2	11 15 5 1 3 15 4 10	x √	3 15 4 10
3	8 4 1 7 9 3 1 7	√ x	8 4 1 7
4	10 6 1 13 15 12 9 2	√ x	10 6 1 13
5	4 14 1 7 9 11 7 2	x √	9 11 7 2
6	11 15 5 1 3 18 4 10	x √	3 18 4 10
7	4 14 7 1 7 10 6 1 13	x √	10 6 1 13
8	9 3 1 7 2 15 13 3	x √	2 15 13 3
9	12 15 5 3 15 12 9 2	x √	15 12 9 2
10	9 11 7 2 8 4 1 7	√ x	9 11 7 2

específicos, de tal forma que o objetivo da criação de novos indivíduos à partir dos antigos é alcançado.

Uma vez que os pais são selecionadas pela seleção do tipo torneio, eles irão cruzar e, portanto, produzir descendentes (filhos). Na pesquisa que se propõe, dois pais serão tomados para produzir dois filhos e um único ponto de cruzamento é adotado para esta finalidade. Seleção aleatória do ponto de cruzamento é realizado usando uma função que gera um número inteiro no intervalo de 1 a (S-1) para cada operação de crossover, onde S é o número de estações de monitoramento.

Um filho gerado é denominado "perfeito" se todos os seus genes são diferentes. Para o problema que se investiga um "filho com defeito" é uma solução que contém um determinado nó mais de uma vez, o que implicaria uma duplicação de pontos de monitoramento. No entanto, se qualquer número de dois ou mais genes parece ser o mesmo após o cruzamento, o filho é denominado "defeituoso". Durante a operação de crossover, apenas filhos perfeitos são autorizados a entrar no processo de mutação, no caso de uma criança defeituoso, o pai correspondente (em vez do filho com defeito) é levado para o processo de mutação. O processo de mutação consiste apenas de indivíduos perfeitos. A Tabela 6 mostra a aplicação do crossover para os pais obtidos na Tabela 5.

Tab. 6: Seleção de Pais por Torneio

	Pais Tab. 5 com ponto de Cross	Filhos obtidos por Crossover	Status da seleção	Submeter a Mutação
1	2 15↓ 13 3	2 15 4 10	Perfeito	2 15 4 10
2	3 13↓ 4 10	3 13 13 3	Falho	3 1↓ 4 10
3	8 4 1↓ 7	8 4 1 13	Perfeito	8 4 1 13
4	10 6 11↓ 13	10 6 11 7	Perfeito	10 6 11 7
5	9 11↓ 7 2	9 11 4 10	Perfeito	9 11 4 10
6	3 15↓ 4 10	3 15 7 2	Perfeito	3 15 7 2
7	10↓ 6 1 13	10 15 13 3	Perfeito	10 15 13 3
8	2↓ 15 13 3	2 6 1 13	Perfeito	2 6 1 13
9	15 12↓ 9 2	15 12 7 12	Perfeito	15 12 7 12
10	15 12↓ 7 12	9 11 9 2	Falho	15 12 7 12

O símbolo (↓) representa o ponto onde é feito o crossover

### 3.4.4 Mutação

Após o cruzamento, uma operação denominada "mutação" é aplicada. Como o princípio fundamental do AG é a sua aleatoriedade, é bem possível que a solução da população inicial não seja levada para várias populações adiante. Isto é devido ao fato de que os pais foram obrigados a morrer, e por isso não apareçam novamente. O operador de mutação pode ajudar a recuperar elementos que desapareceram a fazer parte da população. A taxa com que a mutação é aplicada precisa ser muito baixa. Sugere-se estar dentro da faixa de 1-3%, entretanto, estudos realizados por Tate e Smith [12] sugerem uma taxa de mutação mais alta para a codificação não binária das cadeias (cromossomos). Assim, neste estudo, a mutação é realizada a uma taxa de 5%. Durante a mutação, um número real aleatório é gerado no intervalo de 0 a 1, para cada gene do cromossomo. Se o número aleatório gerado é menor que 0,05, o valor correspondente do gene é alterado por um valor aleatório, mas este deve ser diferente dos valores dos genes do restante do cromossomo. A Tabela 7 ilustra a aplicação da mutação para a população obtida na Tabela 6.

## 3.5 Escolha das estações de monitoramento

### 3.5.1 Construção da matriz com as melhores soluções.

Uma vez que a operação de mutação é concluída, o melhor membro da população é selecionado e armazenado em uma matriz "de melhores soluções". Desta forma, uma matriz é obtida, contendo apenas as melhores soluções em cada população. As melhores soluções representam, portanto, as soluções que foram selecionados em cada geração, com base na aptidão de cada um de seus membros.

Para a rede hipotética, uma matriz das melhores soluções é mostrada na Tabela 8 após a

Tab. 7: Efeito da aplicação do operador mutação

Índice do indivíduo	Pais obtidos da após o croosover - TAB. 6.				População pós mutação			
1	2	15	4	10				
No. Randômico	0,753	0,235	0,625	0,859	2	15	4	10
2	3	15	4	10				
No. Randômico	0,545	0,458	<b>0,015</b>	0,075	3	15	<b>6</b>	10
3	8	4	1	13				
No. Randômico	0,458	0,214	0,958	0,145	8	4	1	13
4	10	6	1	7				
No. Randômico	0,329	<b>0,010</b>	0,385	0,495	10	<b>2</b>	1	7
5	9	11	4	10				
No. Randômico	0,215	0,859	0,439	0,218	9	11	4	10
6	3	15	7	2				
No. Randômico	<b>0,033</b>	0,782	0,529	0,125	<b>12</b>	15	7	2
7	10	15	13	3				
No. Randômico	0,354	0,235	0,625	0,8598	10	15	13	3
8	2	6	1	13				
No. Randômico	0,214	0,354	0,852	0,756	2	6	1	13
9	15	12	7	2				
No. Randômico	0,325	<b>0,024</b>	0,665	0,682	15	<b>14</b>	7	2
10	9	11	7	2				
No. Randômico	0,215	0,710	0,257	<b>0,045</b>	9	11	7	<b>5</b>

Os números em negrito na segunda coluna indicam os números gerados aleatoriamente e que satisfazem o critério estabelecido para a mutação. Em negrito na coluna três estão os genes modificados pela mutação.

Tab. 8: Vetores das melhores soluções para cenário 1

Número da População	Vetor de melhor solução	Função de Ajuste
1	3 15 4 10	59,55
2	15 4 10 6	65,55
3	15 4 10 6	68,55
4	8 10 3 15	69,30
5	10 6 15 8	71,85
6	6 15 8 10	72,90
7	15 8 10 6	77,85
8	6 15 8 10	81,30
9	6 15 8 10	81,45
10	11 15 8 10	86,40
11	8 10 11 15	89,10
12	8 10 11 15	84,30
13	8 10 11 15	93,75
14	<b>15 8 10 11</b>	<b>96,90*</b>
15	11 15 8 10	96,43
16	11 15 8 10	95,55
17	15 8 10 11	95,40
18	15 8 10 11	87,90
19	15 8 10 11	87,45
20	15 8 10 11	93,30

\* - Indica o valor da máxima adequação (aplicada ampliação de escala)

execução do procedimento para 20 iterações (gerações), quatro estações de monitoramento, e uma população composta por 10 indivíduos.

### 3.5.2 Escolha da solução ótima

Uma vez que a "matriz de melhores soluções" é obtida, a solução ideal pode ser identificada com base na maior adequação entre os melhores selecionados nessa matriz. É bastante óbvio que a melhor solução ou ótimo após 20 gerações é (15 8 10 11), que foi selecionada com base em seu valor mais elevado de aptidão de 96,90, dentre todos os membros da matriz.

## 4 Contribuição esperada

Uma metodologia baseada em AG será desenvolvida, como a ilustrada com o auxílio do caso hipotético, para identificar estações de monitoramento da qualidade da água em uma rede de distribuição. Os resultados desta pesquisa podem contribuir significativamente no sentido de assegurar a qualidade da água a ser entregue aos consumidores através das rede de distribuição, através da identificação de locais apropriados das estações de controle sobre a água de toda rede. No trabalho, as estações de monitoramento da qualidade da água são localizadas com

base na vazão que chega em cada nó assumindo que a qualidade da água em um nó a jusante seja inferior a qualidade da água em um nó de montante.

## 5 Referências bibliográficas

- [1] B. H. Lee, “Locating Monitoring Stations in Water Distribution Networks”, Ph.D. Dissertation, Environmental Health Sciences, The University of Michigan, 1990.
- [2] B. H. Lee, and R.A. Deininger, “Optimal locations of Monitoring Stations in Water Distribution system”, *J. Envir. Engrg., ASCE*, 118(1) (1992), pp. 4–16
- [3] F.W. Pontius, *Water Quality and Treatment*. (4 th Edn), New York: (AWWA), McGraw-Hill, (1990).
- [4] A., Kumar, M.L., Kansal, and G. Arora, “Identification of Monitoring Stations in Water Distribution System”, *J. Envir. Engrg., ASCE*, 123 (8) (1997). pp. 746–752.
- [5] A. Kessler, Ostfeld, and G. Sineri, “Detecting Accidental Contaminations in Municipal Water Networks”, *J. Water Resour. Planning and Management, ASCE*, 124(4) (1998), pp. 192–198.
- [6] L.F.R. Reis, R.M., Porto, and F.H. Chaudhry, “Optimal Location of Control Valves in Pipe Networks by Genetic Algorithm”, *Journal of Water Resources Planning and Management, ASCE*, 123(6) (1997), 314–326.
- [7] R.W. Meier, and B.D. Barkdoll, “Sampling design for network model calibration using Genetic Algorithm”, *J. Water Resour. Plang. and Mgmt., ASCE*, 126(4) (2000), pp. 245–250.
- [8] P.F., Boulos, and T. Altman, “Explicit Calculation of Water Quality Parameters in Pipe Distribution Network.” *Journal of Civil Engineering Systems*, 10(1), (1993), pp. 187–206.
- [9] L.A. Rossman, *EPANET – User’s Manual*. Cincinnati, Ohio: United States Environmental Protection Agency (USEPA), 2000.
- [10] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Mass. Addison – Wesley, (1989).
- [11] D.A. Savic, and G.A. Walters, “Genetic Algorithms for Least–Cost Design of Water Qistribution Networks”, *Journal of Water Resources Planning and Management, ASCE*, 123 (2) (1997), pp. 67–77.
- [12] D.M. Tate, and A.E. Smith, “Expected Allele Coverage and the Role of Mutation in Genetic Algorithms”, *Proceedings of the Fifth International Conference on Genetic Algorithm*, University of Illinois at Urban-Champaign, July 17–21, 1993, pp. 31–37. Paper Received 4 February 2002, Revised 2 June 2002; Accepted 23 October 2002.