

## 4

### Modelo Inteligente de Previsão (MIP)

#### 4.1

##### Modelo Proposto

Como já mencionado, o objetivo deste trabalho é a criação de uma metodologia, ou mais especificamente, um modelo que traga algum tipo de melhoria ou inovação no campo das previsões de vazões fluviais ou, simplesmente, um novo tipo de abordagem, podendo ser usado para prever vazões em quaisquer regiões dos aproveitamentos hidroelétricos das bacias hidrográficas do país.

Neste capítulo, é apresentado esse modelo, aqui denominado “Modelo Inteligente de Previsão de Vazão” ou MIP, cuja composição inicial possui 3 módulos. Conforme mostra a figura 4.1, os dados ou séries históricas são arrumados e tratados pelo módulo 1. Após a saída do módulo 1, os dados têm dois caminhos possíveis: ou seguem até o módulo 2, onde são analisados e decompostos em componentes wavelets (aproximações e detalhes), para depois seguirem para o módulo 3, ou da saída do módulo 1 se dirigem diretamente ao módulo 3, onde são feitas as previsões relativas as 15 variáveis de saída. Essa alternativa de se tratar ou não no módulo 2 (wavelets), os dados oriundos do módulo 1, tem o seguinte objetivo: avaliar os resultados das previsões com e sem o módulo 2, desvendando-se assim, o grau de melhoria das previsões, proporcionado pela inclusão desse módulo.

Os resultados das previsões (15 séries de previsão) são comparados, então, com as séries históricas de vazões observadas, sendo em seguida calculado o erro MAPE.

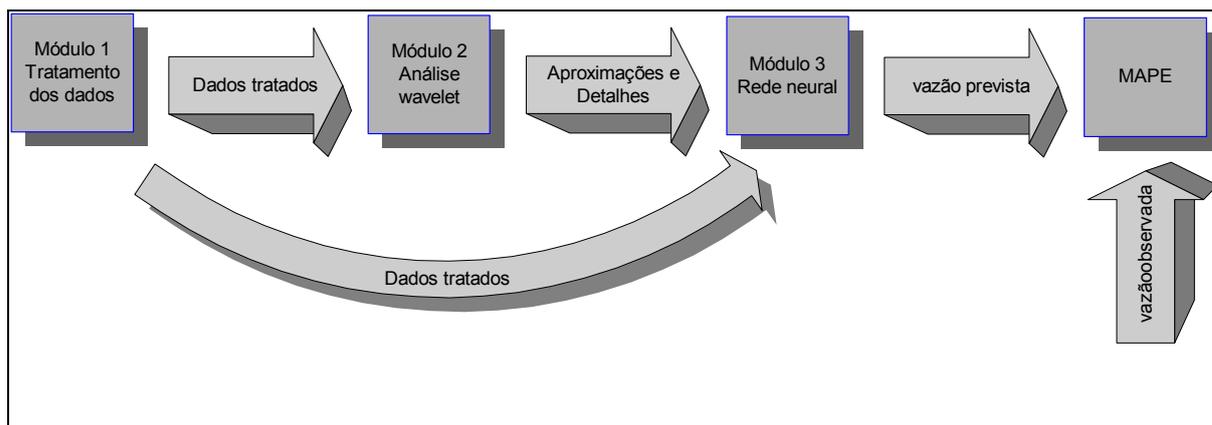


Figura 4.1: Modelo MIP composto pelos módulos 1,2 e 3

Conforme já salientado na seção 1.3 desta dissertação, as 15 variáveis de saída são as previsões de vazão de 1 até 12 dias à frente e 3 previsões médias semanais. Mais especificamente:

- 3 previsões de vazão natural média semanal – entende-se como vazão natural média semanal, uma média das vazões naturais incrementais médias diárias, calculada para 3 períodos distintos de 7 dias cada, a saber:
  - do 4º ao 10º dia (1º sábado a 2ª sexta);
  - do 5º ao 11º dia (1º domingo ao 2º sábado);
  - do 6º ao 12º dia (1ª segunda ao 2º domingo).
- Previsões de vazões naturais incrementais médias diárias do 1º ao 12º dia (12 previsões) - o primeiro dia de previsão, ou melhor dizendo, o primeiro dia a ser previsto, é sempre uma **quarta-feira**, tendo como dados mais recentes os do dia anterior, terça-feira. Então, em relação às previsões diárias do 1º ao 12º dia, estar-se-á sempre na manhã de uma quarta-feira, querendo-se prever a vazão natural incremental média diária desta mesma quarta-feira (1º dia de previsão) e a vazão natural incremental média diária dos próximos 11 dias (quinta até o 2º domingo). As 3 previsões médias semanais também são feitas a partir de uma quarta-feira, estando-se nessa quarta-feira.

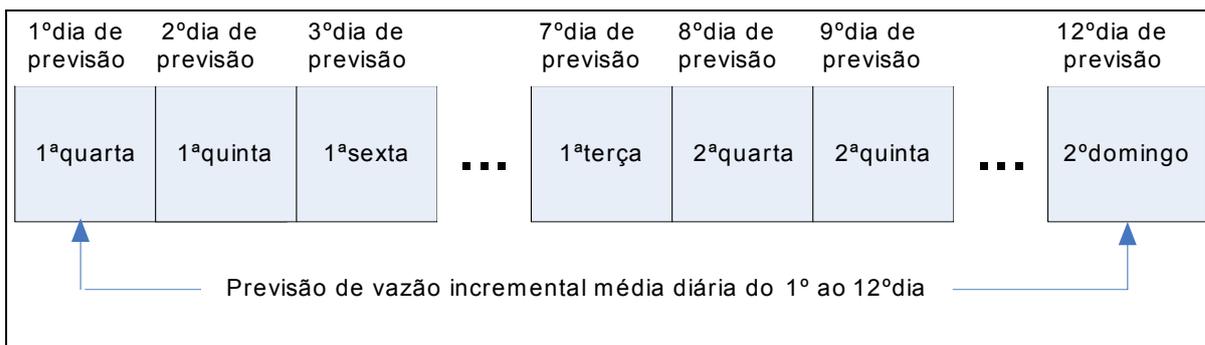


Figura 4.2: Previsão de vazão incremental média diária do 1º ao 12º dia

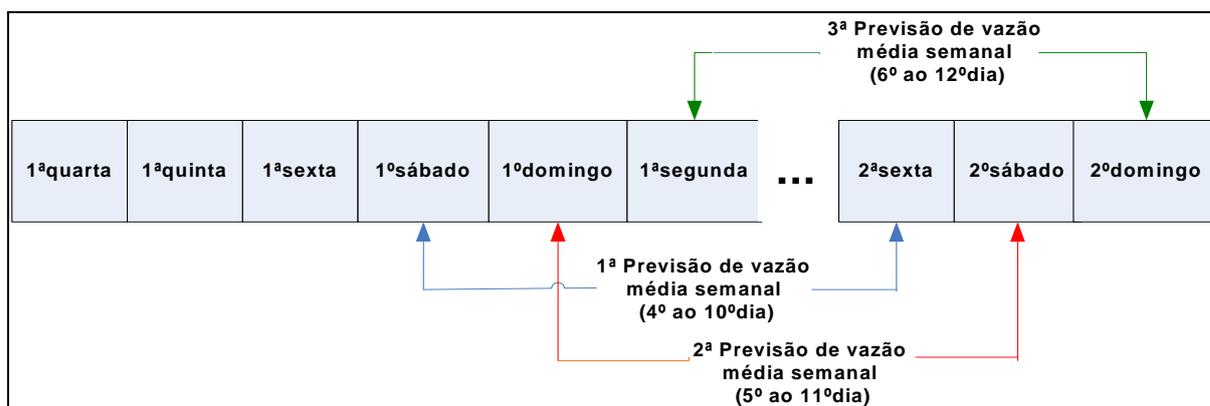


Figura 4.3: Três previsões de vazões médias semanais

A seguir procede-se o detalhamento dos 3 módulos constituintes do modelo MIP.

## 4.2

### Módulo 1 – Módulo de Tratamento dos Dados

Neste módulo, todos os dados disponíveis são tratados e arrumados de modo que se possa proceder às operações relativas aos módulos 2 e 3. Neste módulo são efetuadas ações como:

#### a) Cálculo da *Chuva Média de Thiessen da Bacia*:

Numa bacia fluvial, onde haja a necessidade de controle de vazões nos locais de aproveitamento, são extremamente importantes as informações contidas no histórico de chuvas da área de captação dessa bacia. Esse histórico é formado por um conjunto de séries históricas de precipitação, onde cada série é registrada por um posto pluviométrico situado num ponto específico da bacia. Uma maneira

mais eficaz de se trabalhar com essas várias séries de chuvas medidas pelos postos é transformar essas várias séries em uma única série de chuvas, que represente o efeito conjunto de todas as séries isoladas.

Neste cálculo, ao invés de utilizarem-se diretamente as chuvas diárias observadas em cada posto pluviométrico, é calculada, para cada dia, a chuva média diária da bacia ou chuva de Thiessen (existem três métodos para o cálculo da chuva média: método da Média Aritmética, método de Thiessen e método das Isoietas).

O efeito é como se existisse um único posto que medisse a chuva média de toda a bacia. O método de Thiessen consiste em relacionar a cada posto pluviométrico um peso correspondente a sua área de domínio dentro da bacia. Essa área, por sua vez, é calculada pelo método dos polígonos de Thiessen (ver Apêndice 2) e o peso de cada posto é dado pelo resultado da divisão da área de domínio do posto pela área total do trecho da bacia. Para cada dia, cada peso é multiplicado pelo valor observado da chuva do dia pelo respectivo posto. Somam-se então, todos os valores resultantes dessas multiplicações e o resultado é a chuva média do dia para todo o trecho da bacia. Essa chuva média é comumente chamada “chuva de Thiessen” da bacia.

b) **Completamento de dados faltantes**, ou seja, quando uma determinada série histórica apresenta falta de dados em alguns dias. No caso deste trabalho, este completamento se refere as  $n$  séries históricas de vazões fluviais ( $n$  postos pluviométricos na bacia) e as  $p$  séries históricas pluviométricas da bacia ( $p$  postos pluviométricos). O procedimento de completamento de dados faltantes pode ser efetuado por meio de diversos métodos [86], tais como:

- Caso dois postos apresentem falta de dados e o coeficiente de correlação linear entre suas séries seja alto (próximo a 1), essas séries podem ser completadas por intermédio da equação de regressão linear formada pelos dados históricos destes postos;
- Para o caso particular de séries históricas de chuvas, o preenchimento das faltas pode ser feito por método (descrito através de exemplo no Apêndice 3), onde são mudados os pesos de Thiessen dos postos que não apresentam falta de registro para um determinado dia.

### c) **Confecção das matrizes de dados**

O agrupamento das séries históricas em forma de matrizes é fundamental para a seleção de variáveis, que é uma das funções do módulo 1, e que na realidade trata da seleção das séries históricas que formarão as linhas das matrizes de dados que posteriormente serão analisadas e tratadas nos módulos 2 e 3. Essas matrizes se subdividem em 2 grupos: matrizes de entrada e matrizes de saída. Inicialmente, cada matriz de entrada tem como colunas as séries históricas relativas a um determinado grupo de dados, grupos estes que serão usados para se fazer as previsões de vazões.

Sendo assim, na matriz relativa aos dados de vazões fluviais, cada uma de suas linhas representa a série histórica de um dos postos fluviométricos. Na matriz relativa aos dados pluviométricos, cada linha representa a série histórica de um posto de medição de chuva. As matrizes de saída, por sua vez, têm como linhas as séries históricas das vazões a serem previstas. Todas essas matrizes, pelo fato de representarem somente uma variável de saída, apresentam somente uma linha (uma série histórica).

Depois do processo de seleção de variáveis, as matrizes de entrada se reduzirão a apenas uma matriz de séries históricas para cada uma das vazões a serem previstas (15 vazões). Pelo fato de representarem séries históricas, todas as matrizes, tanto de entrada como de saída, antes e depois do processo de seleção de variáveis, necessitam ter o mesmo número de colunas, porque cada coluna representa o valor diário de uma chuva ou vazão, por exemplo.

É importante lembrar que o processo de seleção de variáveis reduzirá o número de linhas de cada matriz, fazendo com que, como já foi dito, essas linhas restantes sejam concatenadas em uma única matriz de entrada para cada uma das 15 saídas (15 vetores de saída) a serem previstas. Cabe lembrar, ainda, que o primeiro dia de previsão, ou melhor dizendo, o 1º dia a ser previsto, é sempre uma **quarta-feira**, estando-se na manhã desta quarta-feira, e tendo, então, como dados mais recentes, os dados do dia anterior, terça-feira. Como já foi dito anteriormente, a escolha de quarta como o dia atual escolhido para o modelo MIP, reside no fato do ONS preparar suas previsões sempre nas manhãs de quarta-feira.

Para efeito de ilustração, supõe-se que existam os seguintes registros históricos de chuvas e vazões de uma bacia hipotética X:

- Chuvas registradas em 23 postos pluviométricos distribuídos pela bacia no período entre 23/01/1987 a 08/02/2008;
- Registro das vazões naturais incrementais entre as UHE “Fulano”, “Beltrano” e “Cicrano” da bacia X no período de 27/07/1972 a 06/01/2008;
- Registro de vazões naturais registradas por 8 postos fluviométricos no período de 05/04/1994 a 08/02/2008;
- Registro de chuvas previstas pelo modelo ETA no período entre 01/01/1996 a 08/02/2008 – o modelo ETA realiza previsões do total de chuva acumulada em 24 horas, realizadas semanalmente, às quartas-feiras, para um horizonte de 10 dias à frente, incluindo quarta-feira, em formato de grade, com resolução espacial de 40 km x 40 km. Os dias previstos, então, são a 1ª quarta (1º dia de chuva prevista), 1ª quinta, 1ª sexta, sábado, domingo, segunda, terça, 2ª quarta (oito dias à frente), 2ª quinta e 2ª sexta (10ª chuva prevista).

Sendo assim, pelo fato de que todas as matrizes de dados necessitam ter o mesmo tamanho, é sempre utilizado, inicialmente, como referência, o período de interseção temporal entre todas as séries, que neste caso é 01/01/1996 – segunda-feira a 06/01/2008 – domingo. Isso não significa, conforme será visto mais adiante, que todas as séries tenham de se restringir a este período.

Serão então, montadas as seguintes matrizes de dados:

#### **Matrizes de entrada**

- **Matriz Vinc** (vazão incremental): essa matriz é de suma importância, pois representa os valores históricos anteriores da vazão a ser prevista. Optando-se por formar essa matriz somente com o último valor histórico de vazão incremental antes da previsão (terça-feira), fica então a matriz restrita ao período de 02/01/1996 – terça-feira a 01/01/2008 – terça-feira, o que resulta numa matriz Vinc de vazão incremental **1 x 626**, já que há 626 terças-feiras nesse período.

- **Matriz Th** (chuvas de Thiessen): depois do cálculo da chuva média de Thiessen da bacia para cada dia, essa matriz pode assim ser montada:

- 1 matriz contendo as 7 últimas chuvas de Thiessen (quarta a terça) formando uma matriz onde a 1ª linha contém todas as últimas chuvas registradas (chuvas de terça), a 2ª linha contém todas as penúltimas

chuvas registradas (chuvas de segunda-feira) e por fim a 7ª linha possui todas as chuvas registradas 7 dias atrás (chuvas de quarta-feira);

- 1 matriz contendo  $n$  linhas dentre as 7 da matriz acima enunciada, formando uma matriz  $n \times 250$ .

Optando-se por montar uma matriz com todas as 7 últimas chuvas de Thiessen (quarta a terça), o período temporal dessa matriz, será, em princípio, de 27/12/1995 – quarta – a 01/01/2008 – terça. Logo, há 626 quartas, quintas, sextas, sábados, domingos, segundas e terças entre 27/12/1995 e 01/01/2008. Assim, forma-se a Matriz Th  $7 \times 626$ , onde cada linha representa o histórico de uma variável, ou melhor, uma das 7 últimas chuvas de Thiessen (o histórico de todas as chuvas de Thiessen de terça-feira, ou o histórico de todas as chuvas de Thiessen de segunda-feira, ou ..., ou o histórico de todas as chuvas de Thiessen de quarta-feira) .

Serão realizadas para a Matriz Th, então, 15 seleções de variáveis, uma para cada previsão a ser feita.

– **Matriz ETA:** sendo o objetivo do modelo ETA prever as 10 chuvas seguintes, sendo a primeira dessas dez a chuva prevista ETA de quarta-feira (também o primeiro dia de previsão do modelo MIP), é razoável que se escolha a matriz ETA representando um período histórico contendo todos os dez dias à frente de previsão ETA de chuvas. Como acontece na matriz de Thiessen, mais tarde, a seleção de variáveis escolherá, dentre essas 10 chuvas previstas ETA, aquelas que mais influenciam nas vazões incrementais de 1 até 12 dias à frente e nas 3 vazões médias semanais. Também, como acontece na matriz de Thiessen, serão realizadas 15 seleções de variáveis, uma para cada previsão a ser feita. Assim, o período dessa matriz, atentando para o fato de que ela deve ter o mesmo número de colunas das demais, é de 03/01/1996 – quarta-feira a 11/01/2008 – sexta-feira. Com isso, há 626 grupos de dez chuvas de previsão ETA, formando uma Matriz ETA  $10 \times 626$ , onde cada linha representa o histórico de uma variável, ou melhor, as chuvas previstas ETA de um dos 10 dias à frente (histórico de previsão de chuva ETA de todas as primeiras quartas, ou o histórico de previsão de chuva ETA de todas as primeiras quintas, ..., ou o histórico de previsão de chuva ETA de todas as segundas quintas ou finalmente, o histórico de previsão de chuva ETA de todas as segundas sextas).

- **Matriz Flu** (vazões fluviais): sendo 8 o número de postos fluviométricos na bacia X, por exemplo, é razoável que a matriz de dados de vazões fluviais seja montada com os dados históricos de vazões de todos os 8 postos. A dúvida é quanto aos dias dessas séries históricas que serão usados. Ou seja, pode-se usar somente o último dado diário de vazão – terça-feira – de cada posto antes do 1º dia a ser previsto (quarta-feira). Alternativamente, podem ser usados os dois (segunda e terça), três (domingo, segunda, terça) ou até mais dias anteriores. O modelo MIP utiliza, em princípio, somente as últimas vazões fluviais registradas (terça-feira). No entanto, como já foi visto nas demais matrizes, a seleção de variáveis escolherá, dentre essas 8 vazões fluviais registradas de terça-feira, aquelas que mais influenciam nas vazões incrementais de 1 até 12 dias à frente e nas 3 vazões médias semanais. Serão realizadas, para a matriz Flu, então, 15 seleções de variáveis, uma para cada previsão a ser feita.

O período utilizado para cada um dos 8 postos será 02/01/1996 – terça-feira a 01/01/2008 – terça-feira. Surge, então, a matriz Flu de tamanho **8 x 626**, onde cada linha representa o histórico de vazões de terça de um dos 8 postos fluviais.

#### **Matrizes de saída**

Como são 15 as variáveis de saída, serão 15 as matrizes de saída:

- 12 **matrizes Vm<sub>di</sub>** (i=1,2,...12), relativas aos 12 dias de vazões a serem previstas, ou 12 dias de vazões observadas.

- Vmd1 – vazões observadas de todas as quartas-feiras no período de 03/01/1996 – quarta-feira a 02/01/2008 – quarta-feira, **1 x 626**;
- Vmd2 – vazões observadas de todas as quintas-feiras no período de 04/01/1996 – quinta-feira a 03/01/2008 – quinta-feira, **1 x 626**;
- Vmd3 – vazões observadas de todas as sextas-feiras no período de 05/01/1996 – sexta-feira a 04/01/2008 – sexta-feira, **1 x 626**;
- E assim sucessivamente até Vmd12 - vazões observadas de todos os domingos no período de 14/01/1996 – domingo a 13/01/2008 – domingo, **1 x 626**.

- 3 **matrizes Vm<sub>si</sub>** (i=1,2,3), relativas às 3 vazões médias semanais a serem previstas ou 3 vazões médias semanais observadas.

- Vms1 – vazão média semanal calculada entre os períodos 06/01/1996 – sábado a 12/01/1996 – sexta-feira e 05/01/2008 - sábado a 11/01/2008 – sexta-feira, 626 períodos – 626 médias, **1 x 626**;
- Vms2 – vazão média semanal calculada entre os períodos 07/01/1996 – domingo a 13/01/1996 – sábado e 06/01/2008 – domingo a 12/01/2008 – sábado, 626 períodos – 626 médias, **1 x 626**;
- Vms3 – vazão média semanal calculada entre os períodos 08/01/1996 – segunda a 14/01/1996 – domingo e 07/01/2008 – segunda a 13/01/2008 – domingo, 626 períodos – 626 médias, **1 x 626**;

#### d) Normalização dos dados

Após a montagem das matrizes, o próximo passo seria a seleção de variáveis. Tendo em vista que as variáveis ou séries históricas das matrizes de dados, notadamente chuva e vazão, são grandezas completamente distintas (apesar de serem correlacionadas) e tendo em vista que o processo de seleção de variáveis escolhido é o LSE, método que avalia o grau de importância das variáveis de um processo em relação as variáveis de saída por meio da medição da variação das saídas quanto à variação das variáveis do processo, essa seleção só pode ser executada após um processo de normalização das séries históricas (matrizes). Essa normalização será aproveitada quando da entrada e tratamento das séries históricas selecionadas nos módulos 2 e 3.

A normalização utilizada é dada pela equação abaixo:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (4.1)$$

onde  $x_{norm}$  é o valor normalizado,  $x$  é o valor original,  $x_{min}$  o valor mínimo da série histórica, e  $x_{max}$  é o valor máximo.

e) **Seleção de variáveis**, ou seja, é feita a seleção das séries históricas que serão as entradas do módulo 2 (wavelets) e módulo 3 (redes neurais). Mais especificamente, na verdade, ocorrem 2 fases de seleção de variáveis. Na primeira fase são selecionadas, dentro de cada matriz de entrada, as linhas (que representam valores de variáveis) que mais influenciam cada uma das variáveis ou matrizes de saída, também já anteriormente descrita. Na segunda fase, com as variáveis selecionadas de cada matriz de entrada, monta-se uma matriz resultante. Como são 15 as variáveis de saída, são 15 as matrizes resultantes. Procedese,

então, uma segunda seleção de variáveis dentro de cada uma dessas 15 matrizes resultantes.

O método de seleção de variáveis escolhido é o LSE, pois, ao contrário de outros métodos, não supõe relação linear entre entrada e saída, linearizando possíveis relações não lineares. Este método já foi descrito em 3.1.2.

O processo de seleção de variáveis ficará claro, quando da realização do estudo de caso, adiante, no capítulo 5, mais especificamente em 5.2.3.

### 4.3

#### **Módulo 2 – Módulo Wavelet**

A introdução da análise wavelet (representada por este módulo) no MIP é uma tentativa de fazer com que o módulo 3 possa entender melhor os padrões das séries históricas. Como será visto adiante, realmente, a decomposição de algumas (somente algumas) séries históricas em sinais de alta e baixa frequência melhora significativamente o desempenho do módulo 3.

O desenvolvimento acontece da seguinte forma: os dados já pré-tratados no módulo 1 são transformados por meio da metodologia wavelet, ou seja, os dados são encarados pelo módulo como sinais e são transformados em componentes wavelets – componentes de aproximação e detalhes. É importante salientar, mais uma vez, que essa transformação é benéfica – no sentido de melhorar o desempenho do módulo 3 – somente para algumas séries de dados.

No modelo MIP, o módulo de wavelets decompõe os dados ou séries temporais que vêm do módulo 1, em princípio, em 2 níveis de coeficientes wavelets, podendo chegar a no máximo 4 níveis [84,87]. Conforme visto anteriormente (seção 3.3), o nível 1 contém um componente de detalhe – D1 – e o nível 2 contém um componente de aproximação – A2 – e um de detalhe – D2.

São utilizadas para a decomposição das séries temporais as famílias wavelets Daubechies, Biortogonal e Coiflets [85], tendo em vista essas famílias possuírem importantes características, tais como suporte compacto, análise ortogonal e biortogonal, possibilidade de reconstrução exata e presença de transformada discreta.

De início, o modelo procura descobrir a melhor estrutura de decomposição wavelet para  $V_{ms1}$ , entendendo que para as outras previsões não haverá muitas

mudanças (entende-se como melhor estrutura aquela que implicará em menores erros MAPE no módulo 3).

#### 4.4

### Módulo 3 – Módulo de Redes Neurais

Esse é o módulo que faz as previsões, sendo composto por 15 redes neurais. Neste módulo, todos os dados, depois de terem passado pelo módulo 1 e alguns pelo módulo 2, são analisados por uma das 15 redes neurais que executarão a previsão. Essas 15 redes neurais dizem respeito aos dois horizontes de previsão já descritos no módulo 1: previsões de vazão natural média diária de 1 até 12 dias à frente (12 redes neurais) e 3 previsões de vazão natural média semanal (3 redes neurais).

Como alguns dados advêm do módulo 2 e outros do módulo 1, o módulo 3 opera como se fosse composto por 2 núcleos de 15 redes neurais cada, onde um núcleo é responsável pela recepção e análise dos dados oriundos do módulo 2 e o outro responsável por aqueles oriundos diretamente do módulo 1 (figura 4.4).

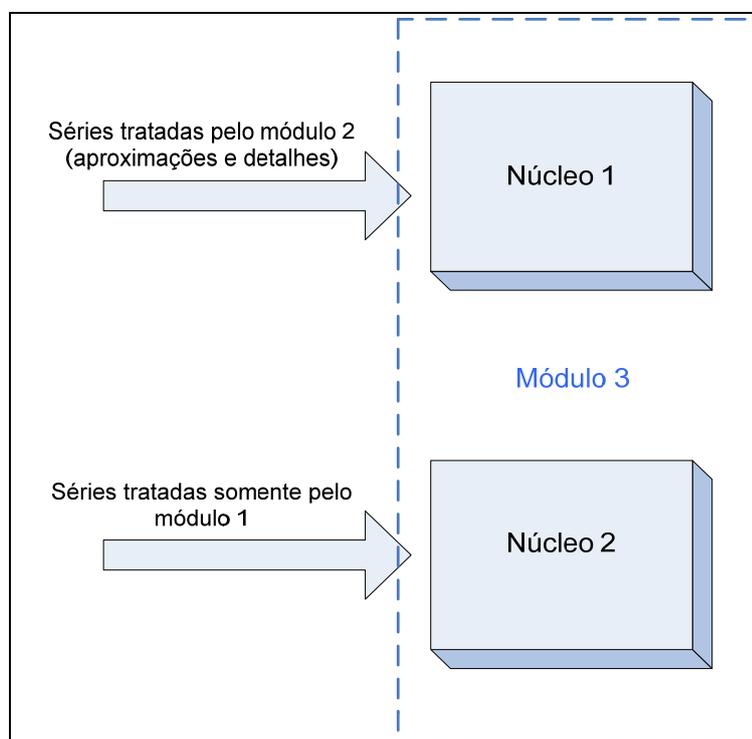


Figura 4.4: Módulo 3 com seus dois núcleos fictícios

Na realidade, só existe um núcleo, que é o próprio módulo 3 composto por 15 redes neurais, cada uma responsável por um horizonte de previsão (figura 4.5).

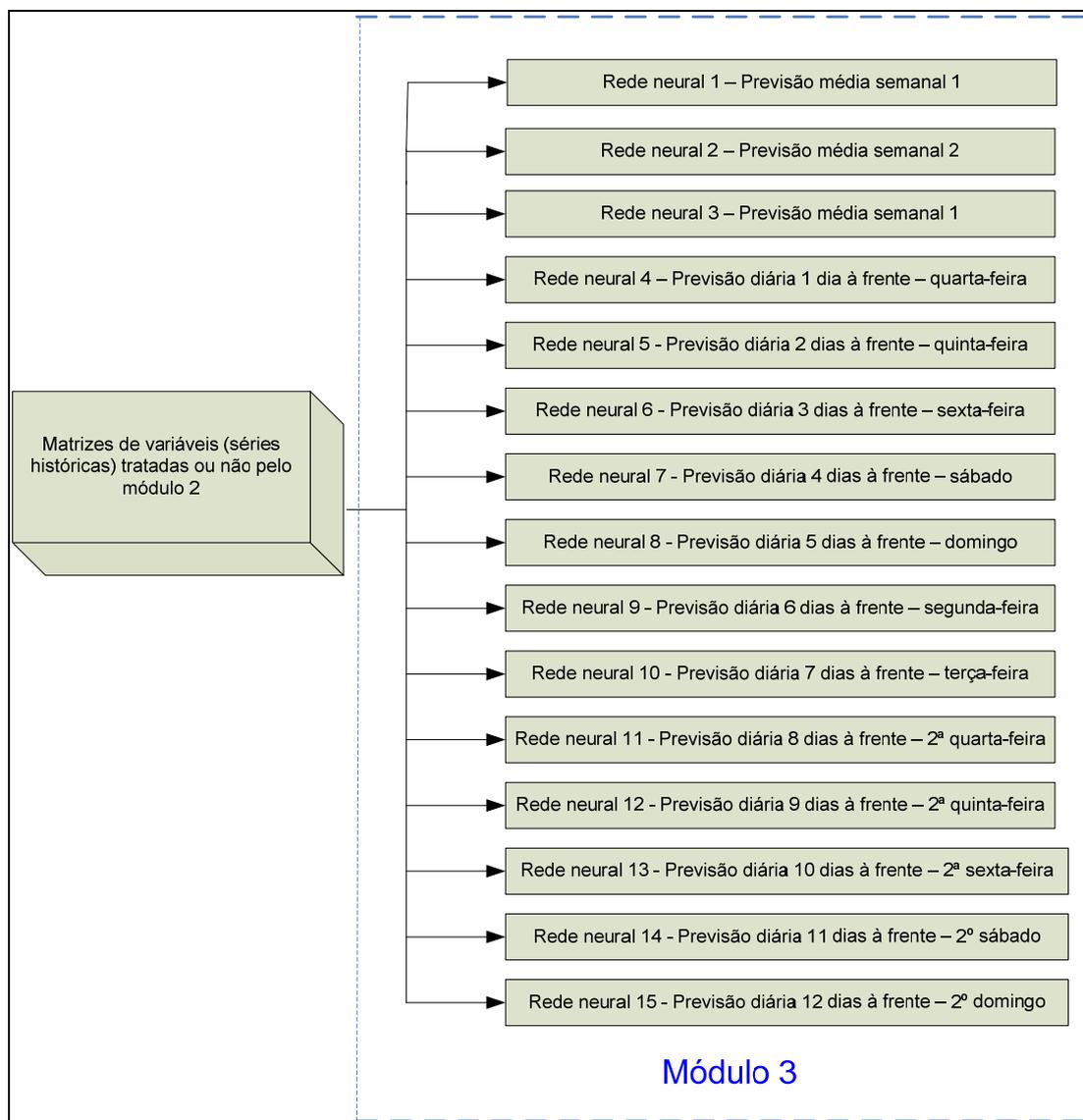


Figura 4.5: Módulo 3 com suas 15 redes neurais