

4

O Processo Estocástico Neural no Tratamento de Séries de Afluências Mensais

Escolheu-se nessa tese aplicar o processo estocástico neural (PEN) para realizar a modelagem de séries de afluências mensais, dentro do contexto do planejamento da operação energética do sistema hidrotérmico brasileiro [36, 21, 37, 53].

O Brasil possui uma sofisticada e eficiente estrutura de fornecimento de eletricidade com forte predominância da geração por parte de usinas hidrelétricas, com grandes reservatórios de regularização plurianual, localizadas em diferentes bacias hidrográficas [60]. As usinas hidrelétricas possuem acoplamento temporal e espacial, onde o controle da vazão dos reservatórios é de fundamental importância para evitar desperdício e obter o máximo de energia contida, exigindo um esforço do setor elétrico no sentido de desenvolver ferramentas e técnicas de modelagem de operação de reservatórios, de geração de séries sintéticas de afluências, etc. [32].

O planejamento da operação energética do sistema hidrotérmico brasileiro é realizado pelo Operador Nacional do Sistema Elétrico (ONS) [50], cujo objetivo principal é minimizar o valor esperado do custo total da operação, considerando a garantia de suprimento de energia, sendo este custo dependente das vazões dos rios brasileiros. Como não se conhecem as vazões afluentes futuras que ocorrerão aos reservatórios durante as decisões operativas, este problema é inerentemente estocástico.

O tratamento das séries de afluências é importante no planejamento da operação do SIN para determinar o quanto se deve utilizar dos estoques de energia, em forma de água armazenada, nos reservatórios das usinas hidrelétricas. Caso se utilize toda a água dos reservatórios no momento presente e caso ocorra baixas afluências no futuro, isso implicará no acionamento das usinas termoelétricas, o que aumentará o custo no futuro e o risco da demanda não ser totalmente atendida (custo de *déficit*). Por outro lado, ao se poupar a água disponível nos reservatórios no presente, e as afluências no futuro forem altas, as capacidades máximas de armazenamento dos reservatórios das usinas

hidrelétricas serão atingidas, obrigando-as a verter água, o que representa um desperdício de energia. Por isso, um bom planejamento do SIN envolve a decisão sobre o quanto se utilizar dos estoques de energia, em forma de água, dos reservatórios das usinas hidrelétricas, que está intrinsecamente ligada à incerteza das tendências hidrológicas [54, 38].

Uma das características das séries de aflúências é a sua natureza não estacionária, devido aos períodos de cheia e seca do ano. Por este motivo, é de fundamental importância o estudo e desenvolvimento de modelos capazes de capturar essa natureza, associada à sazonalidade e elevado nível de incerteza, presente nas aflúências dos rios brasileiros [35].

As séries de aflúências mensais são séries sazonais com períodos de 12 meses, as quais, geralmente, apresentam correlações periódicas, e os modelos de séries temporais que são sugeridos para analisá-las são os modelos auto-regressivos periódicos [39]. Atualmente essas séries são tratadas por metodologias baseadas nos modelos de Box & Jenkins [9]. A estocasticidade das aflúências é modelada por cenários hidrológicos gerados sinteticamente por um modelo estocástico periódico auto-regressivo de ordem p - PAR(p) [37].

Dado que o problema do planejamento da operação do SIN é um problema de grande porte, necessita-se da adoção de simplificações que consistem na divisão em subproblemas com diferentes horizontes de planejamento (tais como médio e curto prazo), com diferentes graus de detalhamento do problema e representação da estocasticidade das aflúências, como ilustra a Figura 4.1.

Observe que à medida que o horizonte de planejamento aumenta, a representação das incertezas aumenta e o grau de detalhe da representação do sistema elétrico diminui. Nesse estudo de caso decidiu-se focar no planejamento de médio prazo.

Atualmente, o SIN é segmentado em quatro subsistemas correspondentes aos sistemas interligados: Sul, Sudeste/Centro-Oeste, Nordeste e Norte. Por essa razão, no planejamento de médio prazo ocorre a agregação dos reservatórios das usinas em reservatórios equivalentes de energia, um para cada subsistema. Ocorre também a agregação das aflúências às usinas em Energias Naturais Afluentes (ENA). Esta energia corresponde à estimativa da energia que pode ser gerada com todas as vazões afluentes a cada um dos reservatórios que compõem aquele reservatório equivalente, segundo uma política de operação.

Desde quando ocorreu uma crise energética, meados de 2001 e 2002, onde o governo foi obrigado a anunciar um plano de racionamento de energia, tornou-se evidente a necessidade de discussão e revisão dos modelos usados no

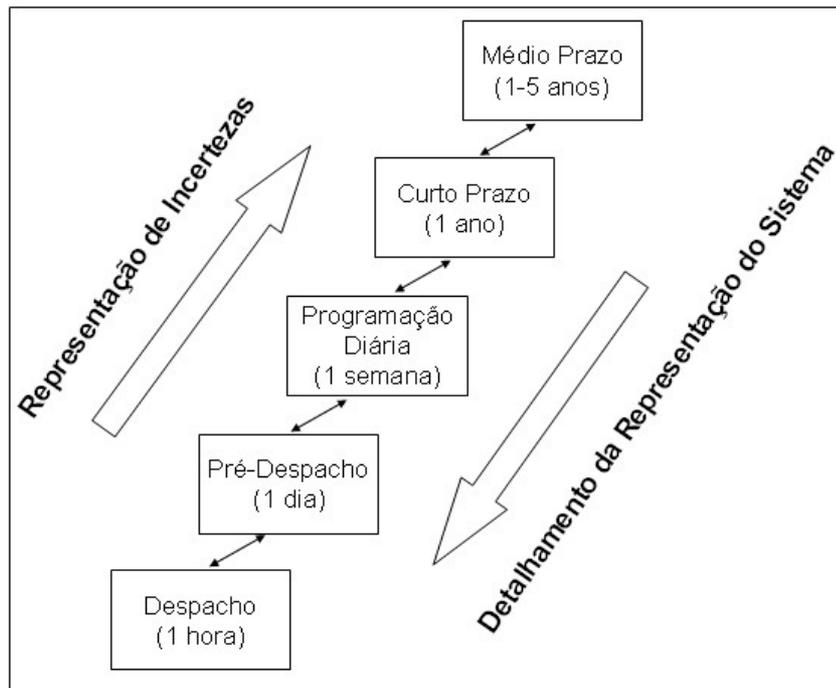


Figura 4.1: Modelagem de Sistemas Hidrotérmicos no Planejamento da Operação Energética

setor elétrico brasileiro [60]. Isso motivou, nessa tese, a utilização do modelo Processo Estocástico Neural (PEN) para tratar as incertezas das afluências através da geração de um conjunto de cenários hidrológicos, para serem usados no planejamento da operação energética de médio prazo. Como o PEN é um modelo não-linear, devido ao uso das redes neurais artificiais (RNAs) na sua formação, ele pode capturar as características da série histórica de afluências mensais que os modelos tradicionais lineares não conseguem. No caso, a proposta é utilizar o modelo PEN para simular cenários de ENA. Portanto, deve-se ajustar um modelo PEN para cada subsistema do SIN.

Conforme comentou-se, atualmente, os cenários de ENA são gerados pelo modelo PAR(p) e são utilizados pelo modelo de planejamento de médio prazo empregado pelo ONS, chamado NEWAVE. Como a execução desses programas ocorre em conjunto [41], não é possível utilizar o conjunto de cenários gerados pelo PEN no NEWAVE. Assim, nesta tese, os cenários do PEN foram utilizados no modelo de otimização monocritério da operação de sistemas hidrotérmicos utilizando algoritmos genéticos, desenvolvido em [60].

A fim de avaliar as características das séries geradas pelo PEN, são aplicados alguns testes de hipótese relevantes com o intuito de verificar a aderência desse conjunto de séries sintéticas às séries históricas de ENA. A idéia é aplicar os mesmos testes ao conjunto de séries sintéticas gerado pelo

modelo utilizado atualmente, o PAR(p), e comparar os resultados dos dois modelos para a análise de desempenho do PEN. Contudo, como o PAR(p) gera séries sintéticas dentro do NEWAVE, decidiu-se nesse trabalho implementar uma versão do PAR(p) segundo a descrição obtida em [37].

4.1

Processo Estocástico Neural na Modelagem de Séries Hidrológicas

Para cada subsistema do Sistema Interligado Nacional (SIN) existe um histórico consolidado de Energia Natural Afluente (ENA) que apresenta valores desde 1931 até 2005 ¹. A Figura 4.2 apresenta as séries históricas de cada subsistema: Sudeste/Centro-Oeste, Sul, Nordeste e Norte.

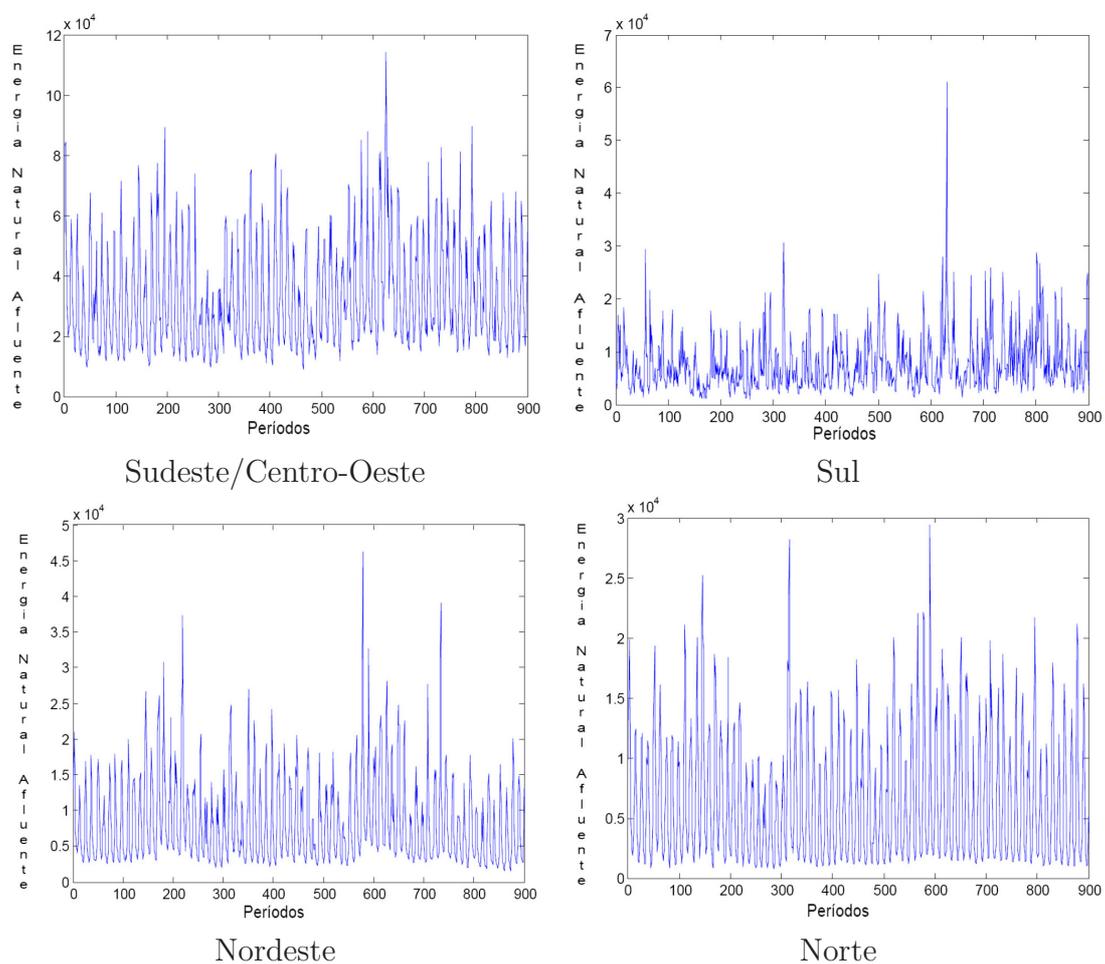


Figura 4.2: Séries Históricas de Energia Natural Afluente

Ao fazer um estudo da variação de ENA por mês ao longo dos anos, observou-se que no subsistema Sudeste/Centro-Oeste existe um ano atípico

¹Atualmente o histórico apresenta valores de 1931 até 2008. Contudo foi mantido o histórico com dados até 2005 para fazer comparações com experimentos realizados desde o início desse trabalho.

onde choveu muito em todos os meses, como pode ser visto na Figura 4.3. A figura apresenta as curvas de cada ano das séries históricas de ENA. Os pontos em destaque em cada mês indicam o valor máximo de ENA já registrado até agora. Observe que no subsistema do Sudeste/Centro-Oeste todos máximos dos meses ocorreram no mesmo ano, que no caso foi o ano de 1983. Além disso, para alguns meses, como por exemplo junho, o valor do máximo é muito atípico quando comparado com os valores dos outros anos. Através de experimentos preliminares, verificou-se que esse ano atípico realmente dificultava a estimação do modelo Processo Estocástico Neural (PEN) (o aprendizado das redes neurais que estavam trabalhando nessa série histórica). Decidiu-se então fazer um estudo das características da série histórica, onde esse ano de 1983 foi retirado do histórico e foi colocada no lugar a média dos valores dos anos de 1982 e 1984, para manter os 75 anos no histórico. Como os experimentos desse estudo tiveram melhor desempenho em relação aos experimentos com a série histórica original, manteve-se o histórico de ENA do subsistema Sudeste/Centro-Oeste dessa forma nesse trabalho.

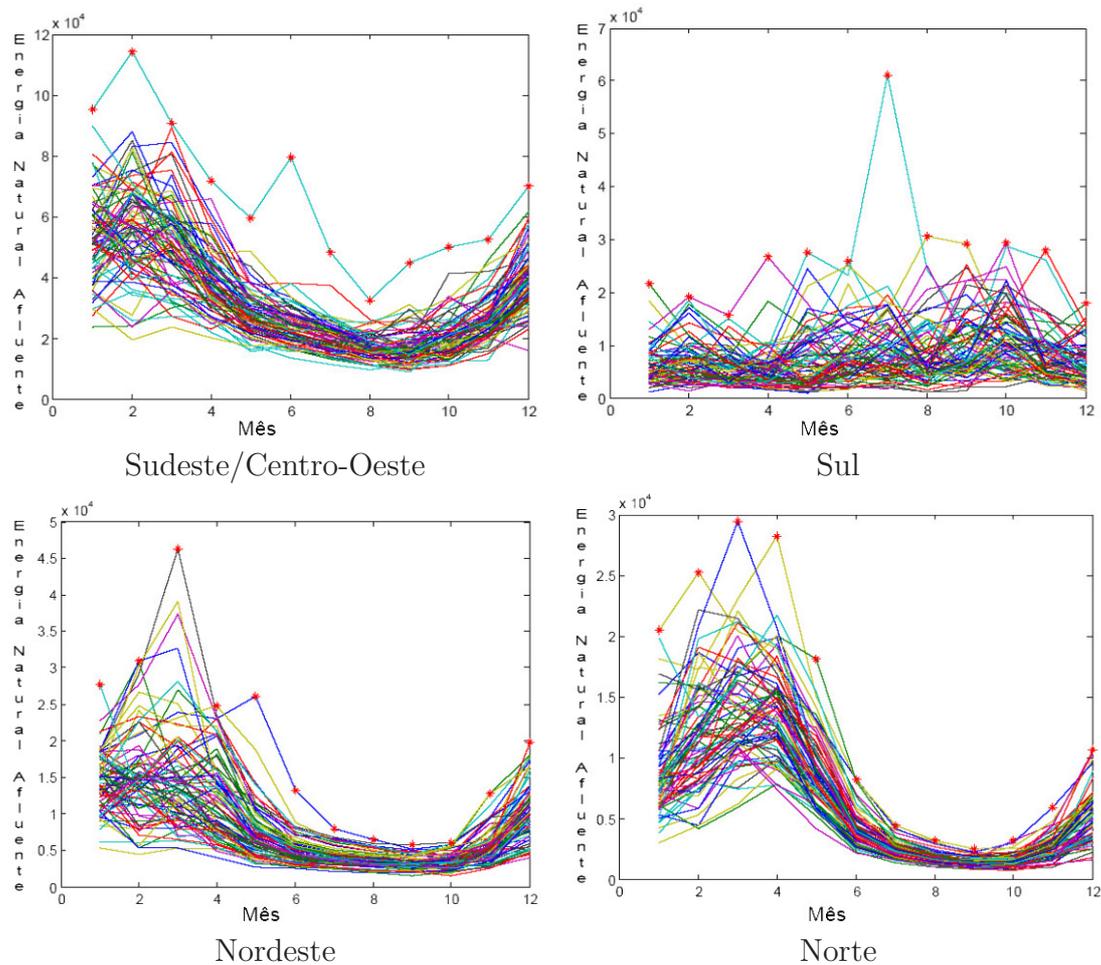


Figura 4.3: Dados Anuais das Séries Históricas de ENA

A modelagem do processo estocástico neural (PEN) para o tratamento de séries de ENA é composta de 4 sistemas $PEN(p, l)$, um para cada subsistema do SIN (a fim de capturar as características específicas da sua série histórica), conforme é indicado na Figura 4.4.



Figura 4.4: Modelagem do Processo Estocástico Neural no tratamento de Séries de Afluências Mensais

Como as séries de ENA são séries sazonais com períodos de 12 meses, os sistemas $PEN(p, l)$ são compostos por 12 componentes estocásticas (CEs), uma para cada mês. A disposição dessas CEs é apresentada no diagrama de blocos da Figura 4.5 (que é similar à Figura 3.12, sendo $s = 12$). É preciso fazer um ajuste do modelo PEN à série de ENA em que ele será empregado. Como comentou-se na Seção 3.1, esse ajuste deve seguir três etapas importantes:

1. Determinação da estrutura do modelo
2. Estimação dos seus parâmetros
3. Validação dos resíduos gerados

4.1.1

Determinação da Estrutura do Processo Estocástico Neural modelado para a Série de Energia Natural Afluente

Considere $Z(t)$ como sendo o valor de energia natural afluente (ENA) para o índice de tempo t dado por:

$$t = (r - 1) \cdot s + m$$

onde $m = 1 \dots s$ corresponde ao mês, $s = 12$ e r corresponde ao ano.

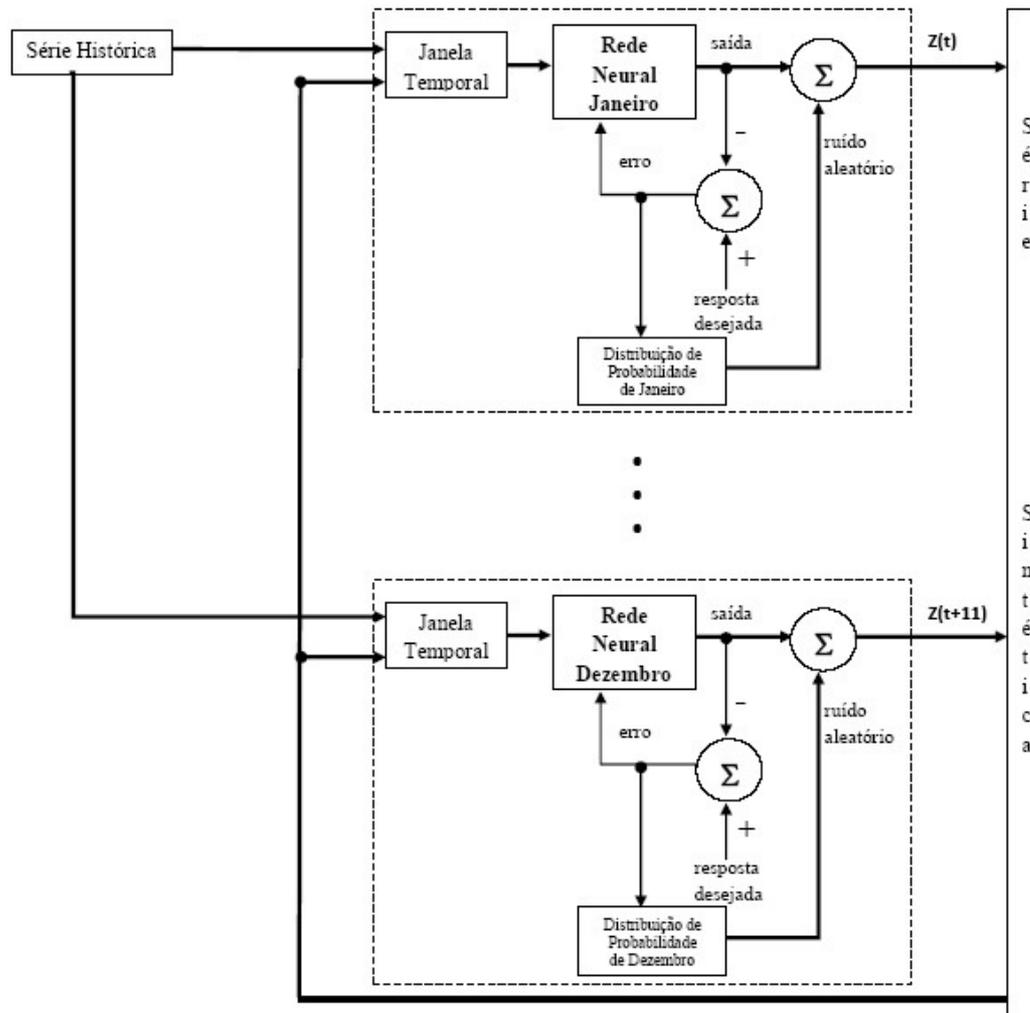


Figura 4.5: Diagrama do Processo Estocástico Neural na geração de Séries Sintéticas de Afluências Mensais

Para cada mês m , a janela temporal da rede neural artificial (RNA) da componente estocástica (CE) m do Processo Estocástico Neural (PEN) é formada pelo valor de Energia Natural Afluente (ENA):

- **no instante $t - s$** : $Z(t - s)$ - que corresponde à informação do mês defasada de um período. Isso reforça o aprendizado do comportamento periódico do mês m que a RNA representa; e
- **os valores dos p_m meses anteriores ao mês m** : $Z(t - 1), Z(t - 2), \dots, Z(t - p_m)$ - que representam os termos auto-regressivos da série e p_m é chamada de ordem do mês m .

Tem-se que a ordem p_m pode variar de 1 até 11, que corresponde ao *lag* (número de termos passados) máximo permitido para manter a sazonalidade anual.

Como comentou-se na Seção 3.2, para cada RNA de um modelo PEN é preciso definir o número de entradas e o número de neurônios na camada oculta. Como visto, o número de entradas está ligado à ordem p_m de cada mês m . Dado que as séries de ENA são séries mensais não estacionárias, devido aos períodos de cheia e seca do ano, através de experimentos preliminares teve-se informações de que alguns meses normalmente precisam de *lags* menores que outros. Logo, decidiu-se testar empiricamente 4 tipos de ordem para cada mês:

1. defasagem de 3 meses
2. defasagem de 6 meses
3. defasagem de 9 meses
4. defasagem de 11 meses

O número de neurônios na camada oculta de cada RNA também é escolhido de forma empírica. Como o número máximo de entradas de uma RNA pode ser 12, decidiu-se testar para cada ordem definida configurações desde 1 até 20 neurônios na camada oculta.

Assim, são determinadas as estruturas do modelo $PEN(p, l)$ onde são testadas configurações de RNA com a ordem p ($p = 3, 6, 9$ ou 11) e o número l de neurônios na camada oculta ($l = 1 \dots 20$), para cada ordem p .

4.1.2

Estimação dos parâmetros do Processo Estocástico Neural

Para poder gerar cenários de Energia Natural Afluente (ENA), os parâmetros do Processo Estocástico Neural (PEN), que correspondem aos pesos sinápticos das RNAs, precisam passar pela etapa de estimação (através do treinamento das RNAs).

No estudo de caso desse trabalho, os pesos das RNAs foram ajustados através do algoritmo de treinamento *Levenberg-Marquardt* (LM) [25], uma variação do algoritmo *backpropagation*, comentado na Seção 2.4.1.

Os dados que compõem os conjuntos de padrões entrada-saída utilizados pelo algoritmo de treinamento são retirados do histórico de ENAs. O histórico foi dividido em 2 conjuntos: treinamento e validação. Como comentou-se na Seção 3.3, na etapa de validação do treinamento das RNAs do PEN gera-se um cenário de x anos com as saídas das RNAs. Por esse motivo, usa-se os x últimos anos do histórico para compor o conjunto de padrões entrada-saída de validação e todos os anos anteriores para o conjunto de padrões entrada-saída de treinamento. Nesse trabalho decidiu-se utilizar a quantidade de x anos no

cenário de validação igual à quantidade de anos dos cenários que serão gerados pelo PEN, que corresponde à quantidade de anos que, geralmente, abrange o horizonte de planejamento de médio prazo - 5 anos. Como os cenários serão utilizados pelo modelo de otimização utilizando algoritmos genéticos e esse modelo não sofre a influência do efeito de “fim de horizonte”, não é preciso gerar 5 anos além do desejado, como ocorre no NEWAVE.

Os dados da série histórica Z são normalizados, dentro do intervalo $[0,1]$, antes de serem usados como padrões de dados no treinamento das RNAs. Essa normalização é feita através da equação 4-1.

$$\tilde{Z}(k) = \frac{Z(k) - Z^{\min}(m)}{Z^{\max}(m) - Z^{\min}(m)} \quad (4-1)$$

onde $Z(k)$ é o k -ésimo valor do mês m , $\tilde{Z}(t)$ é o seu valor normalizado, $Z^{\min}(m)$ corresponde ao menor valor do mês m presente no conjunto de treinamento diminuído de 10%, e $Z^{\max}(m)$ corresponde ao maior valor do mês m presente no conjunto de treinamento acrescido de 10%. Essa folga de 10% permite que se normalize corretamente, dentro do intervalo $[0, 1]$ um dado futuro menor/maior que o menor/maior valor que o registrado no conjunto de treinamento. Caso seja gerado um valor menor/maior que a folga de 10%, para não ficar muito divergente dos outros valores gerados, ele será considerado um *outlier* e será substituído por outro valor.

Porém, para realizar o cálculo das métricas de desempenho que avaliam o treinamento das RNAs, é preciso fazer a desnormalização do seu valor de saída, a fim de compará-la com o valor original (desnormalizado) da série histórica Z . A equação 4-2 apresenta a desnormalização da saída y_{out} da RNA do mês m .

$$Y = y_{out} \cdot (Z^{\max}(m) - Z^{\min}(m)) + Z^{\min}(m) \quad (4-2)$$

onde y_{out} é o valor obtido na saída da RNA, Y é o seu valor desnormalizado, $Z^{\min}(m)$ e $Z^{\max}(m)$ são, respectivamente, o menor e o maior valor do mês m presentes no conjunto de treinamento.

Conforme comentou-se na seção anterior, para cada mês m treina-se RNAs com 4 tipos de ordem ($p = 3, 6, 9$ ou 11) e, para cada ordem, 20 possibilidades de neurônios na camada oculta das RNAs que o compõem ($l = 1 \dots 20$), dando um total de 80 configurações a serem avaliadas para cada mês m .

Além disso, cada uma dessas 80 configurações de RNA são treinadas 10 vezes, com diferentes inicializações dos pesos sinápticos. Avalia-se o desempenho dessas 10 inicializações para cada configuração. Dessa forma, evita-se que a escolha de uma configuração de RNA seja influenciada pelo mal desempenho

devido a uma inicialização ruim dos pesos sinápticos.

Como foi descrito na Seção 3.3, é possível interromper o treinamento de uma configuração de RNA de duas formas: de forma coletiva, onde as RNAs de todos os meses são salvas na mesma época (ver Figura 3.10), ou de forma individual por mês, onde as RNAs são salvas em épocas distintas (ver Figura 3.11). Portanto, para cada configuração de RNA do mês m treinada, salvam-se 2 RNAs cujos parâmetros de cada RNA salva são determinados de acordo com o método de interrupção utilizado.

Após o término da etapa de treinamento de todas essas redes, escolhe-se a melhor configuração, de acordo com uma medida de desempenho estabelecida, para representar a RNA da componente estocástica (CE) do mês m que fará parte do Processo Estocástico Neural $PEN(p, l)$.

Medidas de Desempenho para escolha da Rede Neural de cada mês

Nesse estudo de caso usou-se diferentes medidas de desempenho para escolher a rede neural artificial (RNA) da componente estocástica (CE) do mês m do Processo Estocástico Neural (PEN), dentre todas as configurações de RNAs treinadas.

O MAPE mensal de validação, calculado de forma individual por mês através da equação 3-9 é uma das medidas de desempenho utilizadas. Dentre todas as configurações treinadas para o mês m , escolhe-se a que fornecer o menor MAPE mensal de validação. Com essa medida escolhe-se para representar a RNA do mês m a configuração de rede que conseguir a melhor generalização em relação à série de ENA do mês m . No decorrer do texto, essa medida de desempenho será referenciada como **MAPE val. mês**.

O MAPE do cenário de validação, calculado através da equação 3-8 que envolve todos os meses é uma outra medida de desempenho utilizada. Nesse caso, avalia-se o ajuste de todos os parâmetros do PEN e não a de uma CE m . Dentre todos os cenários gerados, um para cada conjunto de 12 configurações de RNAs do PEN, escolhe-se aquele conjunto de 12 RNAs que conseguir o menor MAPE de cenário de validação. Com essa medida de desempenho escolhe-se para representar a RNA do mês m uma configuração de rede baseada no comportamento das redes dos outros meses. Essa medida de desempenho será referenciada como **MAPE val. cen**.

Como o conjunto de validação contém apenas 5 anos, decidiu-se verificar o desempenho das RNAs em relação a um número maior de padrões de dados. Por essa razão, unificou-se então os conjuntos de padrões de treinamento e validação, como ilustra a Figura 4.6, para calcular outras medidas de

desempenho do modelo.

Dados de entrada					Dados de saída
$Z(k-s)$	$Z(k-1)$	$Z(k-2)$...	$Z(k-p_m)$	$Z(k)$

Conjunto de
Treinamento

Conjunto de
Validação

Figura 4.6: Unificação dos Padrões de Dados de Treinamento e Validação

Com todas as 80 configurações de RNAs treinadas criou-se, com esse padrão de dados unificado, um cenário de forma análoga ao cenário de validação, como apresenta a Figura 4.7. Esse cenário criado com os padrões de dados unificados é disparado com as entradas do primeiro ano, gerando valores até se completar 74 anos de cenário, que corresponde à quantidade de dados de padrão unificado para a avaliação de desempenho. Posteriormente, o cenário é comparado com as saídas desejadas do padrão de dados unificado. Da mesma forma que ocorre com o MAPE de validação, é possível obter dois tipos de MAPE no cenário de padrões de dados unificados:

1. **Por cenário:** calcula-se o MAPE percorrendo serialmente o cenário.
2. **Por mês:** compara-se os 74 valores do mês m presentes no cenário com as 74 saídas do mês m presentes no padrão de dados unificado.

Contudo, como foi visto, o treinamento das RNAs pode ser interrompido de 2 formas, e em cada tipo de parada é possível salvar uma RNA para cada mês m (conforme as Figuras 3.11 e 3.10). Logo decidiu-se criar o cenário de padrão de dados unificado utilizando os dois tipos de RNAs:

- O conjunto das 12 RNAs salvas na mesma época. Nesse caso o cenário será referenciado como **cenário unificado**.
- O conjunto das 12 RNAs salvas, independentemente, em épocas distintas. Nesse caso o cenário será referenciado como **cenário unificado de redes diferentes**.

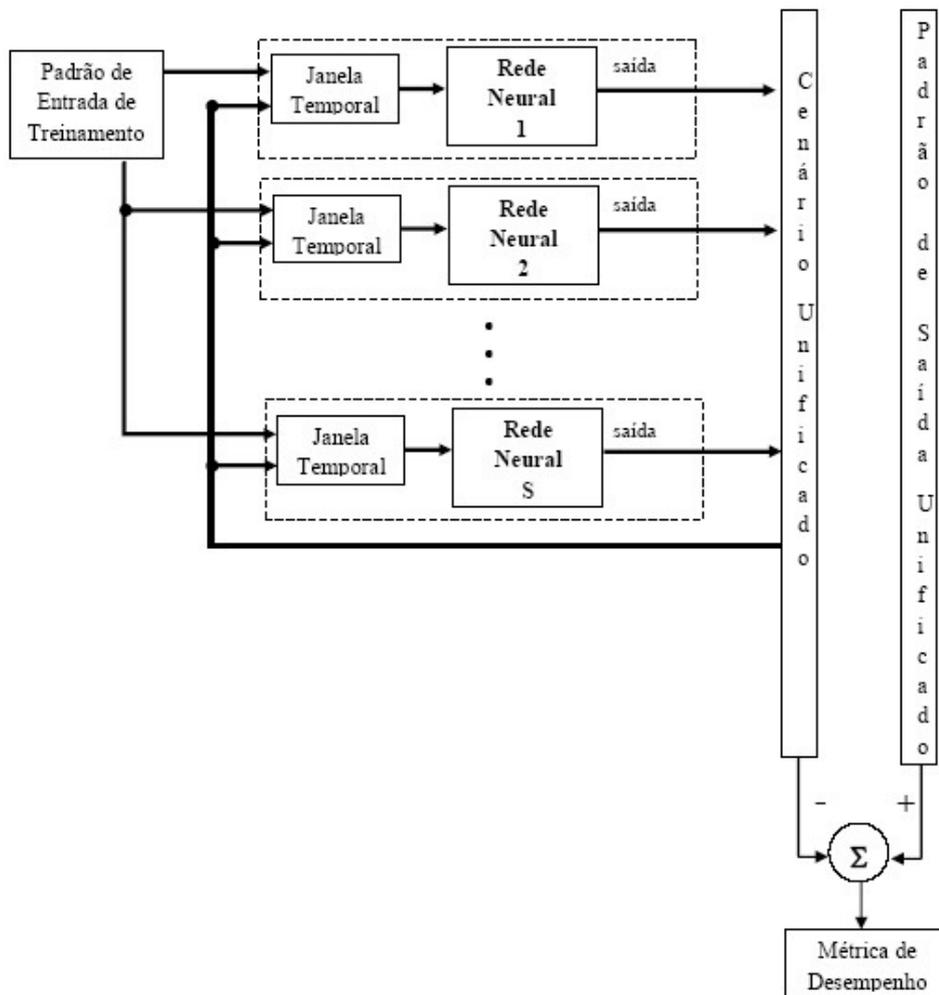


Figura 4.7: Geração do Cenário Unificado

A partir de cada um desses cenários pode-se obter: um MAPE individual por mês e um MAPE de cenário. As medidas de desempenho do cenário unificado são referenciadas, respectivamente, como **MAPE uni. mês** e **MAPE uni. cen.** E as medidas de desempenho do cenário unificado de redes diferentes são referenciadas como **MAPE uni.dif. mês** e **MAPE uni.dif. cen.**

Obtém-se com cada configuração de RNA treinada uma série de resíduos de validação, que corresponde às diferenças entre as saídas da RNA ajustada à série de ENA do mês m com as saídas desejadas desse mês m presentes no conjunto de validação. A média da série de resíduos de validação de cada mês é também uma outra medida de desempenho utilizada na escolha da configuração da RNA de cada mês. O ideal é que os resíduos dessa série sejam nulos, para indicar que a RNA consegue generalizar muito bem, isto é, que a RNA acertou a resposta de padrões desconhecidos em 100%. Porém, isso nunca ocorre na prática, então a idéia é escolher a média que se aproxime mais do valor nulo.

Dessas medidas de desempenho também podem ser obtidas: uma média mensal de resíduos de validação, uma por RNA; e uma média dos resíduos obtidos através do cenário de validação, referentes à diferença $Z(i) - C(i)$ presente na equação 3-8. Assim, essas medidas de desempenho são denominadas **Média val. mês** e **Média val. cen.** De maneira análoga, decidiu-se criar uma série de resíduos mensais através da diferença entre o cenário de padrões unificados e a saída desejada presente no conjunto de padrões de dados unificados. Como foram criados dois tipos de cenários de padrões unificados, referenciados como cenário unificado e cenário unificado redes diferentes, para cada um dos cenários calculou-se as séries de resíduos mensais. Dessas séries de resíduos obtidas, decidiu-se calcular uma média mensal e uma média do cenário. Para o cenário unificado as medidas de desempenho foram denominadas **Média uni. mês** e **Média uni. cen.** Já para o cenário unificado de redes diferentes as medidas de desempenho foram denominadas **Média uni.dif. mês** e **Média uni.dif. cen.**

Resumindo, foram estabelecidas 12 medidas de desempenho para serem analisadas, com o intuito de escolher uma configuração de rede para representar a RNA da componente estocástica (CE) do mês m que irá compor o $PEN(p, l)$:

1. **MAPE val. mês:** MAPE mensal calculado individualmente por mês;
2. **MAPE val. cen:** MAPE calculado sobre todo o cenário de validação criado;
3. **MAPE uni. mês:** MAPE mensal calculado individualmente por mês no cenário unificado;
4. **MAPE uni. cen:** MAPE calculado sobre todo o cenário unificado criado;
5. **MAPE uni.dif. mês:** MAPE mensal calculado individualmente por mês no cenário unificado redes diferentes;
6. **MAPE uni.dif. cen:** MAPE calculado sobre todo o cenário unificado redes diferentes;
7. **Média val. mês:** Média mensal dos resíduos de validação, individual por mês;
8. **Média val. cen:** Média dos resíduos do cenário de validação;
9. **Média uni. mês:** Média mensal dos resíduos do cenário unificado, calculado individualmente por mês;

10. **Média uni. cen:** Média dos resíduos do cenário unificado;
11. **Média uni.dif. mês:** Média mensal dos resíduos do cenário unificado redes diferentes, calculado individualmente por mês;
12. **Média uni.dif. cen:** Média dos resíduos do cenário unificado redes diferentes.

4.1.3

Validação dos resíduos gerados

Uma vez definida a RNA do mês m , utilizando uma das medidas de desempenho descritas na seção anterior, ajusta-se a série de resíduos obtidos entre a saída da RNA e os padrões de saída do conjunto de treinamento à uma distribuição teórica de probabilidade. Verifica-se a aderência entre a série de resíduos e a distribuição teórica de probabilidade através do teste de Kolmogorov-Smirnov. A distribuição teórica de probabilidade que conseguir o menor erro de estimativa através desse teste é adotada como a distribuição de probabilidade dos resíduos que compõem a componente estocástica (CE) do mês m no Processo Estocástico Neural (PEN).

Como os padrões de treinamento são formados por 69 anos do histórico, tem-se apenas 69 valores de resíduo para se fazer o ajuste a uma distribuição de probabilidade teórica, o que pode não ser uma quantidade de dados suficiente. Lembrando que cada configuração de RNA foi treinada 10 vezes, com a finalidade de avaliar diferentes inicializações de pesos sinápticos, tem-se para cada treinamento de uma configuração 69 resíduos de treinamento. Dessa forma, decidiu-se testar 2 formas de ajustar os resíduos a uma distribuição de probabilidade teórica:

1. ajustando os 69 resíduos obtidos no treinamento da RNA usada no mês m ;
2. ajustando os 690 resíduos obtidos dos 10 treinamentos da configuração da RNA do mês m .

Para cada tipo de ajuste de distribuição determina-se todas as distribuições de probabilidade do modelo PEN, uma para cada mês. Dessa forma, obtém-se dois tipos de modelo $PEN(p, l)$, cada um com um conjunto de distribuições de probabilidade. Cada modelo $PEN(p, l)$ é utilizado para gerar diferentes séries sintéticas de ENA. As séries sintéticas são geradas através do encadeamento das redes como ilustra a Figura 4.5, onde apenas as setas grossas é que fazem parte do fluxo de geração de cenários. Cada série $Z(t)$ de ENA

é gerada pela equação 3-5, que corresponde a:

$$Z(t) = \varphi_{out} \left(\sum_{i=1}^{l_m} \omega_{out,i} \cdot \varphi \left[\omega_{i,0} Z(t-s) + \left(\sum_{j=1}^{p_m} \omega_{i,j} Z(t-j) \right) + \theta_i \right] + \theta_{out} \right) + \alpha(t)$$

O valor aleatório $\alpha(t)$ presente nesta equação, para o mês m , consiste de um sorteio oriundo da distribuição de probabilidade cujos resíduos do mês m se ajustaram. Para realizar esse sorteio utiliza-se o método de amostragem Quase Monte Carlo com seqüência de Sobol (apresentado na Seção 2.2), a fim de capturar elementos uniformemente distribuídos dentro da distribuição de probabilidade. Como a equação 3-5 representa um valor de ENA, esta equação não pode resultar em valores negativos. Assim, sorteia-se valores para $\alpha(t)$ até que a restrição $Z(t) = y_{out} + \alpha(t) > 0$ seja atendida, sendo y_{out} a saída da RNA do mês m que o instante de tempo t representa.

4.2

Mantendo a correlação espacial entre os Subsistemas na série de Afluência Mensal

4.2.1

Modelo PAR(p) na Geração de Séries Sintéticas de Afluências Mensais

O modelo atualmente utilizado para representar as incertezas hidrológicas no planejamento da operação energética de médio prazo é o PAR(p), descrito na Seção 2.1.2. O PAR(p) gera cenários de séries hidrológicas de intervalo de tempo mensal, e, uma importante característica desse modelo é que ele preserva não só a estrutura de correlação temporal mensal das afluências mas também a correlação espacial entre os subsistemas do Sistema Interligado Nacional (SIN) [38].

Cada série sintética gerada corresponde a um sorteio, realizado através do modelo PAR(p_m) ajustado para o estudo do comportamento estocástico da série de Energia Natural Afluente (ENA) de um dado subsistema.

Manipulando a equação 2-10 do modelo, é possível isolar o valor das vazões, como ilustra a expressão 4-3 [53]:

$$Z_t = \mu_m + \phi_1^m \cdot \sigma_m \cdot \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \cdot \sigma_m \cdot \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m \cdot a_t \quad (4-3)$$

onde,

Z_t : é uma série sazonal de 12 períodos;

s : é o total de períodos ($s = 12$);

t : é o índice de tempo, $t = 1, 2, \dots, s \cdot r$

r : é o número de anos;

m : é o mês ($m = 1 \dots 12$);

μ_m : é a média histórica do mês m ;

σ_m : é o desvio padrão histórico do mês m ;

p_m : é a ordem do modelo auto-regressivo do mês m ;

ϕ_i^m : é i -ésimo coeficiente auto-regressivo do mês m ;

a_t : série de resíduos independentes com média zero e variância $\sigma_a^{2(m)}$.

O modelo de geração de séries sintéticas é aplicado diretamente à série original e é capaz de lidar com resíduos que apresentam um forte coeficiente de assimetria. Ocorre que em sistemas com reservatórios em cascata, as vazões incrementais em um período podem ser muito pequenas [38], e para evitar que a equação 4-3 gere vazões mensais negativas, ajustou-se uma distribuição Lognormal com três parâmetros aos resíduos mensais a_t [37].

Assim, a variável ξ_t da equação 4-4 têm distribuição Normal com média μ_ξ e variância igual a $\sigma_\xi^{2(m)}$.

$$\xi_t = \ln(a_t - \Delta) \quad (4-4)$$

onde a_t é o resíduo mensal e o parâmetro Δ é estimado segundo a equação 4-5 [53].

$$\Delta = - \left(\frac{\mu_m}{\sigma_m} \right) - \phi_1^m \cdot \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) - \dots - \phi_{p_m}^m \cdot \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \quad (4-5)$$

Definindo b_t como elemento de uma distribuição Normal com média zero e variância 1 ($\sim N(0, 1)$), tem-se a equação 4-6 [49].

$$b_t = \frac{\xi_t - \mu_\xi}{\sigma_\xi} \quad (4-6)$$

Os parâmetros μ_ξ e $\sigma_\xi^{2(m)}$ são estimados de forma a preservar os momentos dos resíduos, através das relações dadas pelas equações 4-7, 4-8 e 4-9 [13]:

$$\mu_\xi = 0,5 \cdot \ln \left(\frac{\sigma_a^m}{\theta \cdot (\theta - 1)} \right) \quad (4-7)$$

$$\sigma_\xi^{2(m)} = \ln(\theta) \quad (4-8)$$

$$\theta = 1 + \frac{\sigma_a^{2(m)}}{\Delta^2} \quad (4-9)$$

onde $\sigma_a^{2(m)}$ é estimada utilizando-se a expressão dada pela equação 2-16:

$$\sigma_a^{2(m)} = 1 - \phi_1^m \cdot \rho^m(1) - \phi_2^m \cdot \rho^m(2) - \dots - \phi_{p_m}^m \cdot \rho^m(p_m)$$

Os resíduos ξ_t são espacialmente não correlacionados, porém, para se gerar afluições mensais multivariadas, é necessário transformá-los em resíduos espacialmente correlacionados, representados por W_t na equação 4-10 [53]:

$$W_t = D \cdot \xi_t \quad (4-10)$$

onde D é uma matriz quadrada de dimensão igual ao número de subsistemas equivalentes no planejamento de médio prazo [38]. Ela pode ser considerada uma matriz triangular inferior e ser estimada por decomposição espectral ou através da equação 4-11 [37]:

$$D \cdot D^T = U \quad (4-11)$$

sendo D^T a matriz transposta de D e U a estimativa da matriz de correlação espacial entre as ENAs dos reservatórios equivalentes dos pares de subsistemas. Portanto, um elemento $u_{i,j}$ da matriz U é a correlação espacial de *lag* zero entre as ENAs dos reservatórios equivalentes dos subsistemas i e j .

A equação 4-11 pode ser resolvida através da obtenção das matrizes de auto-valores e auto-vetores da matriz U , como ilustra a equação 4-12 [45].

$$U = P \cdot L^{1/2} \cdot L^{1/2} \cdot P^T \quad (4-12)$$

onde P e L são, respectivamente, as matrizes de auto-vetores e auto-valores de U . Através das equações 4-11 e 4-12 obtém-se a matriz de carga $D = P \cdot L^{1/2}$ [49].

Isolando o resíduo mensal a_t na equação 4-4, obtém-se a equação 4-13:

$$a_t = e^{\xi_t} + \Delta \quad (4-13)$$

Colocando ξ_t em evidência na equação 4-6 e substituindo na equação 4-13, obtém-se:

$$a_t = e^{b_t \cdot \sigma_{\xi_t} + \mu_{\xi_t}} + \Delta$$

Como se deseja manter a correlação espacial entre as ENAs, utiliza-se o resíduo espacialmente correlacionado, definido em 4-10, no lugar do valor aleatório $\sim N(0, 1)$, obtendo-se a equação 4-14, a qual fornece os resíduos usados pelo modelo PAR(p) para gerar séries sintéticas de afluições mensais multivariadas positivas.

$$a_t = e^{W_t \cdot \sigma_{\xi_t} + \mu_{\xi_t}} + \Delta \quad (4-14)$$

Portanto, para a geração de cenários de Energia Natural Afluente utilizando a distribuição Lognormal de três parâmetros, deve-se seguir os seguintes passos para cada instante de tempo t :

1. Sortear os valores $b_t \sim N(0, 1)$;
2. Utilizar esses valores na equação 4-6 para encontrar os resíduos ξ_t ;
3. Multiplicar esses resíduos pela matriz D para encontrar os resíduos correlacionados W_t como informa a equação 4-10;
4. Através das equações 4-5, 4-7, 4-8 e 4-9 encontrar os valores dos parâmetros da lognormal que fornece o resíduo mensal a_t como demonstra a equação 4-14;
5. Adicionar o resíduo a_t na série sintética Z_t conforme informa a equação 4-3.

4.2.2

Mantendo a correlação espacial entre os Subsistemas nos cenários gerados pelo Processo Estocástico Neural

O processo estocástico neural (PEN) na sua formação original não preserva a correlação espacial entre os subsistemas. Ajusta-se um PEN para cada subsistema e as RNAs do PEN de um dado subsistemas são treinadas independentemente das RNAs de outro subsistema, ou seja, o resíduo $\alpha(t)$ presente na equação 3-4 é individual por mês e por subsistema.

Afim de manter a correlação espacial entre os subsistemas, decidiu-se implementar duas variações do PEN:

1. Ajustando uma distribuição Lognormal com três parâmetros para substituir o resíduo mensal α
2. Aglomerando os 4 subsistemas em um único PEN, fazendo as suas RNAs terem 4 saídas, uma para cada subsistema.

4.2.3

Processo Estocástico Neural com distribuição Lognormal de três parâmetros

A idéia é utilizar como resíduo mensal um valor vindo da distribuição Lognormal de três parâmetros ajustada ao modelo PEN. Esse ajuste é feito de forma análoga ao modelo PAR(p), apresentado na Seção 2.1.2.

Na Seção 4.2.1 foi apresentado um esquema para gerar cenários de vazão utilizando a distribuição Lognormal de três parâmetros, a ser seguido em cada instante de tempo t .

Apenas no passo 2, é preciso realizar modificações para se aplicar a distribuição Lognormal de três parâmetros no modelo Processo Estocástico Neural (PEN), onde a equação 4-6 consiste em estimar valores para os parâmetros média μ_ξ e variância $\sigma_\xi^{2(m)}$ dos resíduos ξ . Esses parâmetros, μ_ξ e $\sigma_\xi^{2(m)}$, são estimados de forma a preservar os momentos dos resíduos, através das equações 4-8 e 4-7, que dependem dos parâmetros Δ e $\sigma_a^{2(m)}$.

O parâmetro Δ corresponde ao terceiro parâmetro da distribuição Lognormal, que tem o objetivo de garantir que no processo de geração de séries sintéticas não ocorram valores negativos de energia natural afluyente (ENA) aos subsistemas. O parâmetro Δ toma a forma de um limite inferior para os ruídos, ajustável a cada situação da parcela determinística do modelo [53]. Logo, ele pode ser estimado segundo a equação 4-5. Observe que essa equação é formada pela soma do valor negativo dos termos auto-regressivos da série, num dado mês, ponderados pelos parâmetros do modelo.

No modelo PEN, a saída de cada rede neural artificial (RNA) que o compõe é dada pelos termos auto-regressivos da série (do mês que a rede representa) ponderados por parâmetros do modelo (os pesos sinápticos da rede), como mostra a equação 3-3. Por isso, mantendo a analogia, no PEN o parâmetro Δ é estimado pela equação 4-15.

$$\Delta = -y_{out}(t) \quad (4-15)$$

onde y_{out} é a saída da RNA do mês m determinado pelo instante de tempo t .

O parâmetro $\sigma_a^{2(m)}$, segundo [30, 53], é estimado pela equação 2-16. Através de experimentos verificou-se que esse parâmetro pode ser aproximado pela variância dos resíduos obtidos como a diferença entre uma previsão feita pelo PAR(p) e o valor real dessa previsão, obtida do histórico.

Mantendo a analogia, no PEN os resíduos de diferença entre previsão e valor real são dados pela série de resíduos mensais de treinamento de cada RNA. Portanto, o parâmetro $\sigma_a^{2(m)}$ no PEN é fornecido pela variância dos resíduos mensais do conjunto de treinamento da RNA do mês m .

Com base nisso, seguem-se todos os passos até o passo 5, onde o resíduo mensal a_t obtido irá substituir o resíduo mensal $\alpha(t)$ na equação 3-4.

Apesar de não ser a melhor forma de se embutir a correlação espacial entre os subsistemas nos cenários do Processo Estocástico Neural, essa modelagem foi avaliada por ser de fácil implementação e para fazer uma análise de como a correlação espacial influencia a aderência dos cenários com as séries históricas de ENA.

4.2.4

Processo Estocástico Neural com Redes Neurais de 4 saídas

Outra idéia é deixar que as próprias redes neurais (RNAs) do processo estocástico neural (PEN) aprendam a correlação espacial entre os subsistemas. Para isso modela-se um único PEN para todos os subsistemas e as RNAs do PEN fornecem como saída as séries sintéticas dos 4 subsistemas.

A Figura 4.8 ilustra a RNA do mês de janeiro. Observe que a RNA recebe como entradas 4 tipos de janela temporal, uma para cada subsistema. Cada janela temporal é formada por valores passados da série de ENA do subsistema ao qual ela representa. O número de termos passados pode ser independente por subsistema, mas nesse trabalho deixou-se idênticos. Ou seja, ao definir que a RNA do mês m tem ordem p_m , todas as janelas temporais usam p_m como a quantidade de termos passados da sua série.

A Figura 4.9 ilustra como fica um neurônio na camada oculta, sendo que Z_1 representa a série do Sudeste, Z_2 a série do Sul, Z_3 a série do Nordeste e Z_4 a série do Norte. A saída desse neurônio é dada pela equação 4-16.

$$y_i = \varphi \left(\sum_{\gamma=1}^4 [\omega_{i,(\gamma-1) \cdot p_m + (\gamma-1)} \cdot Z_\gamma(t-s) + \lambda] + \theta_i \right) \quad (4-16)$$

sendo

$$\lambda = \left(\sum_{j=1}^{p_m} \omega_{i,(\gamma-1) \cdot p_m + (\gamma-1) + j} \cdot Z_\gamma(t-j) \right)$$

onde φ é a função de ativação do neurônio i , γ representa o subsistema, p_m é a ordem do mês m , $\omega_{i,j}$ é o peso sináptico da conexão entre a entrada j e o neurônio i , com $j = 0 \dots 4 \cdot p_m + 3$, $Z_\gamma(t-j)$ é o valor da série do subsistema γ no instante de tempo $t-j$, $s = 12$ é o total de meses e θ_i é o bias desse neurônio.

A RNA possui 4 neurônios na camada de saída, e cada um fornece a saída para uma série Z_γ , onde $\gamma = 1 \dots 4$ representa os subsistemas. Considerando que essa RNA contém L neurônios na camada oculta, a saída de cada neurônio da camada de saída é dada pela equação 4-17.

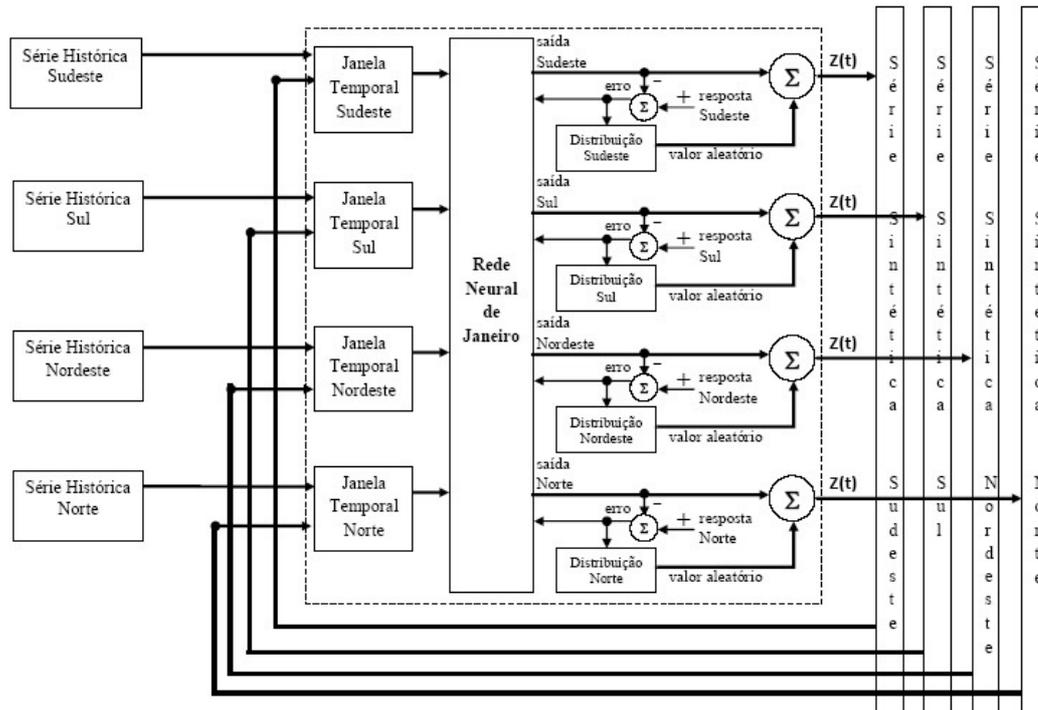


Figura 4.8: Rede Neural de Janeiro com 4 saídas

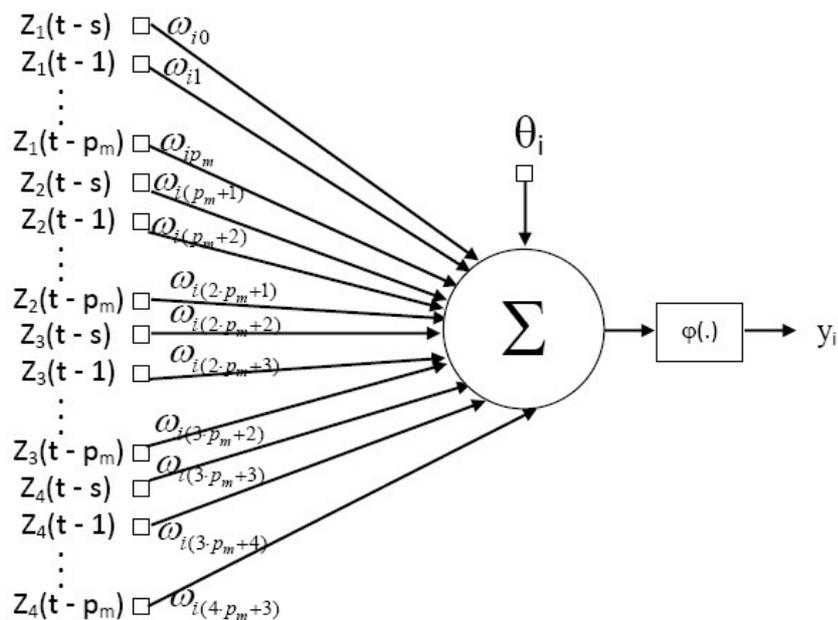


Figura 4.9: Neurônio da camada oculta da rede neural de 4 saídas

$$y_{out}(\gamma) = \varphi_{out} \left(\sum_{i=1}^L \omega_{out,(\gamma-1) \cdot L+i} \cdot y_i + \theta_{out}(\gamma) \right) \quad (4-17)$$

onde $y_{out}(\gamma)$ é a saída do neurônio do subsistema γ , φ_{out} é a função de ativação dos neurônios da camada de saída representados por out , $\omega_{out,(\gamma-1) \cdot L+i}$ é o peso sináptico da conexão entre a entrada i (que corresponde a saída do neurônio i da camada oculta) e o neurônio out do subsistema γ e $\theta_{out}(\gamma)$ é o bias do neurônio do subsistema γ .

Durante o treinamento dessa RNA, para cada padrão de dados entrada-saída k , é calculado o erro entre o valor real da série do subsistema γ , representado aqui por $d(k)$, e a saída da rede fornecido para o subsistema γ ao receber as entradas do k -ésimo padrão de dados. Esse erro é calculado pela equação 4-18.

$$e_{out}(\gamma)(k) = d(\gamma)(k) - y_{out}(\gamma)(k) \quad (4-18)$$

Portanto, é gerada uma série de resíduos para cada subsistema γ e para cada série é ajustada uma distribuição de probabilidade teórica. Da distribuição do subsistema γ é sorteado um valor aleatório para ser adicionado à saída do neurônio desse subsistema para compor a sua série sintética, de acordo com a equação 4-19.

$$Z_{\gamma}(t) = y_{out}(\gamma) + \alpha_{\gamma}(t) \quad (4-19)$$

onde $Z_{\gamma}(t)$ é a série do subsistema γ gerada no instante de tempo t e $\alpha_{\gamma}(t)$ é o valor aleatórios proveniente da distribuição de probabilidade ajustada aos resíduos obtidos da diferença entre a saída deseja e a saída da RNA do subsistema γ , representada aqui por $y_{out}(\gamma)$.

Essa modelagem é feita para as 12 RNAs do PEN, uma para cada mês. Observe que as séries dos subsistemas são simuladas em paralelo através da execução seqüencial da equação 4-19. Isto é, a cada instante de tempo t definido por:

$$t = (r - 1) \cdot s + m$$

onde $r = 1 \dots n$ é o número de anos a ser gerados, $m = 1 \dots s$ é o mês, $s = 12$ é o total de meses das séries, a RNA do mês m executa a equação 4-19 para gerar as séries dos 4 subsistemas no mês m do ano r .

4.3 Resumo

Nesse capítulo foi apresentada a modelagem do Processo Estocástico Neural para ser utilizado na geração de cenários de afluências mensais, in-

cluindo uma versão adaptada com uma distribuição de probabilidade Lognormal de três parâmetros, a qual incorpora a correlação espacial entre os subsistemas nos resíduos usados na geração de séries sintéticas de vazão, e uma versão que incorpora numa mesma rede neural do PEN as entradas/saídas de todos os subsistemas do SIN.

O próximo capítulo apresenta os resultados obtidos no estudo de caso dessa tese, mostrando as configurações dos modelos utilizados para a geração de cenários de energia natural afluyente (PAR(p) e PEN) e os resultados das análises desses cenários.