

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA
FILHO”**

**FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**REDES NEURAIS ARTIFICIAIS COMO FERRAMENTA PARA
PROGNOSE DE CRESCIMENTO E MELHORAMENTO
GENÉTICO FLORESTAL**

William de Medeiros Silva

Engenheiro Florestal

2019

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA
FILHO”**

**FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**REDES NEURAIS ARTIFICIAIS COMO FERRAMENTA PARA
PROGNOSE DE CRESCIMENTO E MELHORAMENTO
GENÉTICO FLORESTAL**

William de Medeiros Silva

Orientador: Prof. Dr. Rinaldo Cesar de Paula

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Agronomia (Genética e Melhoramento de Plantas).

2019

S586r

Silva, William de Medeiros

Redes neurais artificiais como ferramenta para prognose de crescimento e melhoramento genético florestal / William de Medeiros Silva. -- Jaboticabal, 2019

75 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Rinaldo Cesar de Paula

1. Melhoramento florestal. 2. Redes neurais artificiais. 3. BLUP. 4. Interação G x A. 5. Prognose da produção floresta. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TÊSE: REDES NEURAIS ARTIFICIAIS COMO FERRAMENTA PARA PROGNOSE DE CRESCIMENTO E MELHORAMENTO GENÉTICO FLORESTAL

AUTOR: WILLIAM DE MEDEIROS SILVA

ORIENTADOR: RINALDO CESAR DE PAULA

Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA (GENÉTICA E MELHORAMENTO DE PLANTAS), pela Comissão Examinadora:


Prof. Dr. RINALDO CESAR DE PAULA
Departamento de Produção Vegetal (Fitotecnia) / FCAV / UNESP - Jaboticabal


Prof. Dr. BRUNO ETTORE PAVAN
Departamento de Fitotecnia, Tecnologia de Alimentos e Sócio Economia / Ilha Solteira/SP


Prof. Dr. GUSTAVO VITTI MÓRO
Departamento de Produção Vegetal (Fitotecnia) / FCAV / UNESP - Jaboticabal


Pesquisador Dr. PAULO HENRIQUE MÜLLER DA SILVA
Instituto de Pesquisas e Estudo Florestais-IPEF / Piracicaba/SP


Prof. Dr. SANDRA HELENA UNEDA TREVISOLI
Departamento de Produção Vegetal (Fitotecnia) / FCAV / UNESP - Jaboticabal

Jaboticabal, 09 de agosto de 2019

DADOS CURRICULARES DO AUTOR

WILLIAM DE MEDEIROS SILVA – filho de Maria Inês de Medeiros e Raimundo Nonato Guedes da Silva, nascido em 07 de agosto de 1985 em Taubaté, interior do estado de São Paulo. Ingressou no curso de Engenharia Florestal na Universidade Federal do Piauí, Câmpus de Bom Jesus (UFPI/CPCE), em abril de 2007. Como aluno de graduação estagiou nas áreas de silvicultura e viveiros florestais, foi monitor da disciplina de melhoramento florestal e presidente do Centro Acadêmico de Engenharia Florestal da UFPI/CPCE. Obteve o título de Engenheiro Florestal em novembro de 2013. Em março de 2014 ingressou no curso de Pós-graduação em Agronomia (Sistemas de Produção Vegetal), nível Mestrado, pela Universidade Estadual Paulista “Julio de Mesquita Filho” - Faculdade de Engenharia de Ilha Solteira (UNESP/FEIS), quando trabalhou com tamanho de parcelas experimentais e seleção simultânea de caracteres em progênies de *Eucalyptus camaldulensis* DEHNN, sendo bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), tendo como orientador o Prof. Dr. Bruno Ettore Pavan, obtendo o título de Mestre em fevereiro de 2016. Em março de 2016 ingressou no curso de Pós-graduação em Agronomia (Genética e Melhoramento de Plantas), nível Doutorado, pela Universidade Estadual Paulista “Julio de Mesquita Filho” – Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal (UNESP/FCAV), sendo bolsista CAPES, tendo como orientador o Prof. Dr. Rinaldo Cesar de Paula. É chefe assistente no Grupo Escoteiro Carcarás de Jaboticabal (339-SP) realizando trabalho voluntário na comunidade.

“Venha, procure, porque a busca é o
alicerce da sorte: todo sucesso depende
de colocar o coração naquilo que você
quer”

(Rumi)

AGRADECIMENTOS

Aos meus pais Nonato e Inês, pelo apoio e amor incondicionais. Vocês são minha fortaleza e refúgio.

À minha querida esposa Bruna, por ter sonhado junto comigo e enfrentado cada obstáculo durante a caminhada até aqui, se mantendo firme ao meu lado todos esses anos, me dando forças para superar os momentos de dificuldade e me fazendo acreditar que chegaríamos ao final desta difícil e gratificante etapa.

À minha amada filha Giovanna, o melhor de mim, que com sua doçura torna as coisas fáceis mesmo quando a vida fica difícil. Poder te dar esse exemplo de conquista vale mais do que qualquer título.

Ao Professor Dr. Rinaldo Cesar de Paula, pelos ensinamentos, pelo exemplo de profissionalismo e dedicação, pela amizade e pela orientação, que muitas vezes foi além dos assuntos acadêmicos.

Ao amigo Roberto Carlos, funcionário do Horto Florestal da Unesp de Jaboticabal, pela ajuda e apoio na condução do experimento de campo, pelos ensinamentos práticos e pela amizade.

Aos colegas do Laboratório de Sementes e Melhoramento Florestal, Bruna, Patrícia, Tamara, Robson e, em especial, aos amigos João Pedro, João Vitor e Márcio Araújo pela ajuda nos trabalhos de campo e laboratório, pelo bom convívio e pelos momentos de descontração.

À Universidade Estadual Paulista “Julio de Mesquita Filho” – Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal – Unesp/FCAV, e ao Programa de Pós-Graduação em Agronomia (Genética e Melhoramento de Plantas) pelos ensinamentos oferecidos.

Aos professores da Unesp/FCAV, em especial aos do Programa de Pós-Graduação em Agronomia (Genética e Melhoramento de Plantas), com os quais tive a oportunidade de aprender muito durante as disciplinas que cursei.

Ao Instituto de Pesquisas e Estudos Florestais (IPEF) e ao programa de Tolerância de *Eucalyptus* Clonais aos Estresses Hídrico, Térmico e Biótico (TECHS) pelo fornecimento dos dados.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Enfim, a todos que contribuíram de alguma forma com o desenvolvimento deste trabalho e com minha formação profissional e pessoal. Meus sinceros agradecimentos!

SUMÁRIO

RESUMO	iii
ABSTRACT	iv
CAPÍTULO 1 – Considerações gerais	1
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1. O Setor Florestal Brasileiro e a Cultura do Eucalipto	3
2.2. Prognose da Produção Florestal	4
2.3. Seleção de Genótipos	5
2.4. Interação G x A e Análise de Estabilidade e Adaptabilidade	7
2.5. Redes Neurais Artificiais (RNA)	8
2.5.1. Neurônios biológicos e neurônios artificiais	9
2.5.2. Perceptrons de múltiplas camadas	11
2.5.3. Algoritmo de retropropagação do erro (<i>backpropagation</i>)	12
2.6. Redes Neurais Artificiais Aplicadas à Ciência Florestal	14
3. REFERÊNCIAS	15
CAPÍTULO 2 – Prognose do diâmetro de árvores individuais de eucalipto em diversas condições climáticas utilizando redes neurais artificiais	21
RESUMO.....	21
CHAPTER 2 – Prognosis of the diameter of individual eucalyptus trees in various climatic conditions using artificial neural networks	22
ABSTRACT	22
1. Introdução	23
2. Material e métodos.....	24
3. Resultados e discussão.....	29
4. Conclusão	33
5. Referências	34
CAPÍTULO 3 – Redes Neurais Artificiais na predição de valor genético em <i>Eucalyptus sp.</i>	37
RESUMO.....	37
CHAPTER 3 – Artificial Neural Networks in the prediction of genetic value in <i>Eucalyptus sp.</i>	38

ABSTRACT	38
1. Introdução	39
2. Material e métodos	40
3. Resultados e discussão.....	45
4. Conclusão	52
5. Referências	53
CAPÍTULO 4 – Redes Neurais Artificiais para seleção simultânea de progênies de eucalipto quanto à produtividade, estabilidade e adaptabilidade	56
RESUMO.....	56
CHAPTER 4 – Artificial Neural Networks for simultaneous selection of eucalyptus progenies for productivity, stability and adaptability	57
ABSTRACT	57
1. Introdução	58
2. Material e métodos.....	59
3. Resultados e discussão.....	63
4. Conclusão	70
5. Referências	71
CAPÍTULO 5 – Considerações Finais	74

REDES NEURAIS ARTIFICIAIS COMO FERRAMENTA PARA PROGNÓSE DE CRESCIMENTO E MELHORAMENTO GENÉTICO FLORESTAL

RESUMO – O eucalipto é a cultura de maior destaque para o setor florestal brasileiro. No entanto, a expansão do setor para áreas com condições climáticas limitantes ao desenvolvimento da cultura e a instabilidade climática atual, são alguns dos fatores que têm comprometido o desenvolvimento desta cultura no país nos últimos anos. Assim, é importante a busca contínua por ferramentas que possibilitem a prognose de crescimento, a seleção de indivíduos e famílias e a análise do comportamento de genótipos de eucalipto frente às variações ambientais de forma cada vez mais acurada. Desta forma, o objetivo geral deste trabalho foi testar o desempenho das Redes Neurais Artificiais (RNA) na modelagem de crescimento de clones de eucalipto, na predição de valores genéticos de indivíduos e famílias, e na seleção quanto à produtividade, estabilidade e adaptabilidade de progênies de *Eucalyptus sp.* Para a prognose de crescimento foram utilizados dados de 18 clones comerciais de *Eucalyptus* em diferentes estados do Brasil, e para a estimação de valor genético e análise de produtividade, estabilidade e adaptabilidade foram utilizados dados de testes de progênies de *Eucalyptus grandis*. Neste trabalho foram testadas diferentes arquiteturas de RNA do tipo múltiplas camadas com o algoritmo de aprendizado de retropropagação do erro e função de ativação do tipo tangente hiperbólica. O modelo desenvolvido para prognose do diâmetro à altura do peito (DAP) de árvores individuais em um local foi capaz de manter boa acurácia ao ser aplicado em outros locais. As RNA com três camadas ocultas foram as que apresentaram melhor desempenho na predição do valor genético de árvores individuais e de famílias, assim como na seleção das progênies quanto à produtividade, estabilidade e adaptabilidade. RNA do tipo múltiplas camadas proporcionam bom desempenho na prognose de crescimento, na predição de valor genético e na seleção quanto à produtividade, estabilidade e adaptabilidade em plantios de eucalipto, constituindo-se uma ferramenta promissora para uso nos programas de melhoramento florestal.

Palavras-chave: BLUP, inteligência artificial, Interação G x A, inventário florestal, melhoramento de plantas, perceptrons de múltiplas camadas

ARTIFICIAL NEURAL NETWORKS AS A TOOL FOR GROWTH PROGNOSIS AND FORESTRY GENETIC IMPROVEMENT

ABSTRACT – Eucalyptus is the most important crop of the most important for the Brazilian forest sector. However, the expansion of the sector to areas with climatic conditions limiting the development of the crop and current climate instability are some of the factors that have compromised the development of this culture in the country in recent years. Thus, it is important to continuously search for tools that allow the prognosis of growth, the selection of individuals and families and the analysis of the behavior of eucalyptus genotypes in the face of environmental changes in an increasingly accurate way. Thus, the general objective of this work was to test the performance of artificial neural networks (ANN) in the modeling of growth of eucalyptus clones, prediction of genetic values of individuals and families, and selection of productivity, stability and adaptability of progenies of *Eucalyptus sp.* For the prognosis of growth, data from 18 commercial Eucalyptus clones were used in different states of Brazil, and for genetic value estimation and productivity, stability and adaptability analysis data from *Eucalyptus grandis* progenies were used. In this work, different ANN architectures of the multilayer type were tested with the backpropagation error algorithm and hyperbolic tangent activation function. The model developed for prognosis of the diameter at breast height (DBH) individual trees in one place was able to maintain good accuracy when applied at other sites. The three-layered ANN was the ones that presented the best performance in predicting the genetic value of individual trees and families, as well as progeny selection for productivity, stability and adaptability. ANN of the multilayer type present good performance in growth prognosis, prediction of genetic value and selection in productivity, stability and adaptability in eucalyptus plantations, constituting a promising tool for use in forest improvement programs.

Keywords: BLUP, artificial intelligence, G x E interaction, forest Inventory, plant breeding, multilayer perceptrons

CAPÍTULO 1 – Considerações gerais

1. INTRODUÇÃO

Espécies arbóreas do gênero *Eucalyptus* são as mais plantadas no mundo (Becerra et al., 2018). No Brasil, a área cultivada com este gênero representa cerca de 72% das florestas plantadas no país, e vem aumentando a cada ano (IBÁ, 2017). Isto se deve, entre outros fatores, à boa capacidade produtiva, ao ciclo de corte relativamente curto, ao potencial para plantio em áreas com diferentes condições ambientais e às boas propriedades da madeira e fibra das espécies deste gênero (Maier et al., 2017). Além disso, os avanços nos programas de melhoramento genético e nas técnicas de manejo florestal têm contribuído significativamente, ao longo dos anos, para que a eucaliptocultura se firme como principal cultura florestal do país. No entanto, a estimativa de aumento da demanda por madeira nas próximas décadas (IBÁ, 2017), associada à redução da produtividade do eucalipto no Brasil nos últimos anos, devido principalmente a fatores ambientais (Gonçalves et al., 2013; Silva et al., 2016), exige que os pesquisadores busquem continuamente por novas ferramentas que possibilitem a manutenção do aumento da produtividade para suprir essa demanda crescente.

Um planejamento florestal adequado é importante para a otimização da produção das florestas, no entanto, o longo ciclo de corte das espécies florestais torna-se um complicador importante. Neste sentido, deve-se buscar continuamente por modelos de crescimento que possibilitem a prognose da produção florestal ao longo do tempo, de maneira cada vez mais precisa. As principais ferramentas utilizadas para este fim são os tradicionais modelos de regressão que, dependendo da categoria, permitem estimar o crescimento total do povoamento florestal, o crescimento por classe de diâmetro ou o crescimento de cada árvore individualmente (Campos e Leite, 2013).

Outro fator essencial para o aumento contínuo da produtividade florestal é a precisão na seleção e recomendação de genótipos superiores, e para tanto, é fundamental considerar o mérito genético das árvores no momento da seleção. Neste sentido, os pesquisadores buscam o uso de estimadores de parâmetros e

valores genéticos cada vez mais acurados. O procedimento REML/BLUP, que associa o método de estimação de componentes de variância REML (*Restricted Maximum Likelihood* – Máxima Verossimilhança Restrita), proposto por Patterson e Thompson (1971), ao método de estimação de valor genético BLUP (*Best Linear Unbiased Prediction* – Melhor Predição Linear não Viesada), proposto por Henderson na década de 40, está entre os mais utilizados para seleção de indivíduos e/ou famílias em programas de melhoramento florestal (Gallo et al., 2018; Kainer et al., 2018; Resende et al., 2018).

Sabendo-se que a resposta de um genótipo para um determinado caráter é resultado da combinação dos efeitos genéticos, ambientais e da interação genótipo x ambiente (G x A), ou seja, das respostas diferenciadas dos genótipos frente às variações ambientais, e considerando a instabilidade climática atual associada à tendência da expansão do setor florestal para novas fronteiras agrícolas, com condições edafoclimáticas cada vez mais específicas e limitantes ao desenvolvimento do eucalipto, faz-se necessário a realização de estudos que possibilitem a seleção de genótipos quanto à estabilidade e adaptabilidade, ou seja, a capacidade dos genótipos de apresentarem comportamento previsível frente às variações ambientais, e de responderem positivamente quando essas variações forem favoráveis (Santos et al., 2016). O método GGE Biplot, que possibilita a análise de adaptabilidade e estabilidade por meio de abordagem gráfica (Yan e Tinker, 2006), e o método da Média Harmônica da Performance Relativa dos Valores Genéticos (MHPRVG), que permite a seleção simultânea quanto à produtividade, estabilidade e adaptabilidade dos genótipos (Resende, 2007), têm-se destacado no melhoramento florestal (Correia et al., 2010; Engel et al., 2016; Oliveira et al., 2018; Santos et al., 2015).

Esses métodos têm em comum o fato de serem baseados em estatísticas convencionais e, portanto, apresentam certas exigências quanto a pressuposições em relação aos dados. No entanto, nem sempre é possível atender tais pressuposições, principalmente no caso de experimentos florestais, onde o período de duração e as áreas necessárias para realização dos mesmos são geralmente grandes. Assim, as Redes Neurais Artificiais (RNA) surgem como alternativa para os mais diferentes tipos de análises do setor florestal. As RNA são modelos inspirados

nos elementos básicos que formam a estrutura dos neurônios biológicos, e têm a capacidade de aprender por meio de um conjunto de exemplos e generalizar a informação dando respostas coerentes para dados não conhecidos, e diferentemente das análises estatísticas comumente utilizadas na experimentação florestal, não possuem nenhum tipo de pressuposição, uma vez que seus resultados dependem do aprendizado e não da distribuição dos dados em si (Cruz e Nascimento, 2018).

Assim, o objetivo geral deste trabalho foi testar o desempenho das RNA na modelagem de crescimento de povoamentos clonais de eucalipto em diferentes locais do Brasil, na predição de valores genéticos de indivíduos e famílias de *Eucalyptus*, e na seleção de progênies de *Eucalyptus* quanto à produtividade, estabilidade e adaptabilidade.

2. REVISÃO DE LITERATURA

2.1. O Setor Florestal Brasileiro e a Cultura do Eucalipto

O setor brasileiro de florestas plantadas, que apresenta a maior produtividade de madeira do mundo, atingiu um total de 7,84 milhões de hectares plantados em 2016, sendo responsável por mais de 90% de toda madeira produzida para fins industriais, por 6,1% do PIB industrial do país e pela geração de, aproximadamente, 3,7 milhões de postos de trabalho considerando os empregos diretos, indiretos e os gerados pelo efeito renda. Além do desenvolvimento econômico, as florestas plantadas desempenham, também, um papel importante na prestação de serviços ambientais, diminuindo a pressão sobre as florestas nativas brasileiras e conservando, para cada hectare de floresta plantada, 0,7 hectare em forma de área de preservação permanente, reserva legal, e/ou reserva particular de patrimônio natural, preservando assim a biodiversidade do país (IBÁ, 2017).

Dentre as culturas mais importantes para o setor florestal, a do eucalipto é a de maior destaque ocupando 72,7% (5,7 milhões de hectares) de toda a área de florestas plantadas no Brasil (IBÁ, 2017). Entretanto, a produtividade do eucalipto tem sido comprometida nos últimos anos no país devido, principalmente, à expansão

do setor florestal para áreas com condições climáticas cada vez mais específicas e limitantes ao desenvolvimento da cultura, e à instabilidade climática atual em áreas tradicionalmente cultivadas com eucalipto (Gonçalves et al., 2013; Silva et al., 2016). Além disso, apesar do Brasil ser o país com menor rotação do mundo para o eucalipto (IBÁ, 2017), o tempo também tem sido um fator crítico, dificultando diversos processos essenciais para o desenvolvimento desta cultura, como a seleção de novos materiais genéticos devido à demora dos ciclos de seleção, assim como o estabelecimento de planos adequados de manejo.

Estima-se que para atender a demanda por madeira para uso industrial e geração de energia nos próximos anos, em um cenário de baixo carbono, energias renováveis e desmatamento líquido zero, será necessário um plantio adicional de 250 milhões de hectares de eucalipto em todo o mundo até 2050 (IBÁ, 2017), e apesar do melhoramento genético e o manejo florestal terem contribuído muito ao longo dos anos para que o Brasil alcançasse a atual posição de destaque no cenário mundial entre os países mais eficientes na produção de eucalipto (Castro et al., 2016), fica evidenciada a necessidade de avanços contínuos nas técnicas de melhoramento e manejo florestal.

2.2. Prognose da Produção Florestal

Procedimentos que permitam antecipar o processo de seleção num programa de melhoramento florestal e facilitar o planejamento e o manejo das florestas são essenciais para o avanço da eucaliptocultura (Corrêa et al., 2017). Dentre estes procedimentos destacam-se a seleção precoce, em que os caracteres avaliados em idades juvenis servem como critério de seleção das árvores, permitindo a antecipação dos ganhos genéticos e tornando os programas de melhoramento mais eficientes em relação ao tempo e aos custos necessários (Massaro et al., 2010; Moraes et al., 2014; Pinto et al., 2014), e os modelos de crescimento que permitem a prognose da produção florestal ao longo do tempo, facilitando o planejamento e o manejo das florestas.

Os tradicionais modelos de regressão utilizados para prognose de crescimento florestal podem ser classificados como modelos em nível de

povoamento total, modelos de distribuição por classe diamétrica ou modelos em nível de árvores individuais, de acordo com o nível de detalhamento das informações (Campos e Leite, 2013). Os modelos em nível de povoamento total constituem a classe de modelos mais utilizada pelas empresas florestais e predizem o crescimento das florestas por unidade de área a partir de dados como idade do plantio, área basal, índice do local, entre outros (Azevedo et al., 2016; Campos e Leite, 2013; Castro et al., 2013). A modelagem de distribuição por classe diamétrica tem como foco principal a compreensão da dinâmica e do potencial de uso das florestas visando à obtenção de multiprodutos de madeira. Esta compreensão é obtida por meio do uso de uma função densidade de probabilidade que permite descrever o comportamento atual e futuro da distribuição dos diâmetros em classes com amplitudes previamente determinadas, de acordo com as características do povoamento (Araújo Júnior et al., 2013; Miranda et al., 2018; Schmidt et al., 2019). Já os modelos em nível de árvores individuais simulam o crescimento de cada árvore individualmente, e por meio da soma dos valores estimados das árvores é possível obter, por exemplo, a estimativa de crescimento por classe de diâmetro e o crescimento total do povoamento, com a vantagem de gerar maior detalhamento de informações sobre a dinâmica de crescimento da população. No entanto, a prognose de crescimento de florestas em nível individual ainda é pouco realizada em relação às demais categorias de modelos, e isso se deve principalmente à maior complexidade dos modelos de árvores individuais e à falta de estudos que comprovem a eficiência destes modelos para plantios comerciais (Castro et al., 2013; Martins et al., 2014).

2.3. Seleção de Genótipos

No melhoramento genético é fundamental que a população a ser melhorada disponha de variabilidade genética e que a frequência de alelos favoráveis, e conseqüentemente as suas combinações, aumentem a cada ciclo de seleção, permitindo que os ganhos com a seleção sejam contínuos ao longo dos ciclos. Para tanto, é importante que o método de seleção seja o mais acurado possível, e os métodos que consideram o mérito genético tendem a ser mais eficientes em relação

aos que levam em consideração apenas o fenótipo dos indivíduos (Paiva et al., 2002; Miranda et al., 2015).

Neste sentido, em testes de progênies de espécies florestais, são comumente empregados os métodos de seleção entre e dentro de famílias, no qual se identificam primeiramente as melhores famílias e posteriormente os melhores indivíduos dentro destas famílias, e o método de seleção combinada, baseado em um índice que leva em consideração o comportamento dos indivíduos e das famílias (Rosado et al., 2009). Esses dois métodos de seleção normalmente se baseiam em duas fontes de informação: o desvio individual em relação à média da família no bloco (efeito do indivíduo dentro da parcela), e o desvio da média da família em relação à média geral do experimento (efeito da progênie). No entanto, o modelo estatístico normalmente utilizado para descrever uma observação em um indivíduo em experimentos de campo, além destas informações, contém os efeitos da parcela e do bloco, e um procedimento ótimo de seleção, capaz de prever o valor genético dos indivíduos, é dado pelo índice multiefeitos (IME), que considera todos os efeitos aleatórios do modelo, possibilitando a maximização da correlação entre os valores genéticos verdadeiros e os preditos (Resende e Higa, 1994).

O IME, para o caso de dados balanceados, equivale ao método de predição de valor genético BLUP (Best Linear Unbiased Prediction – Melhor Predição Linear não Viesada), proposto por Henderson na década de 40, e muito utilizado atualmente no melhoramento florestal (Dias et al., 2018; Kainer et al., 2018; Silva et al., 2017). No entanto, para situações de desbalanceamento, o BLUP é mais adequado por ajustar os dados para os efeitos fixos e considerar as diferentes quantidades de dados associadas aos diferentes candidatos à seleção (Resende, 2002). Este método presume o conhecimento dos componentes de variância, mas como isso não é possível, são utilizadas as estimativas desses componentes, usualmente obtidas por meio do método REML (Restricted Maximum Likelihood – Máxima Verossimilhança Restrita), proposto por Patterson e Thompson (1971), ideal para a estimação de componentes de variância com dados desbalanceados. Assim, o método REML/BLUP é o procedimento ótimo para avaliação genética de espécies florestais, pois lidam naturalmente com as situações de desbalanceamento comumente encontradas em experimentos com plantas perenes (Resende, 2002).

2.4. Interação G x A e Análise de Estabilidade e Adaptabilidade

A resposta de um genótipo para um determinado caráter é resultado da combinação dos efeitos genotípicos, ambientais e da interação genótipo x ambiente (G x A). Esta interação é decorrente das diferentes respostas dos genótipos frente às mudanças ambientais e pode indicar que os melhores indivíduos em um determinado ambiente podem não sê-los em outro. Este efeito está presente mesmo quando os genótipos são avaliados em um único ambiente ou local, mas sua estimativa só é possível quando a avaliação é realizada em mais de um ambiente, e caso não seja considerada adequadamente, a interação G x A pode ser um complicador na seleção (Resende, 2002).

Apesar da sua importância, a simples estimativa da interação G x A não fornece informações detalhadas sobre o comportamento dos genótipos nos diferentes ambientes (Teodoro et al., 2016). Assim, para minimizar os efeitos negativos e, sempre que possível, aproveitar os efeitos positivos da interação, faz-se necessário a interpretação de análises que medem a sensibilidade diferencial dos genótipos aos diferentes ambientes, conectando esta variação a mecanismos biológicos específicos (Leon et al., 2016). Neste sentido, estudos que gerem informações sobre a capacidade dos genótipos apresentarem comportamento previsível frente às variações ambientais (estabilidade), e de responderem positivamente quando essas variações forem favoráveis (adaptabilidade), são essenciais para os programas de melhoramento genético (Santos et al., 2016).

Ao longo dos anos, pesquisadores têm-se dedicado ao desenvolvimento de ferramentas destinadas ao detalhamento de informações sobre as respostas dos genótipos diante das variações ambientais, e os procedimentos têm evoluído desde métodos baseados em análise de variância conjunta, passando por métodos baseados em regressão para análise de estabilidade e adaptabilidade, pelos métodos não paramétricos, métodos multiplicativos e métodos com abordagem dos modelos mistos (Resende, 2004).

Entre os métodos mais utilizados no melhoramento florestal para avaliação do comportamento dos genótipos frente às variações ambientais a partir de análise de estabilidade e adaptabilidade, destacam-se o método GGE Biplot, que consiste em

um sistema de análises que permite responder questões sobre o comportamento dos genótipos ao longo do gradiente ambiental por meio de abordagem gráfica (Correia et al., 2010; Santos et al., 2015; Yan e Tinker, 2006), e o método da Média Harmônica da Performance Relativa dos Valores Genéticos (MHPRVG), preditos pelo Método REML/BLUP (Resende, 2002), que permite a seleção simultânea quanto à produtividade, estabilidade e adaptabilidade dos genótipos (Dias et al., 2018; Engel et al., 2016; Oliveira et al., 2018).

2.5. Redes Neurais Artificiais (RNA)

A inteligência computacional é a área da ciência que busca o desenvolvimento de sistemas inteligentes por meio de técnicas inspiradas na natureza, como a Lógica Fuzzy, a Computação Evolucionária e as Redes Neurais Artificiais (RNA). As RNA são técnicas computacionais inspiradas nos neurônios biológicos, que simulam o cérebro humano, principalmente, quanto à sua capacidade de auto-aprendizagem, ou seja, adquirindo conhecimento por meio de experiência (Cruz e Nascimento; 2018).

A maior motivação dos pesquisadores para trabalharem com RNA é o reconhecimento de que o cérebro humano processa as informações de uma forma completamente diferente das técnicas computacionais convencionais, realizando certos processamentos de maneira muito mais rápida do que qualquer computador digital. Assim, simulando o cérebro humano, as RNA têm a capacidade de processar rapidamente uma grande quantidade de dados e aproximar funções não lineares, mapeando relações de entrada e saída, além de boa capacidade de generalização, baixa suscetibilidade a ruídos e *outliers*, entre outras características (Haykin, 2009), que fazem com que as RNA sejam cada vez mais aplicadas nas diversas áreas do conhecimento, como Economia (Mostafa e El-Masry, 2015), Medicina (Krittanawong et al., 2017; Liu et al., 2018), Agronomia (Fieuzal et al., 2017; Lu et al., 2017; Reis et al., 2016) entre outras.

O primeiro trabalho com redes neurais artificiais foi realizado pelo psiquiatra e neuroanatomista Warren McCulloch e pelo matemático Walter Pitts, que juntos propuseram o primeiro modelo de RNA composto por unidades básicas

relativamente simples (McCulloch e Pitts, 1943). Posteriormente, uma nova abordagem para o problema de reconhecimento de padrões, foi introduzida pelo pesquisador Frank Rosenblatt no seu trabalho sobre perceptron, no qual ele introduziu um modelo que difere do proposto por McCulloch e Pitts por apresentar um método inovador de aprendizagem supervisionada (Rosenblatt, 1958).

Alguns anos mais tarde, outros pesquisadores demonstraram que os perceptrons de camada única, ao contrário do que se pensava até então, tinham limitações fundamentais em relação ao que podiam calcular, não sendo capazes de solucionar problemas que não fossem linearmente separáveis (Minsky e Papert, 1969). Estes pesquisadores não acreditavam na possibilidade de se construir um método de treinamento para redes com mais de uma camada, concluindo então que as RNA seriam sempre suscetíveis a tais limitações. No entanto, em meados da década 1980, foi relatado o desenvolvimento de um algoritmo de aprendizado por retropropagação do erro (*backpropagation*), que permitiu a resolução de problemas quanto ao ajuste dos pesos de camadas intermediárias (Rumelhart et al., 1986). Desta forma, deu-se origem à proposta de modelo de rede neural artificial mais utilizada atualmente, as redes Perceptron de Múltiplas Camadas (*Multilayer Perceptron* - MLP) associada ao algoritmo de aprendizado por retropropagação do erro.

2.5.1. Neurônios biológicos e neurônios artificiais

Os neurônios biológicos, constituintes estruturais do cérebro, podem ser divididos em três seções: o corpo da célula, os dendritos e o axônio, cada um com funções específicas, porém complementares (Figura 1). Os dendritos recebem as informações (impulsos nervosos) oriundas de outros neurônios e as conduzem ao corpo da célula onde os sinais dos dendritos são processados gerando novos impulsos que passam pelo axônio e atingem os dendritos dos neurônios seguintes. A interação entre as extremidades do axônio e os dendritos dos neurônios seguintes é feita por meio das sinapses, que são conexões que podem impor excitação ou inibição ao neurônio receptivo, ou seja, controlam a transmissão dos impulsos (Cruz e Nascimento, 2018; Haykin, 2001).

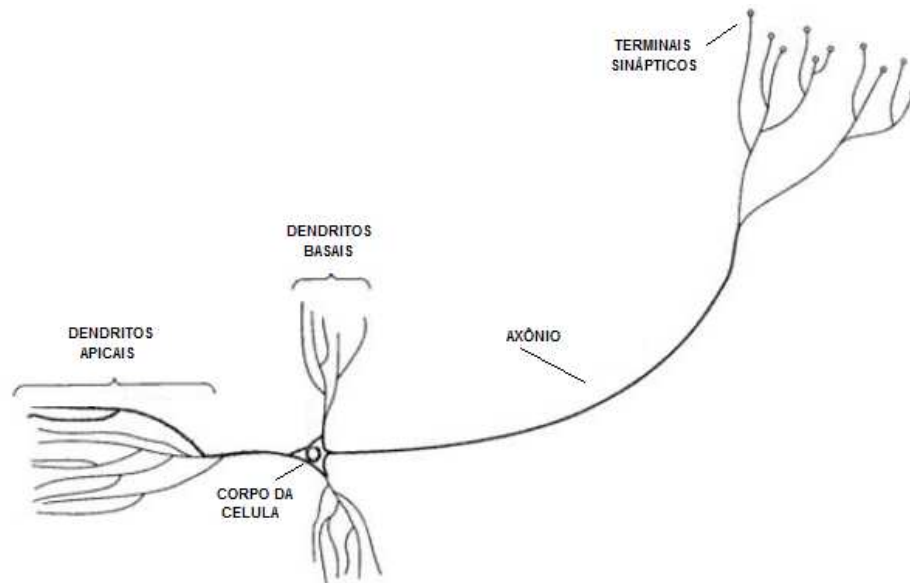


Figura 1. Representação simplificada de um neurônio biológico (Extraído de Haykin, 2001).

Os neurônios artificiais são unidades de processamento de informação inspiradas nos neurônios biológicos, fundamentais para o funcionamento das redes neurais artificiais. Conforme apresentado por McCulloch e Pitts (1943), o neurônio artificial é composto por três elementos básicos: sinais de entrada (X_j) ponderados por pesos que simulam as sinapses ($W_{k,j}$) ligando as entradas ao corpo da célula artificial, um somador (Σ) para os sinais de entrada ponderados pelos respectivos pesos sinápticos, e uma função de ativação $f(x)$ usada para restringir a amplitude de saída de um neurônio, além de um viés (b_k) que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação (Figura 2) (Haykin, 2009).

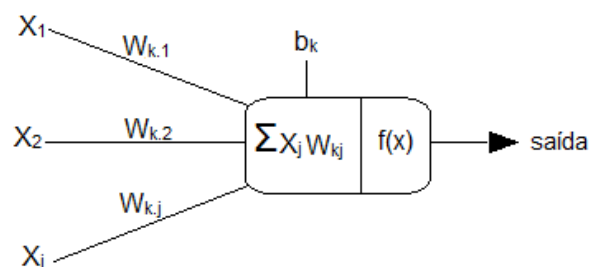


Figura 2. Esquema simplificado de um neurônio artificial.

2.5.2. Perceptrons de múltiplas camadas

As redes perceptrons de múltiplas camadas constituem uma importante classe de redes neurais artificiais e consistem em um conjunto de neurônios artificiais que formam a camada de entrada, uma camada de saída e uma ou mais camadas intermediárias ou ocultas (Figura 3). Os neurônios das camadas ocultas capacitam a rede a aprender tarefas complexas extraindo progressivamente as características mais significativas dos vetores de entrada por meio do seu treinamento de forma supervisionada com o algoritmo de retropropagação do erro. Além disso, as redes de múltiplas camadas possuem as vantagens de exibir um alto grau de conectividade determinada pelas sinapses das redes, e incluir uma função de ativação não linear a cada neurônio, o que permite inserir a não linearidade presente em problemas mais complexos (Haykin, 2001).

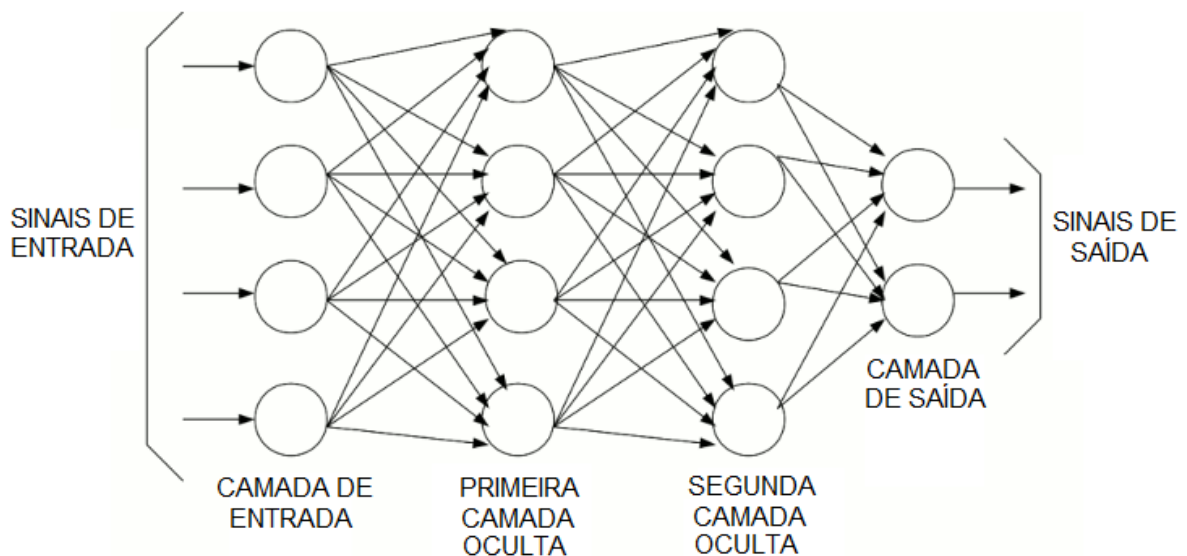


Figura 3. Esquema de uma rede neural artificial de múltiplas camadas com uma camada de entrada, duas camadas intermediárias e uma camada de saída.

O número de camadas intermediárias de uma rede de múltiplas camadas é variável e depende, principalmente, da complexidade do problema a ser resolvido. Redes com uma única camada intermediária geralmente conseguem aproximar qualquer função contínua, entretanto, os neurônios nesta camada tendem a interagir entre si globalmente, de forma que, em situações complexas, essa interação dificulta

a melhoria do ajuste de um determinado ponto sem que outro ponto seja prejudicado, comprometendo desta forma a capacidade de generalização da rede (Haykin, 2001). Geralmente, duas camadas ocultas são suficientes para resolver a maioria dos problemas, no entanto, para dados que apresentam descontinuidade e quando a precisão é tida como um critério essencial, a utilização de uma terceira camada oculta pode melhorar o desempenho da rede (Karsoliya, 2012).

Quanto ao número de neurônios das camadas intermediárias, existem alguns métodos que auxiliam na decisão, mas geralmente a escolha é feita de forma empírica e depende do conhecimento prático do pesquisador. Esta decisão deve ser tomada com cuidado, pois caso o número de neurônios esteja além do necessário, pode ocorrer o problema conhecido como *overfitting*, ou seja, a memorização dos dados de treinamento ao invés da extração das características que possibilitam a generalização (Karsoliya, 2012; Reis et al., 2016). Por outro lado, se o número de neurônios não for suficiente para o problema que está sendo abordado, pode ocorrer o *underfitting*, ou seja, a incapacidade da rede de produzir resultados satisfatórios em virtude do treinamento insuficiente (Panchal e Panchal, 2014).

2.5.3. Algoritmo de retropropagação do erro (*backpropagation*)

Com o algoritmo de aprendizagem por retropropagação do erro, o treinamento se dá, basicamente, em dois passos: passo para frente ou propagação, e passo para trás ou retropropagação. No passo para frente, as entradas são apresentadas à primeira camada da rede e propagadas em direção à camada de saída, de maneira que os sinais de saída de cada neurônio das diferentes camadas são calculados sem que haja alterações nos respectivos pesos sinápticos. Os valores de saída dos neurônios da última camada (camada de saída) são comparados às saídas desejadas emitindo-se um sinal de erro. No passo para trás, o sinal de erro emitido é propagado para as camadas anteriores proporcionando o ajuste dos pesos sinápticos dos neurônios que se dá por meio da regra delta generalizada (Cruz e Nascimento, 2018).

A regra delta, projetada para eliminar a deficiência do algoritmo de treinamento do Perceptron para dados não linearmente separáveis, é resumida na seguinte equação:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n)$$

onde: Δw_{ji} é a correção do peso, η é o parâmetro da taxa de aprendizagem, δ_j é o gradiente local, dado pelo produto da derivada da saída do neurônio pelo sinal de erro (neurônios da camada de saída), ou pelo produto da derivada da saída pela soma ponderada dos gradientes locais calculados para os neurônios da próxima camada oculta (camadas ocultas), y_i é o sinal de entrada do neurônio j , e n é a iteração ou época referente ao ajuste.

De acordo com a regra delta, quanto menor a taxa de aprendizagem (η) menor a variação dos pesos e, conseqüentemente, menor a instabilidade da rede, entretanto, a redução da taxa de aprendizagem ocorre à custa de maior lentidão na convergência do algoritmo. Assim, para aumentar a taxa de aprendizagem sem causar instabilidade na rede, inclui-se o “termo de momento”, dando origem à regra delta generalizada. O uso deste termo é especificado na seguinte equação:

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n)$$

em que: α é usualmente um número entre 0 e 1, chamado de constante de momento, $\Delta w_{ji}(n-1)$ é a correção do peso na iteração precedente, η , δ_j , y_i e n conforme apresentado anteriormente.

Desta forma, quando a derivada parcial tem o mesmo sinal algébrico em iterações consecutivas, a soma ponderada cresce em magnitude, ajustando o peso por um valor grande. Se os sinais forem opostos em iterações consecutivas, a soma ponderada diminui ajustando o peso por um valor menor, causando o efeito estabilizador (Haykin, 2001).

2.6. Redes Neurais Artificiais Aplicadas à Ciência Florestal

As redes neurais artificiais têm sido amplamente utilizadas no setor agrícola nos últimos anos, sendo aplicadas, por exemplo, em problemas que envolvem estimativa do rendimento de grãos de milho (Fieuzal et al., 2017), predição de crescimento de cana-de-açúcar (Ghazvinei et al., 2018), identificação de doenças e pragas de plantas (Fuentes et al., 2018; Lu et al., 2017), classificação de padrões de comportamento de bovinos (Peng et al., 2019), entre outros. Especificamente no setor florestal, no qual os ciclos de corte são longos e a produção sofre grande influência de fatores ambientais e genéticos ao longo dos anos, há um crescente interesse pela utilização das RNA, seja concomitantemente ou mesmo substituindo os métodos tradicionais de análises. Esse interesse se deve, entre outras coisas, à estrutura não linear das RNA que permite captar propriedades mais complexas dos dados (Cruz e Nascimento, 2018). As RNA têm sido utilizadas no setor florestal, entre outras coisas, para análise de diversidade florística (Cabacinha e Lafetá 2017), predição de incêndios florestais (Maeda et al., 2009), identificação de doenças (Poona e Ismail, 2013) e, principalmente, para prognose de crescimento das florestas, (Binoti et al., 2015; Lacerda et al., 2017; Reis et al., 2016, Vieira et al., 2018).

A utilização das RNA como ferramenta do melhoramento genético de plantas é recente e tem apresentado bons resultados, principalmente, em análises de adaptabilidade e estabilidade (Nascimento et al., 2013; Teodoro et al., 2015), seleção de genótipos (Oliveira et al., 2013; Brasileiro et al., 2015, Peixoto et al., 2015) e diversidade genética (Barbosa et al., 2011; Peternelli et al., 2017). No entanto, no que tange ao melhoramento florestal, as análises envolvendo as RNA são escassas e, tendo em vista o potencial desta ferramenta, fica evidenciada a necessidade de avanços nos estudos que permitam testar o seu desempenho no melhoramento genético florestal.

3. REFERÊNCIAS

- Araújo Júnior CA, Leite HG, Castro, RVO, Binoti DHB, Alcântara AEM, Binoti MLMS (2013) Modelagem da distribuição diamétrica de povoamentos de eucalipto utilizando a função gama. **Cerne** 19:307-314.
- Azevedo GB, Oliveira EKB, Azevedo GTOS, Buchmann HM, Miguel EP, Rezende AV (2016) Modelagem da produção em nível de povoamento e por distribuição diamétrica em plantios de eucalipto. **Scientia Forestalis** 44:383-392.
- Barbosa CD, Viana AP, Quintal SSR, Pereira MG (2011) Artificial neural network analysis of genetic diversity in *Carica papaya* L. **Crop Breeding and Applied Biotechnology** 11:224-231.
- Becerra PI, Catford JA, Inderjit, Mcleod ML, Andonian K, Aschehoug ET, Montesinos D, Callaway RM (2018) Inhibitory effects of *Eucalyptus globulus* on understorey plant growth and species richness are greater in non-native regions. **Global Ecology and Biogeography** 27:68-76.
- Binoti MLMS, Leite HG, Binoti DHB, Gleriani JM (2015) Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Cerne**, 21: 97-105.
- Brasileiro BP, Marinho CD, Costa PMA (2015) Cruz CD, Peternelli LA, Barbosa MHP Selection in sugarcane families with artificial neural networks. **Crop Breeding and Applied Biotechnology** 15:72-78.
- Cabacinha CD, Lafetá BO (2017) Floristic diversity and equitability in forest fragments using artificial neural networks. **Ciência Florestal** 27:143-152.
- Campos JCC, Leite HG (2013) **Mensuração florestal: perguntas e respostas** 4ed. Viçosa:UFV, 605 p.
- Castro CAO, Resende RT, Bhering LL, Cruz CD (2016) Brief history of Eucalyptus in Brazil under perspective of biometric advances. **Ciência Rural** 46:1585-1593.
- Castro RVO, Soares CPB, Martins FB, Leite HG (2013) Crescimento e produção de plantios comerciais de eucalipto estimados por duas categorias de modelos. **Pesquisa Agropecuária Brasileira** 48:287-295.
- Corrêa TR, Picoli EAT, Souza GA, Condé SA, Silva NM, Lopes-Mattos KLB, Resende MDV, Zauza EAV, Oda S (2017) Phenotypic markers in early selection for tolerance to dieback in *Eucalyptus*. **Industrial Crops & Products**. 107:130-138.
- Correia I, Alía R, Yan W, David T, Aguiar A, Almeida MH (2010) Genotype x Environment interactions in *Pinus pinaster* at age 10 in a multi-environment trial in Portugal: a maximum likelihood approach. **Annals of Forest Science**, 67:612–612.

Cruz CD, Nascimento M (2018) **Inteligência Computacional Aplicada ao Melhoramento Genético**. UFV, 414p.

Dias PC, Xavier A, Resende MDV, Barbosa MHP, Biernask FA, Estopa RA (2018) Genetic evaluation of *Pinus taeda* clones from somatic embryogenesis and their genotype x environment interaction. **Crop Breeding and Applied Biotechnology** 18: 55-64.

Engel ML, Higa AR, Andrejow GP, Flôres Junior PC, Soares ID (2016) Genetic gain from different selection methods in *Eucalyptus macarthurii* progenies in different environments. **CERNE** 22:299-308.

Fieuzal R, Sicre CM, Baup F (2017) Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. **International Journal of Applied Earth Observation and Geoinformation** 57:14–23.

Fuentes AF, Yoon S, Lee J, Park DS (2018) High-Performance Deep Neural Network-Based Tomato Plant Diseases and Pests Diagnosis System With Refinement Filter Bank. **Frontiers in Plant Science** 9:1162.

Gallo R, Pantuza IB, Santos GA, Resende MDV, Xavier A, Simiqueli GF, Baldind T, Santos OP, Valente BMRT (2018) Growth and wood quality traits in the genetic selection of potential *Eucalyptus dunnii* Maiden clones for pulp production. **Industrial Crops & Products** 123:434–441.

Ghazvinei PT, Darvishi HH, Mosavi A, Yusof KBW, Alizamir M, Shamshirband S, Chau K (2018) Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. **Engineering applications of computational fluid mechanics**. 12:738–749.

Gonçalves JLM, Alvares CA, Higa AR, Silva LD, Alfenas AC, Stahl J, Barros SFF, Lima WP, Brancalion PHS, Hubner A (2013) Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. **Forest Ecology and Management** 301:6-27.

Haykin S (2009) **Neural networks and learning machines**. 3 ed. Prentice Hall, 906p.

Haykin S (2001) **Redes Neurais: princípios e práticas**. 2 ed. Bookman, 899p.

IBA - Indústria Brasileira de Árvores (2017): Anuário estatístico da indústria brasileira de árvores: ano base 2016. Brasília, 80p. Disponível em: http://www.iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2017.pdf 10 de jun. 2018.

Kainer D, Stone EA, Padovan A, Foley WJ, Külheim C (2018) Accuracy of Genomic Prediction for Foliar Terpene Traits in *Eucalyptus polybractea*. **G3: Genes|Genomes|Genetics** 8:2573-2583.

Karsoliya S (2012) Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. **International Journal of Engineering Trends and Technology** 3:714-717.

Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T (2017) Artificial Intelligence in Precision Cardiovascular Medicine. **Journal of the American College of Cardiology**, 69: 2657-2664.

Lacerda THS, Cabacinha CD, Araújo Júnior CA, Maia RD, Lacerda KWS (2017) Artificial neural networks for estimating tree volume in the brazilian savanna. **Cerne**, 23: 483-491.

Leon N, Jannink J, Edwards JW, Kaeppler SM (2016) Introduction to a Special Issue on Genotype by Environment Interaction. **Crop Science** 56:2081:2089.

Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R (2018) Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging. **Magnetic Resonance in Medicine** 79:2379–2391.

Lu Y, Yi S, Zeng N, Liu Y, Zhang Y (2017) Identification of rice diseases using deep convolutional neural networks. **Neurocomputing** 267: 378–384.

Maeda EE, Formaggio AR, Shimabukuro YE, Arcoverde GFB, Hansen MC (2009) Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. **International Journal of Applied Earth Observation and Geoinformation** 11:265-272.

Maier CA, Albaugh TJ, Cook RI, Hall K, Mcinnis D, Johnsen KH, Johnson J, Rubilar RA, Vose M (2017) Comparative water use in short-rotation *Eucalyptus benthamii* and *Pinus taeda* trees in the Southern United States. **Forest Ecology and Management** 397:126–138.

Martins FB, Soares CPB, Silva GF (2014) Individual tree growth models for eucalyptus in northern Brazil. **Scientia Agricola** 71:212-225.

Massaro RAM, Bonine CAV, Scarpinati EA, Paula RC (2010) Viabilidade de aplicação da seleção precoce em testes clonais de *Eucalyptus* sp. **Ciência Florestal** 20:597-609.

Mcculloch WS, Pitts WH (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity, **Bulletin of Mathematical Biophysics** 5:115–133.

Minsky ML, Papert SA (1969) **Perceptrons – An Introduction to Computational Geometry** p.258.

Miranda AC, Moraes MLT, Silva PHM, Sebbenn AM (2015) Ganhos genéticos na seleção pelo método do índice multiefeitos em progênies polinização livre de *Eucalyptus grandis* Hill ex Maiden. **Scientia Forestalis** 43:203-209.

- Miranda R, Fiorentin L, Péllico Netto S, Juvanhol R, Dalla Corte A (2018) Prediction system for diameter distribution and wood production of eucalyptus. **Floresta e Ambiente** 25:1-12
- Moraes CB, Freitas TCM, Pieroni GB, Zimback L, Resende MDV, Mori ES (2014) Estimativas dos parâmetros genéticos para seleção precoce de clones de *Eucalyptus* para região com ocorrência de geadas. **Scientia Forestalis** 42: 219-227.
- Mostafa MM, El-Masry AA (2016) Oil price forecasting using gene expression programming and artificial neural networks. **Economic Modelling** 54: 40–53.
- Nascimento M, Peternelli LA, Cruz CD, Nascimento ACC, Ferreira RP, Bhering LL, Salgado CC (2013) Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. **Crop Breeding and Applied Biotechnology** 13:152-156.
- Oliveira ACL, Pasqual M, Pio LAS, Lacerda WS, Silva SO (2013) Utilização da modelagem matemática (redes neurais artificiais) na classificação de autotetraplóides de bananeira (*Musa acuminata* COLLA) **Bioscience Journal** 29:617-622.
- Oliveira TWG, Paula RC, Moraes MLT, Alvares CA, Miranda AC, Silva PHM (2018) Stability and adaptability for wood volume in the selection of *Eucalyptus saligna* in three environments. **Pesquisa agropecuária brasileira** 53:611-619.
- Paiva JR, Resende MDV, Cordeiro ER (2002) Índice Multifeitos e estimativas de parâmetros genéticos em aceroleira. **Pesquisa agropecuária brasileira** 37:799-807.
- Panchal FS, Panchal M (2014) Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network. **International Journal of Computer Science and Mobile Computing** 3:455 – 464.
- Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. **Biometrika** 58: 545-554.
- Peixoto LA, Bhering LL, Cruz CD (2015) Artificial neural networks reveal efficiency in genetic value prediction. *Genetics and Molecular Research* 14:6796-6807.
- Peng Y, Kondo N, Fujiura T, Suzuki T, Wulandari, Yoshioka H, Itoyama E (2019) Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. **Computers and Electronics in Agriculture** 157:247–253.
- Peternelli LA, Moreira EFA, Nascimento M, Cruz CD (2017) Artificial neural networks and linear discriminant analysis in early selection among sugarcane families. **Crop Breeding and Applied Biotechnology** 17: 299-305.
- Pinto DS, Resende RT, Mesquita AG, Rosado AM, Cruz CD (2014) Seleção precoce para características de crescimento em testes clonais de *Eucalyptus urophylla*. **Scientia Forestalis** 42:251–257.

Poona NK, Ismail R (2013) Discriminating the occurrence of pitch canker fungus in *Pinus radiata* trees using QuickBird imagery and artificial neural networks, **Southern Forests: a Journal of Forest Science** 75:29-40.

Reis LP, Souza AL, Mazzei L, Reis PCM, Leite HG, Soares CPB, Torres CMME, Silva LF, Ruschel AR (2016) Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. **Forest Ecology and Management** 382:161 – 167.

Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília, Colombo: Embrapa Florestas, 975p.

Resende MDV (2004) **Métodos estatísticos ótimos na análise de experimentos de campo**. Brasília, Colombo: Embrapa Floresta, 57p.

Resende MDV (2007) **SELEGEN-REML/BLUP: Sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Brasília, Colombo: Embrapa Florestas, 359 p.

Resende MDV, Higa AR (1994) Estimação de valores genéticos no melhoramento de *Eucalyptus*: seleção em um caráter com base em informações do indivíduo e seus parentes. **Boletim de Pesquisa Florestal** 28/29:11-36.

Resende RT, Carneiro ACO, Ferreira RADC, Kuki KN, Teixeira RU, Zaidan UR, Santos RD, Leite HG, Resende, MDV (2018) Air-drying of eucalyptus logs: Genetic variations along time and stem profile. **Industrial Crops&Products** 124:316–324.

Rosado AM, Rosado TB, Resende Júnior MFR, Bhering LL, Cruz CD (2009) Ganhos genéticos preditos por diferentes métodos de seleção em progênies de *Eucalyptus urophylla*. **Pesquisa Agropecuária Brasileira** 44:1653-1659,

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review** 65:386-408.

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by backpropagating errors. **Nature** 323:533–536.

Santos GA, Nunes ACP, Resende MDV, Silva LD, Higa A, Assis TF (2016) An index combining volume and Pilodyn penetration to study stability and adaptability of *Eucalyptus* multi-species hybrids in Rio Grande do Sul, Brazil. **Australian Forestry** 79:248–255.

Santos GA, Resende MDV, Silva LD, Higa A, Assis TF (2015) Interação genótipos x ambientes para produtividade de clones de *Eucalyptus* L'Hér. no estado do Rio Grande do Sul. **Revista Árvore** 39:81-91.

Schmidt LN, Machado SA, Pelissari AL, Silva GF (2019) Dynamics of eucalyptus diameter distribution in the state of Minas Gerais. **Floresta e Ambiente**, 26:1-9.

Silva AM, Canuto DSO, Moraes MLT (2017) Genetic parameters in riparian reforestation without structure of progenies by mixed models (reml/blup). **CERNE**, 23:501-506.

Silva PHM, Campoe OC, Paula RC, Lee DJ (2016) Seedling Growth and Physiological Responses of Sixteen Eucalypt Taxa under Controlled Water Regim. **Forests** 7:1-13.

Teodoro PE, Barroso LMA, Nascimento M, Torre FE, Sagrilo E, Santos A, Ribeiro LP (2015) Redes neurais artificiais para identificar genótipos de feijão-caupi semiprostrado com alta adaptabilidade e estabilidade fenotípicas. **Pesquisa Agropecuária Brasileira** 50:1054-1060.

Teodoro PE, Bhering LL, Costa RD, Rocha RB, Laviola BG (2016) Mixed models for selection of *Jatropha* progenies with high adaptability and yield stability in Brazilian regions. **Genetics and Molecular Research** 15:1-9.

Vieira GC, Mendonça AR, Silva GF, Zanetti SS, Silva MM, Santos AR (2018) Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. **Science of the Total Environment** 619:1473–1481.

Yan W, Tinker NA (2006) Biplot analysis of multi-environment trial data: Principles and applications. **Canadian Journal of Plant Science** 86:623-645.

CAPÍTULO 2 - Prognose do diâmetro de árvores individuais de eucalipto em diversas condições climáticas utilizando redes neurais artificiais

RESUMO - O desenvolvimento de um modelo de rede neural artificial (RNA) para prognose de crescimento de eucalipto em locais onde não há plantios já estabelecidos é de extrema importância. Assim, o objetivo deste trabalho foi modelar o crescimento de um povoamento clonal de eucalipto empregando redes neurais artificiais, e avaliar a capacidade da aplicação do modelo em outros locais com diferentes materiais genéticos e condições edafoclimáticas distintas. Os dados utilizados neste estudo foram obtidos a partir de inventários florestais contínuos de 18 clones comerciais de *Eucalyptus* em 17 locais de diferentes estados do Brasil. As redes testadas neste trabalho foram do tipo múltiplas camadas com o algoritmo de aprendizado de retropropagação do erro e função de ativação do tipo tangente hiperbólica. A rede neural que utilizou o diâmetro à altura do peito (DAP), avaliado próximo aos dois anos de idade, e sua respectiva média como variáveis de entrada foi a que apresentou melhor desempenho na prognose do crescimento em DAP. O modelo desenvolvido para prognose do DAP de árvores individuais em um local foi capaz de manter boa acurácia ao ser aplicado em outros locais sem a necessidade de ajustes nos pesos sinápticos.

Palavras-chave: Inteligência artificial, inventário florestal, modelagem do crescimento florestal, RNA, teste clonal

CHAPTER 2 - Prognosis of the diameter of individual eucalyptus trees in various climatic conditions using artificial neural networks

ABSTRACT - The development of an artificial neural network (ANN) model for prognosis of eucalyptus growth in places where there are no established plantations is of extreme importance. Thus, the objective of this work was to model the growth of a clonal eucalypt settlement using artificial neural networks, and to evaluate the capacity of the application of the model in other sites with different genetic materials and different edaphoclimatic conditions. The data used in this study were obtained from continuous forest inventories of 18 commercial Eucalyptus clones at 17 sites in different states of Brazil. The networks tested in this work were of the multilayer type with the algorithm of learning of backpropagation of the error and activation function of the hyperbolic tangent type. The neural network that used diameter at breast height (DBH) obtained close to two years of age and their respective mean as input variables was the most efficient in the prognosis of growth in DBH. The model developed for DBH prognosis of individual trees in one place was able to maintain good accuracy when applied to other sites without the need for adjustments in the synaptic weights.

Keywords: Artificial intelligence, forest inventory, modeling of forest growth, ANN, clonal test

1. Introdução

A cultura do eucalipto no Brasil, responsável por mais de 72% das florestas plantadas no país, além da maior produtividade com média de 35,7 m³/ha ao ano, tem a menor rotação do mundo, com um período médio entre plantio e colheita próximo a seis anos (IBÁ, 2017). Ainda assim, o tempo é um fator crítico na produção do eucalipto, dificultando diversos processos essenciais para o desenvolvimento da cultura como a seleção de novos genótipos devido à demora dos ciclos de seleção, assim como o estabelecimento de planos adequados de manejo. Neste sentido, é comum a busca por modelos de crescimento que possibilitem a prognose da produção florestal ao longo do tempo, facilitando o planejamento e o manejo das florestas.

Entre as principais categorias de modelos, os destinados a prognose de árvores individuais destacam-se por gerar informações detalhadas sobre a dinâmica da estrutura dos povoamentos (Castro et al., 2013). No entanto, em comparação aos tradicionais modelos de regressão, inúmeros autores têm comprovado maior eficiência de ferramentas de inteligência artificial como as redes neurais artificiais (RNA) na prognose do crescimento de florestas (Binoti et al., 2014; Campos et al., 2016; Vieira et al., 2018). As RNA são modelos que se baseiam em uma rede de neurônios biológicos com capacidade de processar rapidamente uma grande quantidade de dados e aproximar funções não lineares, mapeando relações de entrada e saída com base em sua auto-aprendizagem, além de boa capacidade de generalização, baixa suscetibilidade a ruídos e *outliers*, entre outras características (Haykin, 2009) que as tornam mais vantajosas em relação aos modelos de regressão.

A maioria das pesquisas envolvendo a aplicação de redes neurais artificiais na prognose da produção florestal utiliza um único conjunto de dados obtido em um único local ou região, dividindo-se aleatoriamente esse conjunto de forma que parte dos dados seja destinada ao treinamento das redes e a outra parte destinada a validação das mesmas. Porém, tanto para o treinamento da RNA, onde os pesos são ajustados pelo algoritmo de aprendizagem, como para a etapa de validação, onde é testada a capacidade da RNA em produzir saídas adequadas para as

entradas que não estavam na etapa de treinamento (Binoti et al., 2014), é necessária a utilização de um conjunto de dados obtido em florestas com idades já avançadas para que as saídas das redes possam ser comparadas com os valores obtidos nas medições das árvores atestando, assim, a eficiência do modelo.

Atualmente, há tendência de expansão do setor florestal para áreas com condições edafoclimáticas distintas daquelas em que os programas de melhoramento foram implementados e conduzidos. Isto se faz necessário, inclusive, para atender a crescente demanda por produtos de base florestal, advinda do próprio crescimento populacional. Considerando o que foi exposto anteriormente, o desenvolvimento de um modelo de RNA para prognose de crescimento de florestas clonais de eucalipto para esses locais seria, inicialmente, impossível em função da ausência de plantios já estabelecidos. Desta forma, o objetivo deste trabalho foi modelar o crescimento de um povoamento clonal de eucalipto em função de variáveis numéricas empregando redes neurais artificiais, e avaliar a capacidade da aplicação do modelo, sem ajuste dos pesos sinápticos, para outros locais com diferentes genótipos e/ou condições edafoclimáticas distintas.

2. Material e métodos

Os dados utilizados neste estudo foram obtidos a partir de inventários florestais contínuos em 17 locais de diferentes estados do Brasil, com condições climáticas distintas segundo a classificação de Köppen (Figura 1; Tabela 1).

Os dados são de 18 clones comerciais de *Eucalyptus* que fazem parte do projeto TECHS (Tolerância de *Eucalyptus* Clonais aos Estresses Hídrico, Térmico e Biótico). Cada local conta com 11 dos 18 clones, sendo que 4 clones classificados como plásticos, são coincidentes em todos os locais e 7 clones são específicos para a condição climática local (Binkley et al., 2017), com exceção do local 16 que, nesse caso, contou apenas com 10 clones dada a ausência de um clone específico (Tabela 1).

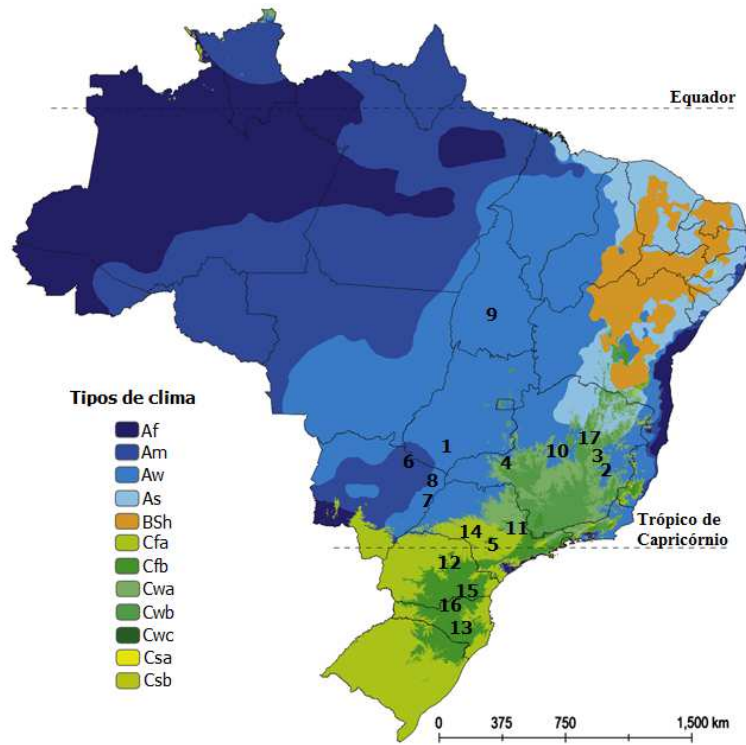


Figura 1. Classificação climática (Alvares et al., 2013) e localização dos 17 locais utilizados neste estudo.

Tabela 1. Localização, descrição climática e clones de cada local presente neste estudo.

LOCAL	CIDADE PRÓXIMA	PRECIP	TEMP	CLONES											
1	Rio Verde - GO	1472	23.95	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
2	Belo Oriente - MG	1236	22.58	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
3	Guanhães - MG	1443	22.23	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
4	Estrela do Sul - MG	1519	23.60	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
5	Botucatu - SP	1314	22.41	A1	C3	F6	I9	J1	K2	L3	M4	N5	O6	Q8	
6	Chapadão do Sul - MS	1665	24.22	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
7	Três Lagoas - MS	1179	25.09	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
8	Inocência - MS	1194	25.15	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
9	Brejinho de Nazaré - TO	1305	27.69	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
10	Três Marias - MG	1216	24.11	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
11	Mogi Guaçu - SP	1535	21.82	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
12	Telemaco Borba - PR	1789	19.87	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
13	Otacílio Costa - SC	1666	16.66	A1	C3	F6	I9	J1	K2	L3	M4	N5	O6	Q8	
14	Borebi - SP	1706	22.75	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	
15	Antônio Olinto - PR	1780	18.53	A1	C3	F6	I9	J1	K2	L3	M4	N5	O6	Q8	
16	Três Barras - SC	1902	17.46	A1	C3	F6	I9	J1	K2	L3	M4	N5	Q8		
17	Bocaiúva - MG	938	24.43	A1	B2	C3	D4	E5	G7	H8	K2	P7	Q8	R9	

PRECIP = Precipitação média anual ; TEMP = Temperatura média anual com base nos 24 meses após o plantio.

O local 1 foi tomado aleatoriamente para desenvolvimento do modelo da rede neural artificial. Para tal, foram consideradas as seguintes variáveis: diâmetro à altura do peito (DAP) de árvores individuais na primeira idade, ou seja, aos 13 meses (DAP_1 , cm), DAP na segunda idade (27 meses) das árvores individuais (DAP_2 , cm), média do clone no local na primeira idade (\bar{x}_1 , cm), média do clone no local na segunda idade (\bar{x}_2 , cm) e DAP individual na terceira idade (50 meses) de todas as árvores vivas dos 11 clones (DAP_3 , cm), totalizando 1324 árvores.

Os dados do local 1 foram divididos em dois conjuntos, um destinado ao treinamento das redes que contou com cerca de 80% das árvores (1086 árvores) e outro destinado a etapa de validação, com aproximadamente 20% dos dados (238 árvores). A divisão dos dados nessa proporção tem apresentado bom desempenho para grandes conjuntos de dados no desenvolvimento de modelos de redes neurais artificiais (Binoti et al., 2015; Nascimento et al., 2013; Reis et al., 2016; Teodoro et al., 2015). Visando testar a capacidade do modelo em prever o DAP de genótipos distintos, o conjunto de dados destinado ao treinamento das redes contou com árvores de 8 clones (B2, C3, G7, H8, K2, P7, Q8 e R9), enquanto o conjunto usado na etapa de validação continha apenas árvores dos outros 3 clones (A1, D4 e E5).

O modelo que representa o neurônio artificial, expresso na equação abaixo, conta com sinais de entrada (X_j) que representam os dendritos de um neurônio biológico, pesos que simulam as sinapses ($W_{k,j}$) ligando as entradas ao corpo da célula artificial, um somador (Σ) para os sinais de entrada ponderados pelas respectivas sinapses do neurônio, uma função de ativação $f(x)$ usada para restringir a amplitude de saída de um neurônio, e um viés (b_k) que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação (Haykin, 2009):

$$Y = f\left(\sum_{j=1}^m W_{k,j} \cdot X_j + b_k\right)$$

As redes utilizadas neste trabalho foram do tipo múltiplas camadas com o algoritmo de aprendizado por retropropagação do erro, função de ativação do tipo tangente hiperbólica, e número máximo de iterações (ou épocas) para cada rede igual a 1.000. As variáveis utilizadas como entrada das redes foram DAP_1 , DAP_2 , \bar{x}_1

e \bar{x}_2 , e a saída da rede foi a variável DAP_3 . Foram considerados quatro modelos segundo as entradas da rede (Tabela 2). Para evitar que variáveis de maior magnitude tivessem maior influência no resultado das redes, os dados foram normalizados no intervalo de 0 a 1 por meio da seguinte equação:

$$X_{\text{norm}} = \left(\frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \right)$$

onde X_{norm} é o valor normalizado, X_i é o valor real a ser normalizado, X_{min} e X_{max} são os valores mínimo e máximo da variável, respectivamente.

Tabela 2. Arquitetura, variáveis de entrada e variáveis de saída das redes neurais artificiais selecionadas na fase de treinamento.

RNA	ARQUITETURA	ENTRADAS	SAÍDA
1	4-5-1	$DAP_1, DAP_2, \bar{x}_1, \bar{x}_2$	DAP_3
2	2-3-1	DAP_1, DAP_2	DAP_3
3	2-3-1	DAP_1, \bar{x}_1	DAP_3
4	2-3-1	DAP_2, \bar{x}_2	DAP_3

RNA = Redes Neurais Artificiais; DAP_1 = Diâmetro a altura do peito (cm) na primeira idade; DAP_2 = Diâmetro a altura do peito na segunda idade; \bar{x}_1 = Média do clone no local na primeira idade; \bar{x}_2 = Média do clone no local na segunda idade; DAP_3 = diâmetro a altura do peito (cm) na terceira idade.

O número de neurônios da única camada oculta de cada modelo foi estabelecido de acordo com o método de Fletcher-Gloss (Silva et al., 2010) considerando o número de variáveis de entrada e saída, de acordo com a condição a seguir:

$$(2 \cdot \sqrt{n} + n2) \leq n1 \leq (2 \cdot n + 1)$$

em que n é o número de entradas da rede, $n1$ é o número de neurônios da camada oculta e $n2$ é a quantidade de neurônios na camada de saída.

Foram treinadas 400 RNA, sendo 100 para cada um dos modelos, das quais apenas uma de cada modelo foi escolhida com base na correlação entre os valores observados e os valores estimados pelas redes. O critério para avaliar a qualidade dos 4 modelos selecionados foram: análise gráfica de dispersão dos resíduos, correlação entre os valores observados e estimados ($r_{y\hat{y}}$), erro médio (ME) e a raiz quadrada do erro quadrático médio (RMSE%), conforme as expressões a seguir:

$$r_{y\hat{y}} = \frac{\text{Cov}(y, \hat{y})}{\sqrt{S^2(y) \cdot S^2(\hat{y})}}$$

em que Cov é a covariância, S^2 a variância, y é o valor observado e \hat{y} o valor estimado pela RNA.

$$\text{ME} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}$$

em que: y_i é o valor observado da i ésima variável; \hat{y}_i é valor estimado da i ésima variável e n é o tamanho da amostra.

$$\text{RMSE\%} = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

em que: y_i é o valor observado da i ésima variável; \hat{y}_i é valor estimado da i ésima variável e n é o tamanho da amostra.

O melhor modelo de RNA, segundo os critérios de avaliação, foi selecionado e aplicado, sem nenhum ajuste dos pesos sinápticos, aos dados dos outros 16 locais avaliados, de acordo com os mesmos critérios ($r_{y\hat{y}}$, ME e RMSE%). As variáveis de entrada e saída da rede não foram obtidas exatamente na mesma idade (meses) em todos os locais, sendo considerado um intervalo entre 24 e 29 meses para as variáveis DAP_2 e \bar{x}_2 , e entre 49 e 53 meses para a variável DAP_3 . Todo o processo das análises foi feito com o software Statistica (Statsoft, 2004).

3. Resultados e discussão

As redes 1 e 4 foram as que apresentaram melhor desempenho na prognose do DAP das árvores na etapa de treinamento, o que fica evidenciado ao analisar as figuras de dispersão dos valores observados e estimados, em que as RNA 1 e 4 apresentam menor dispersão de dados (Figura 2). Entretanto, a RNA 4 apresentou valor de erro médio (ME) mais próximo de zero, indicando menor viés desse modelo em relação aos demais, enquanto a RNA 1 foi a que apresentou o valor mais distante de zero entre os modelos testados nessa fase (Tabela 3).

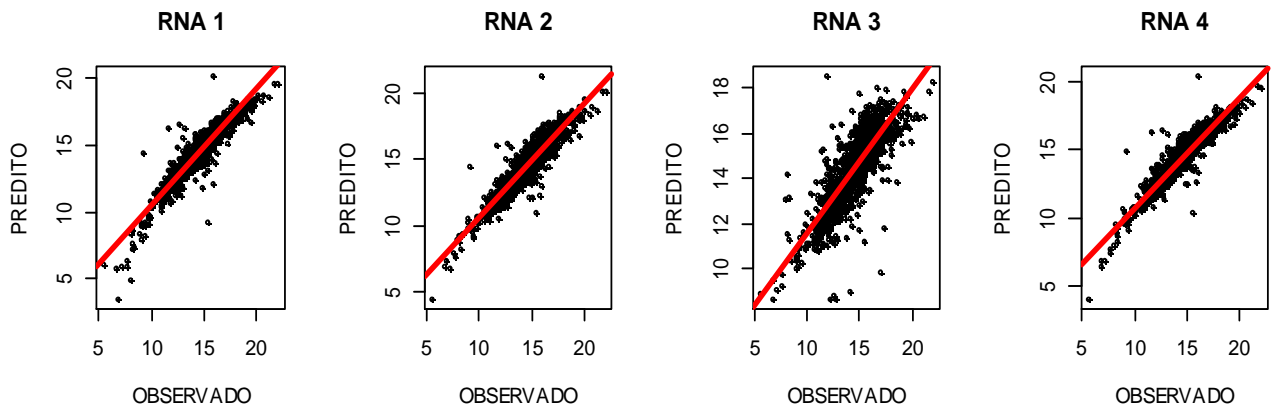


Figura 2. Análise gráfica do diâmetro a altura do peito predito (cm) vs. o diâmetro a altura do peito observado (cm) de árvores individuais de eucalipto na fase de treinamento das redes neurais artificiais.

Tabela 3. Desempenho das redes neurais selecionadas nas fases de treinamento e validação.

RNA	TREINAMENTO			VALIDAÇÃO		
	$r_{y\hat{y}}$	<i>ME</i>	<i>RMSE%</i>	$r_{y\hat{y}}$	<i>ME</i>	<i>RMSE%</i>
1	0.93	0.081	4.75	0.89	0.226	7.18
2	0.91	0.064	5.93	0.81	0.921	9.51
3	0.81	0.068	8.63	0.72	0.818	14.23
4	0.94	-0.024	4.94	0.95	0.064	4.66

$r_{y\hat{y}}$ = Correlação entre os valores observados e preditos; *ME* = Erro médio; *RMSE%* = Raiz quadrada do erro quadrático médio.

A RNA 4 também se mostra superior às demais quanto à capacidade de generalização, uma vez que na etapa de validação das RNA, a mesma manteve os valores de *RMSE%* e *ME* mais baixos, além de maior correlação entre os valores

observados e estimados. As outras RNA apresentaram diminuição na acurácia das estimativas e na correlação entre os valores observados e estimados (Tabela 3). Isso indica maior associação das variáveis envolvidas na RNA 4, que utiliza apenas variáveis obtidas na segunda idade como variáveis de entrada, em relação às demais redes, que utilizam pelo menos uma variável obtida na primeira idade (Tabela 2).

Massaro et al. (2010) obtiveram ótima correlação entre o DAP mensurado próximo aos dois anos de idade com o obtido aos 50 meses, ao investigarem a viabilidade de aplicação da seleção precoce em testes clonais de *Eucalyptus* spp. Li et al. (2017), ao analisarem a correlação entre diferentes idades para caracteres de crescimento em eucalipto, também observaram que o DAP aos dois anos de idade apresentou boa correlação com o DAP em idades mais avançadas, enquanto que em idades inferiores a dois anos essa correlação é consideravelmente reduzida. Assim, as variáveis de entrada na primeira idade podem ter influenciado negativamente nas correlações e na acurácia das redes 1, 2 e 3, principalmente na etapa de validação. Outro fato que contribui para a superioridade da RNA 4, e favorece a sua escolha para a aplicação aos demais locais, é a sua configuração mais simples, com menor número de neurônios, o que facilita a sua aplicação.

Ao aplicar o modelo da RNA 4 para os demais locais, observa-se que o DAP estimado segue a tendência do DAP observado (Figura 3), com correlações positivas entre esses variando de 0,78 a 0,96, mantendo-se acima de 0,90 em 11 dos 16 locais em que o modelo foi aplicado (Tabela 4). O modelo também se mostra com boa precisão, com RMSE% variando entre 5,07 e 14,52 entre os locais, e mantendo-se abaixo de 10 na maioria deles (Tabela 4). Binoti et al. (2015), ao analisarem diferentes configurações de RNA na projeção do volume de clones de eucalipto em um único local, selecionaram redes com RMSE% que variaram de 10,89 a 15,49 e Reis et al (2018), ao testarem RNA na prognose de distribuição diamétrica em um povoamento florestal na Amazônia, selecionaram 5 arquiteturas de redes com RMSE% acima de 15,30 na fase de treinamento.

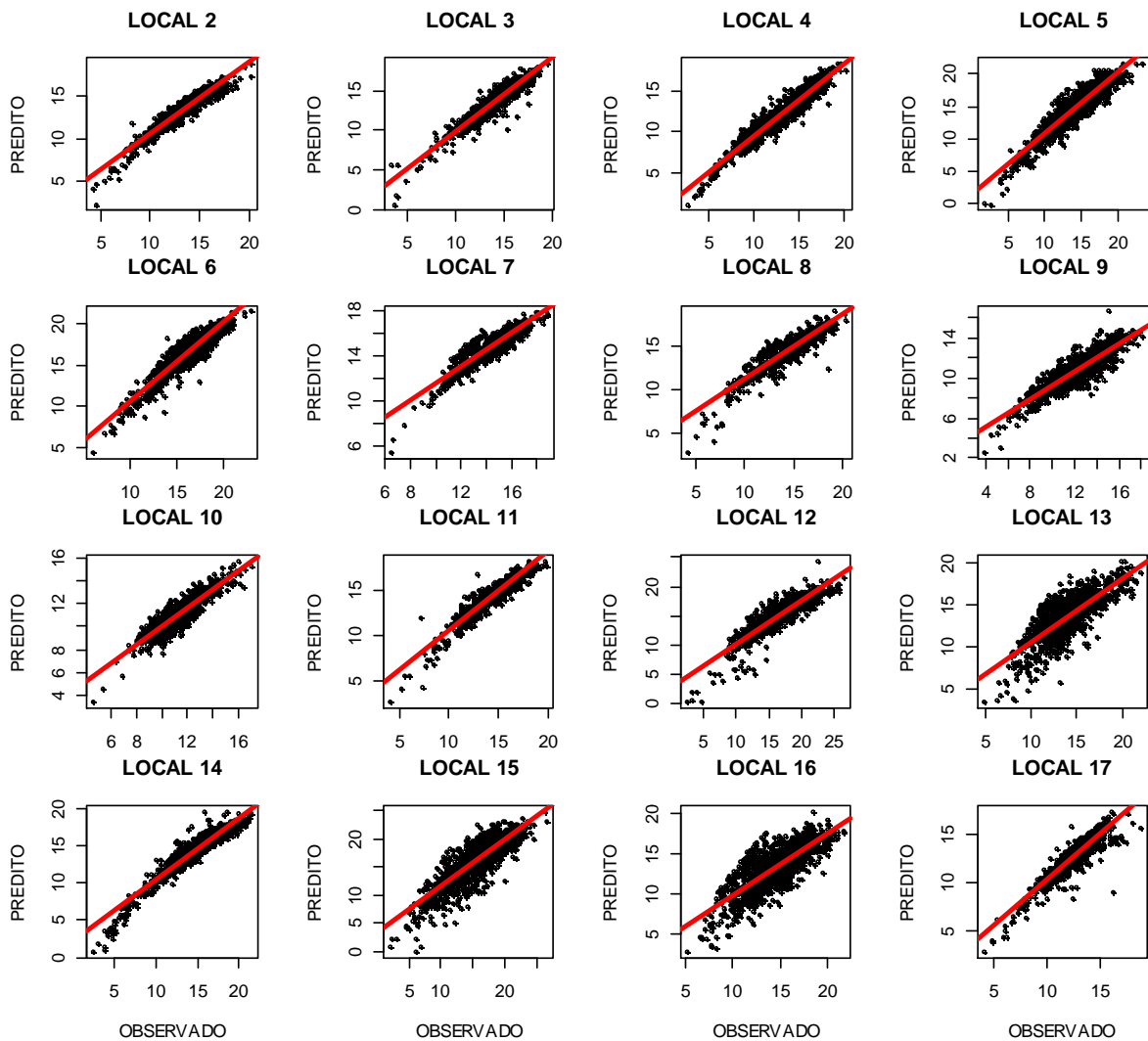


Figura 3. Análise gráfica do diâmetro à altura do peito predito (cm) por um modelo de rede neural artificial vs. o diâmetro à altura do peito observado (cm) de árvores individuais de eucalipto em 16 locais de diferentes estados do Brasil.

Os cinco locais onde a correlação foi inferior a 0,90 são os mesmos em que o modelo se mostrou menos preciso, com valores de RMSE% acima de 10. Quatro desses locais encontram-se abaixo do trópico de Capricórnio (Figura 1) e são os que apresentam menores temperaturas médias anuais (Tabela 1), sendo que, entre esses quatro locais, quanto mais ao sul do Brasil e mais baixa a temperatura média anual, menor foi a correlação entre os dados observados e estimados. O outro local em que a correlação foi inferior a 0,90 e o RMSE% acima de 10 é o que está localizado mais ao norte do Brasil (Figura 1), com a temperatura média anual mais

elevada entre todos os locais avaliados (Tabela 1). Binkley et al. (2017), ao estudarem a interação dos mesmos clones avaliados neste trabalho em 36 ambientes ao longo do Brasil e do Uruguai, observaram uma redução de até 40% na produção de madeira para uma variação de apenas 4°C na temperatura média anual. Entretanto, a diferença na produção entre os locais não explica, necessariamente, a diminuição na correlação e na precisão do modelo, que podem estar associadas ao efeito diferencial dos ambientes sobre os genótipos, ou seja, à interação genótipo x ambiente (IGA).

Tabela 4. Desempenho da rede neural artificial selecionada no local 1 quando aplicada em outros 16 locais de diferentes regiões do Brasil.

LOCAIS	$r_{\hat{y}y}$	<i>ME</i>	<i>RMSE%</i>
2	0.95	-0.087	5.408
3	0.94	0.478	5.950
4	0.96	0.886	9.080
5	0.93	-0.610	9.569
6	0.93	-0.403	5.767
7	0.90	-0.512	6.349
8	0.91	0.019	7.323
9	0.88	1.425	14.346
10	0.90	0.210	6.808
11	0.94	-0.009	5.071
12	0.89	1.607	12.527
13	0.78	0.395	11.776
14	0.95	0.480	6.722
15	0.84	-0.675	14.519
16	0.80	1.234	14.018
17	0.90	-0.258	7.373

$r_{\hat{y}y}$ = Correlação entre os valores observados e preditos; *ME* = Erro médio; *RMSE%* = Raiz quadrada do erro quadrático médio.

A IGA pode ocasionar diferenças nas produções dos locais em proporções distintas para cada genótipo ao longo do tempo, podendo resultar em trocas de posições na classificação dos clones nos diferentes ambientes (Borém e Miranda, 2013), e sua presença afetando a manifestação fenotípica em eucalipto é comumente relatada na literatura (Castro et al., 2018; Nunes et al., 2002; Pupin et al., 2015; Santos et al., 2015).

Ao comparar os locais 5 e 13, que apresentam diferenças acima de 300 mm na precipitação média anual e acima de 5 °C na temperatura média anual (Tabela 1), para o mesmo grupo de genótipos, nota-se uma diferença de 0,15 na correlação entre os dados observados e estimados e inversão na tendência preditiva do modelo. No local 5 os valores preditos superestimaram os dados observados, enquanto no local 13 esses dados foram levemente subestimados (Tabela 4). Da mesma forma, entre os locais 2 e 9, que apresentam uma diferença superior a 5 °C entre as temperaturas médias anuais, as correlações diferiram em 0,07 e a RMSE% em quase 9% para o mesmo grupo de genótipos (Tabela 4). Isso indica que a produção não foi alterada na mesma proporção ao longo do tempo para todos os clones nos diferentes locais, corroborando assim com a idéia de que a interação genótipo x ambiente afetou a correlação e a acurácia do modelo nos diferentes locais.

Por outro lado, com exceção do local 13, os demais locais apresentaram correlação superior a 0,80 entre os dados observados e estimados, assim como a acurácia do modelo se manteve próxima às obtidas por outros autores que comprovaram a eficiência da rede neural artificial na prognose de caracteres de crescimento em espécies arbóreas (Binoti et al., 2015; Lacerda et al., 2017; Vieira et al., 2018). Ainda nesse sentido, os locais 5 e 6, que estão localizados em regiões com climas contrastantes, segundo a classificação de Köppen (Figura 1), e contaram com diferentes grupos de genótipos, apresentaram correlações e acurácias semelhantes (Tabela 4). O uso de genótipos específicos para cada ambiente é uma forma de minimizar os efeitos da interação genótipo x ambiente (Cruz et al., 2004). Assim, o uso de grupos de clones mais adequados para os diferentes tipos de clima, apesar do considerável gradiente ambiental entre os locais estudados, reduziu o efeito da interação genótipo x ambiente, permitindo que o modelo apresentasse bom desempenho quando aplicado em diferentes locais.

4. Conclusão

Redes neurais com uma camada oculta e que utilizam apenas a avaliação do diâmetro à altura do peito (DAP) em idade próxima aos dois anos e sua respectiva

média como variáveis de entrada, apresentaram bom desempenho e podem ser utilizadas na prognose do crescimento em DAP de árvores individuais. É importante salientar que os genótipos devem ser adaptados a cada local para que as suas especificidades representem as características dos ambientes e reduzam o efeito da interação genótipo x ambiente ao longo do tempo.

5. Referências

- Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM (2013) Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift** 22:711–728.
- Binkley D, Campoe OC, Alvares C, Carneiro RL, Cegatta I, Stape JL (2017) The interactions of climate, spacing and genetics on clonal Eucalyptus plantations across Brazil and Uruguay. **Forest Ecology and Management** 405:271-283.
- Binoti MLMS, Binoti DHB, LEITE HG, SILVA AAL, Pontes C (2014) Utilização de redes neurais artificiais para a projeção da distribuição diamétrica de povoamento equiâneos. **Revista Árvore** 38:747-754.
- Binoti MLMS, Leite HG, Binoti DHB, Gleriani JM (2015) Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Cerne** 21:97-105.
- Borém A, Miranda VM (2013) **Melhoramento de Plantas**. 6 ed. Viçosa: UFV, 523 p.
- Castro CEC, Silva PHM, Araújo MJ, Miranda AC, Moraes MLT Paula RC (2018) Adaptabilidade, estabilidade e produtividade de progênies de *Eucalyptus grandis* Hill ex Maiden. **Scientia Forestalis** 46:157-166.
- Castro RVO, Soares CPB, Martins FB, Leite HG (2013) Crescimento e produção de plantios comerciais de eucalipto estimados por duas categorias de modelos. **Pesquisa Agropecuária Brasileira** 48:287-295.
- Campos BPF, Silva GF, Binoti DHB, Mendonça AR, Leite HG (2016) Predição da altura total de árvores em plantios de diferentes espécies por meio de redes neurais artificiais. **Pesquisa Florestal Brasileira** 36: 375-385.
- Cruz CD, Regazzi AJ, Carneiro PCS (2004) **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: UFV, 480 p.
- Haykin S. (2009) **Neural networks and learning machines**. 3 ed. Prentice Hall, 906p.

IBA - Indústria Brasileira de Árvores (2017): Anuário estatístico da indústria brasileira de árvores: ano base 2016. Brasília, 80p. Disponível em: http://www.iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2017.pdf 10 de jun. 2018.

Lacerda THS, Cabacinha CD, Araújo Júnior CA, Maia RD, Lacerda KWS (2017) Artificial neural networks for estimating tree volume in the brazilian savanna. **Cerne** 23: 483-491.

Li C, Weng Q, Chen J, Li M, Zhou C, Chen S, Zhou W, Guo D, Lu C, Chen J, Xiang D, Gan S (2017) Genetic parameters for growth and wood mechanical properties in *Eucalyptus cloeziana* F. Muell. **New Forests**, 48:33–49.

Massaro RAM, Bonine CAV, Scarpinati EA, Paula RC (2010) Viabilidade de aplicação da seleção precoce em testes clonais de *Eucalyptus* spp. **Ciência Florestal**, 20:597-609.

Nascimento M, Peternelli LA, Cruz CD, Nascimento ACC, Ferreira RP, Bhering LL, Salgado CC (2013) Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. **Crop Breeding and Applied Biotechnology** 13:152-156.

Nunes GHS, Rezende GDSP, Ramalho MAP, Santos JB (2002) Implicações da interação genótipos x ambientes na seleção de clones de eucalipto. **Cerne**, 8:049-058.

Pupin S, Santos AVA, Zaruma DUG, Miranda AC, Silva PHM, Marino CL, Sebbenn AM, Moraes MLT (2015) Produtividade, estabilidade e adaptabilidade em progênies de polinização aberta de *Eucalyptus urophylla* S.T. Blake. **Scientia Forestalis** 43: 127-134.

Reis LP, Souza AL, Mazzei L, Reis PCM, Leite HG, Soares CPB, Torres CMME, Silva LF, Ruschel AR (2016) Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. **Forest Ecology and Management**, 382:161 – 167.

Reis LP, Souza AL, Reis PCM, Mazzei L, Binoti DHB, Leite HG (2018) Prognose da distribuição diamétrica na amazônia utilizando redes neurais artificiais e autômatos celulares. **Floresta** 48:93-102.

Silva IN, Spattl DH, Flauzino RA (2010) **Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas**. São Paulo, ARTLIBER, 399p.

STATSOFT. **Statistica** (data analysis software system), version 7. Tulsa: StatSoft Inc., 2004.

Teodoro PE, Barroso LMA, Nascimento M, Torres FE, Sagrilo E, Santos A, Ribeiro LP (2015) Redes neurais artificiais para identificar genótipos de feijão-caupi semiprostrado com alta adaptabilidade e estabilidade fenotípicas. **Pesquisa Agropecuária Brasileira** 50:1054-1060.

Santos GA, Resende MDV, Silva LD, Higa AR, Assis TF (2015) Interação genótipos x ambientes para produtividade de clones de *Eucalyptus* l'hér. No estado do rio grande do sul. **Revista Árvore**, Viçosa-MG, 39:81-91.

Vieira GC, Mendonça AR, Silva GF, Zanetti SS, Silva MM, Santos AR (2018) Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. **Science of the Total Environment**, Amsterdam v. 619–620, p. 1473–1481.

CAPÍTULO 3 - Redes Neurais Artificiais na predição de valor genético em *Eucalyptus sp.*

RESUMO - Os programas de melhoramento genético têm como principal objetivo a seleção e recomendação de genótipos superiores para plantios em escala comercial, e as Redes Neurais Artificiais (RNA) são ferramentas potenciais para a predição do valor genético dos indivíduos auxiliando na seleção. Este trabalho buscou avaliar o desempenho das RNA na predição de valor genético de indivíduos e famílias de *Eucalyptus sp.* Na fase de treinamento foram utilizados dados de experimentos simulados com base em médias, herdabilidades e coeficientes de variação comumente encontrados para o diâmetro a altura do peito em testes de progênies de eucalipto. Os modelos foram validados com dados provenientes de um teste de progênies de *Eucalyptus grandis*, instalado em delineamento de blocos casualizados, com 115 famílias, seis repetições e seis plantas por parcela. As RNA testadas neste trabalho foram do tipo múltiplas camadas com algoritmo de aprendizado por retropropagação do erro. Os modelos de três camadas ocultas, com seis e quatro neurônios por camada, foram os que apresentaram melhor desempenho na predição do valor genético, respectivamente, de árvores individuais e de famílias. RNA do tipo múltiplas camadas apresentam bom desempenho na estimação de valor genético para a seleção individual, entre famílias e entre e dentro de famílias de *Eucalyptus grandis*.

Palavras-chave: BLUP, perceptrons de múltiplas camadas, RNA, seleção entre e dentro de famílias, teste de progênies

CHAPTER 3 - Artificial Neural Networks in the prediction of genetic value in *Eucalyptus sp.*

ABSTRACT - Genetic improvement programs have as main objective the selection and recommendation of superior genotypes for commercial scale plantations, and Artificial Neural Networks (ANN) are potential tools for predicting the genetic value of individuals assisting in selection. This work aimed to evaluate the performance of ANN in the prediction of genetic value of individuals and families of *Eucalyptus sp.* In the training phase, data from simulated experiments were used based on averages, heritabilities and coefficients of variation commonly found for diameter at breast height in tests of eucalyptus progenies. The models were validated with data from a progeny test of *Eucalyptus grandis*, installed in a randomized complete block design, with 115 families, six replicates and six plants per plot. The ANN tested in this work were of the multilayers type with learning algorithm by backpropagation of the error. The models of three hidden layers, with six and four neurons per layer, presented the best performance in predicting the genetic value, respectively, of individual trees and families. ANN of the multilayer present good performance in the estimation of genetic value for individual selection, among families and among and within families of *Eucalyptus grandis*

Keywords: BLUP, multilayer perceptrons, ANN, selection among and within families, progeny test

1. Introdução

O principal objetivo dos programas de melhoramento genético é a seleção e recomendação de genótipos superiores, e para tanto, os pesquisadores buscam, por meio da seleção e recombinação, aumentar a frequência de alelos favoráveis e/ou as suas combinações, visando à melhoria no desempenho dos genótipos para os caracteres de interesse de acordo com o propósito do programa. Entretanto, na identificação de genótipos superiores, é importante considerar informações que expressem o mérito genético dos indivíduos e/ou das famílias, assim, a utilização de estimadores adequados de parâmetros e valores genéticos é um fator de grande importância em estudos de melhoramento florestal, e por isso tem sido bastante discutido em trabalhos na área (Ferreira et al., 2012; Miranda et al., 2015; Mora et al., 2013; Rosado et al., 2009).

Entre os principais métodos de estimação de valor genético, destaca-se o método BLUP (Best Linear Unbiased Prediction – Melhor Predição Linear não Viesada) proposto por Henderson na década de 40, e muito utilizado atualmente no melhoramento florestal (Dias et al., 2018; Kainer et al., 2018; Silva et al., 2017). O BLUP ajusta os dados para os efeitos ambientais identificáveis e simultaneamente prediz os valores genéticos dos indivíduos (Resende, 2002). Este método presume o conhecimento dos componentes de variância, mas como isso não é possível, são utilizadas as estimativas desses componentes, usualmente, obtidas por meio do método REML (Restricted Maximum Likelihood – Máxima Verossimilhança Restrita), proposto por Patterson e Thompson (1971).

Com base nos valores genéticos é possível selecionar as árvores por diferentes métodos de seleção de acordo com o objetivo, a etapa do programa de melhoramento e as demandas de curto, médio e longo prazo. Entre os principais métodos de seleção encontram-se a seleção individual, na qual as árvores são selecionadas com base no ranking dos valores genéticos individuais (Ferreira et al., 2012); a seleção entre famílias, onde as melhores famílias são selecionadas considerando todos os indivíduos das mesmas (Laviola et al., 2014); e a seleção entre e dentro de famílias, na qual selecionam-se as melhores famílias e, posteriormente, os melhores indivíduos dentro dessas famílias (Rosado et al., 2009).

Considerando que a seleção a partir de valores genéticos estimados pelo método REML/BLUP são feitas com base na análise dos modelos mistos, é fundamental que sejam atendidas algumas pressuposições básicas das análises estatísticas convencionais sobre a distribuição e propriedade dos dados. Entretanto, como isso nem sempre ocorre, a confiabilidade das estimativas acabam muitas das vezes sendo comprometidas. Nesse contexto, as Redes Neurais Artificiais (RNA) surgem como uma alternativa na predição do valor genético. As RNA são modelos inspirados nos elementos básicos que formam a estrutura dos neurônios biológicos, e têm entre as suas principais vantagens a capacidade de aproximar funções não lineares captando informações complexas no conjunto de dados, além da não exigência de informações *a priori* sobre a distribuição dos dados e dos resíduos, permitindo que qualquer informação a respeito dos genótipos possa ser utilizada sem qualquer especificação explícita (Haykin, 2009).

As RNA têm sido amplamente utilizadas no setor agrícola nos últimos anos (Leal et al., 2015; Soares et al., 2014; Vieira et al., 2018; Zanuncio et al., 2017), entretanto, os trabalhos sobre o desempenho das RNA na predição de valor genético em espécies florestais ainda são escassos. Assim, o objetivo deste trabalho foi avaliar o desempenho das Redes Neurais Artificiais para prever valores genéticos de indivíduos e famílias de *Eucalyptus sp.*

2. Material e métodos

Para fins de treinamento das Redes Neurais Artificiais foram simulados dados de experimentos em blocos casualizados (DBC) com 100 progênies, seis plantas por parcela e quatro repetições, considerando quatro valores de médias, quatro herdabilidades, e quatro coeficientes de variação, totalizando 64 experimentos simulados (Tabela 1). Os parâmetros utilizados para simulação foram estabelecidos com base em valores comumente encontrados na literatura para a variável diâmetro à altura do peito (DAP) em testes de progênies de *Eucalyptus sp.* (Henriques et al., 2017; Macedo et al., 2013; Moraes et al., 2015; Silva et al., 2016). A simulação dos dados foi realizada com o auxílio do programa computacional GENES (Cruz, 2013).

Tabela 1. Parâmetros estabelecidos para simulação de dados dos 64 experimentos

Exp	MG	h ² (%)	CV(%)	Exp	MG	h ² (%)	CV(%)
1	5	30	10	33	15	30	10
2	5	30	20	34	15	30	20
3	5	30	30	35	15	30	30
4	5	30	40	36	15	30	40
5	5	40	10	37	15	40	10
6	5	40	20	38	15	40	20
7	5	40	30	39	15	40	30
8	5	40	40	40	15	40	40
9	5	50	10	41	15	50	10
10	5	50	20	42	15	50	20
11	5	50	30	43	15	50	30
12	5	50	40	44	15	50	40
13	5	60	10	45	15	60	10
14	5	60	20	46	15	60	20
15	5	60	30	47	15	60	30
16	5	60	40	48	15	60	40
17	10	30	10	49	20	30	10
18	10	30	20	50	20	30	20
19	10	30	30	51	20	30	30
20	10	30	40	52	20	30	40
21	10	40	10	53	20	40	10
22	10	40	20	54	20	40	20
23	10	40	30	55	20	40	30
24	10	40	40	56	20	40	40
25	10	50	10	57	20	50	10
26	10	50	20	58	20	50	20
27	10	50	30	59	20	50	30
28	10	50	40	60	20	50	40
29	10	60	10	61	20	60	10
30	10	60	20	62	20	60	20
31	10	60	30	63	20	60	30
32	10	60	40	64	20	60	40

Exp: experimento, MG: Média geral, h²: herdabilidade a nível de média e CV: coeficiente de variação.

A validação dos modelos de RNA foi realizada utilizando-se dados provenientes de um teste de progênies de *Eucalyptus grandis*, instalado em DBC, com 115 famílias, seis repetições e parcelas experimentais de seis plantas.

Para obtenção das estimativas dos valores genéticos utilizados como variável de saída nas fases de treinamento e validação da RNA e os componentes de variância dos dados utilizados na etapa de validação, foi utilizado o procedimento REML/BLUP (máxima verossimilhança restrita/melhor predição linear não viesada), disponível no software genético-estatístico SELEGEN, desenvolvido por Resende (2007), assumindo-se os dados como sendo de progênes de meios-irmãos com várias observações por parcela, seguindo o modelo proposto por Resende (2007):

$$y = Xr + Za + Wp + e$$

em que y é o vetor de dados, r é o vetor dos efeitos de bloco (assumidos como fixos) somados à média geral, a é o vetor dos efeitos genéticos aditivos individuais (aleatórios), p é o vetor dos efeitos de parcela (aleatórios), e é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

O neurônio artificial é representado pelo seguinte modelo:

$$Y = f\left(\sum_{j=1}^m W_{k,j} \cdot X_j + b_k\right)$$

em que: Y é o valor de saída do neurônio, $f(x)$ é uma função de ativação responsável por restringir a amplitude de saída de um neurônio, $W_{k,j}$ são os pesos sinápticos usados para ponderar os sinais de entrada, X_j são os sinais de entrada, b_k é um viés que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação (Haykin, 2009).

Os modelos de RNA utilizados neste trabalho foram do tipo múltiplas camadas com o algoritmo de aprendizado por retropropagação do erro, função de ativação do tipo tangente hiperbólica e número máximo de iterações (ou épocas) para cada rede igual a 1.000. Para os modelos de predição do valor genético aditivo das árvores individuais, foram utilizados como variáveis de entrada o DAP da árvore, a média do bloco e a média da progênie no experimento. Já para os modelos de predição do valor genético aditivo das famílias, as variáveis de entrada utilizadas foram a média geral do experimento, a média da progênie e a amplitude da progênie dentro do experimento considerando todas as árvores vivas. Para que a diferença na magnitude das variáveis não influenciasse no resultado das RNA, os dados foram normalizados no intervalo de 0 a 1 por meio da seguinte equação:

$$X_{\text{norm}} = \left(\frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \right)$$

onde X_{norm} é o valor normalizado, X_i é o valor real a ser normalizado, X_{min} e X_{max} são os valores mínimo e máximo da variável, respectivamente.

A arquitetura dos modelos, tanto os de árvores individuais como os de famílias, variou quanto ao número de camadas ocultas e ao número de neurônios por camada oculta, sendo de 1 a 3 quanto ao número de camadas ocultas, e de acordo com o estabelecido pelo método de Fletcher-Gloss (Silva et al., 2010) para o número de neurônios, considerando o número de variáveis de entrada e saída, de acordo com a condição a seguir:

$$(2 \cdot \sqrt{n} + n2) \leq n1 \leq (2 \cdot n + 1)$$

em que n é o número de entradas da rede, $n1$ é o número de neurônios da camada oculta e $n2$ é a quantidade de neurônios na camada de saída (Tabela 2).

Tabela 2. Arquitetura, variáveis de entrada e variável de saída das redes neurais artificiais avaliadas.

RNA	ARQUITETURA	ENTRADAS	SAÍDA	RNA	ARQUITETURA	ENTRADAS	SAÍDA
IND 1	3-4-1	DAP, MB, MP	a-l	FAM 1	3-4-1	MG, MP, AP	a-F
IND 2	3-5-1	DAP, MB, MP	a-l	FAM 2	3-5-1	MG, MP, AP	a-F
IND 3	3-6-1	DAP, MB, MP	a-l	FAM 3	3-6-1	MG, MP, AP	a-F
IND 4	3-7-1	DAP, MB, MP	a-l	FAM 4	3-7-1	MG, MP, AP	a-F
IND 5	3-4-4-1	DAP, MB, MP	a-l	FAM 5	3-4-4-1	MG, MP, AP	a-F
IND 6	3-5-5-1	DAP, MB, MP	a-l	FAM 6	3-5-5-1	MG, MP, AP	a-F
IND 7	3-6-6-1	DAP, MB, MP	a-l	FAM 7	3-6-6-1	MG, MP, AP	a-F
IND 8	3-7-7-1	DAP, MB, MP	a-l	FAM 8	3-7-7-1	MG, MP, AP	a-F
IND 9	3-4-4-4-1	DAP, MB, MP	a-l	FAM 9	3-4-4-4-1	MG, MP, AP	a-F
IND 10	3-5-5-5-1	DAP, MB, MP	a-l	FAM 10	3-5-5-5-1	MG, MP, AP	a-F
IND 11	3-6-6-6-1	DAP, MB, MP	a-l	FAM 11	3-6-6-6-1	MG, MP, AP	a-F
IND 12	3-7-7-7-1	DAP, MB, MP	a-l	FAM 12	3-7-7-7-1	MG, MP, AP	a-F

IND = Modelos para predição do valor genético aditivo individual; FAM = Modelos para predição do valor genético aditivo das famílias; DAP = diâmetro à altura do peito; MB = média do DAP do bloco; MP = média do DAP da progênie; MG = média geral do DAP do experimento; AP = amplitude do DAP da progênie; a-l = valor genético aditivo individual para DAP; a-F = valor genético aditivo da família para DAP.

Na etapa de treinamento, para cada um dos modelos, foram treinadas 100 RNA, sendo selecionada a melhor de cada modelo quanto à correlação entre os valores estimados pelo método REM/BLUP e os estimados pelas RNA ($r_{y\hat{y}}$), quanto

ao erro médio (ME) e a raiz do erro quadrático médio (RMSE), conforme as expressões a seguir:

$$r_{y\hat{y}} = \frac{\text{Cov}(y, \hat{y})}{\sqrt{S^2(y) \cdot S^2(\hat{y})}}$$

em que Cov é a covariância, S^2 a variância, y é o valor estimado pelo método BLUP e \hat{y} o valor estimado pela RNA.

$$\text{ME} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}$$

em que: y_i é o valor estimado da i ésima variável pelo método BLUP; \hat{y}_i é valor estimado da i ésima variável pela RNA e n é o tamanho da amostra.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

em que: y_i , \hat{y}_i e n , conforme apresentado anteriormente.

Os modelos selecionados na etapa de treinamento foram avaliados quanto à capacidade de generalização na etapa de validação pelos seguintes critérios: análise gráfica de dispersão; correlação entre os valores estimados pelo método REM/BLUP e os estimados pelas RNA; correlação de postos de Spearman (ρ); porcentagem de coincidência na seleção pelos métodos REML/BLUP e RNA entre os 30 melhores indivíduos para os modelos de árvores individuais, e em três níveis de seleção (10, 20 e 30%) para os modelos de famílias; e pela diferença dos ganhos com a seleção (GS%) estimados pelos dois métodos com a seleção dos 30 melhores indivíduos para os modelos de árvores individuais e nos três níveis de seleção dos modelos de famílias, conforme as seguintes expressões:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)}$$

em que: d é a diferença entre os postos dos indivíduos ou famílias de acordo com a predição do valor genético aditivo pelos métodos BLUP e RNA, e n é o número de indivíduos ou famílias.

$$\text{GS}\% = \frac{\left(\frac{\sum aS}{nS}\right) 100}{DAP}$$

em que: aS é o valor genético aditivo dos indivíduos ou famílias selecionados, nS é o número de indivíduos ou famílias selecionados, e \overline{DAP} é a média do DAP de todos os indivíduos ou famílias.

O melhor modelo de RNA para predição do valor genético aditivo de árvores individuais e o melhor modelo para predição do valor genético aditivo das famílias, foram aplicados conjuntamente para testar o desempenho dos modelos na seleção entre e dentro de famílias. Para tanto, o modelo de famílias foi usado para seleção de 20% das melhores famílias (23 famílias) e o modelo de árvores individuais para seleção de 20% das melhores árvores de cada família selecionada (7 árvores/famílias). Na seleção entre e dentro, o desempenho dos modelos aplicados conjuntamente, foi determinado pela correlação de postos de Spearman (ρ) entre as árvores coincidentes nas seleções realizadas pelos métodos REML/BLUP e RNA, pela porcentagem de coincidência das árvores selecionadas pelos dois métodos, e pela comparação dos GS% estimados pelos dois métodos. Todos os modelos de RNA foram desenvolvidos com o auxílio do software Statistica (Statsoft, 2004).

3. Resultados e discussão

Na fase de treinamento, todas as arquiteturas testadas, tanto dos modelos de árvores individuais como os de famílias, alcançaram correlações positivas e de alta magnitude, indicando boa associação entre as variáveis envolvidas nos modelos de RNA (Tabela 3). O ME se manteve próximo de zero e não indicou tendência clara dos modelos em sub ou superestimar os valores genéticos preditos pelo método REML/BLUP. Já a RMSE, apesar de não indicar grandes magnitudes dos erros em nenhum dos casos, foi maior para os modelos de árvores individuais (Tabela 3). Isso pode ter ocorrido devido ao fato de que a variação entre os dados da etapa de treinamento, tanto das variáveis de entrada como de saída, foi maior nos modelos de árvores individuais, dificultando um pouco mais o reconhecimento de padrões pelas RNA.

Tabela 3. Desempenho das redes neurais artificiais selecionadas nas fases de treinamento.

RNA	r_{yy}	ME	RMSE	RNA	r_{yy}	ME	RMSE
IND 1	96.84	0.019	0.538	FAM 1	97.61	0.007	0.272
IND 2	97.60	-0.009	0.533	FAM 2	97.70	-0.005	0.267
IND 3	96.84	-0.001	0.537	FAM 3	97.60	0.001	0.272
IND 4	96.75	0.003	0.545	FAM 4	97.64	0.002	0.270
IND 5	96.87	-0.021	0.536	FAM 5	97.74	0.000	0.264
IND 6	96.84	0.019	0.538	FAM 6	97.89	-0.006	0.256
IND 7	97.05	-0.007	0.520	FAM 7	97.79	-0.007	0.261
IND 8	97.01	0.011	0.524	FAM 8	97.84	-0.005	0.259
IND 9	97.05	-0.003	0.519	FAM 9	97.80	0.007	0.261
IND 10	96.87	0.004	0.535	FAM 10	97.79	-0.003	0.261
IND 11	97.06	0.010	0.519	FAM 11	97.91	0.007	0.255
IND 12	97.08	-0.002	0.517	FAM 12	97.92	0.000	0.254

IND = Modelos para predição do valor genético aditivo individual; FAM = Modelos para predição do valor genético aditivo das famílias; r_{yy} = Correlação entre os valores preditos pelo método REML/BLUP e pela RNA; **ME** = Erro médio; **RMSE** = Raiz quadrada do erro quadrático médio.

Na análise do experimento utilizado para validação das redes neurais artificiais, os valores encontrados para o coeficiente de variação genotípica entre progênies ($CV_{gp}\%$) e para o coeficiente de variação genética aditiva individual ($CV_{gi}\%$) foram de 8,512 e 17,024 respectivamente. Valores próximos a esses têm sido encontrados em estudos com espécies florestais por diversos autores que evidenciam, nesses casos, a possibilidade de seleção (Hamilton et al., 2015; Moraes et al., 2015; Souza et al., 2011; Zaruma et al., 2015). Desta forma, justifica-se a utilização destes dados para validação das Redes Neurais Artificiais.

Na etapa de validação, as correlações entre os valores estimados pelo método REML/BLUP e pelas RNA mantiveram-se positivas e de alta magnitude, ficando acima de 95% para os modelos de árvores individuais (Tabela 4) e acima de 96% para os modelos de seleção de famílias (Tabela 5). Esses valores foram superiores ao encontrado por Ventura et al. (2012) ao compararem esses dois métodos na predição de valor genético de bovinos.

Tabela 4. Desempenho dos modelos de redes neurais artificiais para seleção de árvores individuais na etapa de validação.

RNA	$r_{y\hat{y}}$ (%)	ρ (%)	% COINCIDENTES	GS (%)	
				BLUP	RNA
IND 1	97.20	99.38	93.33	33.63	78.96
IND 2	97.74	99.46	93.33	33.63	165.59
IND 3	97.85	99.28	93.33	33.63	30.00
IND 4	97.99	99.81	96.66	33.63	57.98
IND 5	95.39	99.60	93.33	33.63	36.31
IND 6	95.67	99.35	86.66	33.63	25.17
IND 7	97.11	99.54	93.33	33.63	115.32
IND 8	97.26	99.00	90	33.63	119.26
IND 9	95.50	99.54	93.33	33.63	28.24
IND 10	95.84	99.70	96.66	33.63	22.91
IND 11	99.81	99.94	96.66	33.63	38.08
IND 12	98.95	99.79	93.33	33.63	10.30

IND = Modelos para predição do valor genético aditivo individual; $r_{y\hat{y}}$ = Correlação entre os valores preditos pelo método REML/BLUP e pela RNA; ρ (%) = correlação de postos de Spearman; % COINCIDENTES = porcentagem de árvores coincidentes entre as 30 selecionadas pelo método REML/BLUP e pela RNA; GS = Ganho com a seleção.

A forte correlação encontrada neste trabalho refletiu no ordenamento dos indivíduos e das famílias, alcançando valores que se mantiveram acima de 94% para ρ nos dois casos (Tabelas 4 e 5). Sabendo que o método REML/BLUP considera uma série de parâmetros genéticos e efeitos ambientais na estimação do valor genético, as altas correlações encontradas na etapa de validação neste trabalho evidenciam que a capacidade das RNA vai além de mapear relações de entrada e saída, e isso ocorre devido à habilidade das RNA de extrair informações não apresentadas de forma explícita na etapa de treinamento (Santos et al., 2005).

Considerando a importância da precisão na escolha de genótipos superiores para os programas de melhoramento, a porcentagem de coincidência entre os indivíduos selecionados pelo método REML/BLUP e os selecionados pela RNA foi vista como um critério de suma importância na escolha dos melhores modelos de RNA. No caso dos modelos de árvores individuais, o número de árvores coincidentes entre as selecionadas pelos dois métodos variou entre 26 (86,66%) e 29 (96,66%). Estes valores são maiores do que os encontrados por Silva et al. (2014) ao determinarem a porcentagem de coincidência na seleção de genótipos

simulados realizada por redes neurais e modelos lineares. Os modelos IND 4, IND 10 e IND 11 foram os melhores quanto a esse critério (Tabela 4), e destes, o modelo IND 11 foi o que apresentou troca de posições menos considerável, excluindo da seleção pela RNA o pior indivíduo entre os trinta selecionados pelo método REML/BLUP e inserindo o primeiro indivíduo abaixo dos selecionados.

Tabela 5. Desempenho dos modelos de redes neurais artificiais para seleção de famílias na etapa de validação.

RNA	% DE COINCIDENTES					GANHO COM A SELEÇÃO (%)					
	$r_{y\hat{y}}$ (%)	ρ (%)	NÍVEL 1	NÍVEL 2	NÍVEL 3	NÍVEL 1		NÍVEL 2		NÍVEL 3	
						BLUP	RNA	BLUP	RNA	BLUP	RNA
FAM 1	97.35	99.68	100	95.65	94.12	22.77	20.12	19.04	17.13	16.53	14.98
FAM 2	99.34	99.89	90.90	100	100	22.77	10.35	19.04	8.90	16.53	7.88
FAM 3	99.10	99.87	90.90	100	100	22.77	7.06	19.04	6.06	16.53	5.34
FAM 4	99.02	99.70	90.90	100	97.05	22.77	8.68	19.04	7.50	16.53	6.67
FAM 5	99.00	99.28	100	95.65	94.12	22.77	13.35	19.04	11.35	16.53	9.89
FAM 6	97.32	99.58	81.82	91.30	97.05	22.77	16.83	19.04	14.73	16.53	13.18
FAM 7	99.11	99.54	100	95.65	94.12	22.77	10.81	19.04	10.23	16.53	9.81
FAM 8	99.34	99.97	100	100	100	22.77	7.97	19.04	6.83	16.53	6.01
FAM 9	99.15	99.88	90.90	100	100	22.77	18.59	19.04	15.91	16.53	14.02
FAM 10	98.58	98.51	90.90	95.65	91.17	22.77	19.10	19.04	16.08	16.53	13.97
FAM 11	96.48	94.17	72.72	78.26	85.29	22.77	12.63	19.04	10.45	16.53	8.84
FAM 12	97.35	99.97	100	100	100	22.77	7.97	19.04	6.83	16.53	6.01

FAM = Modelos para predição do valor genético aditivo de famílias; $r_{y\hat{y}}$ = Correlação entre os valores preditos pelo método REML/BLUP e pela RNA; ρ (%) = correlação de postos de Spearman; % COINCIDENTES = porcentagem de famílias coincidentes entre as selecionadas pelo método REML/BLUP e pela RNA em três níveis de seleção; GS = Ganho com a seleção; NÍVEL 1 = seleção de 10% das famílias; NÍVEL 2 = seleção de 20% das famílias; NÍVEL 3 = seleção de 30% das famílias.

Além do acerto na seleção dos melhores genótipos, é importante que os modelos de RNA não sejam tendenciosos, uma vez que a sub ou superestimação dos valores genéticos podem dificultar a distinção entre os indivíduos ou supervalorizá-los, além de proporcionar estimativas errôneas de ganhos com a seleção comprometendo assim o avanço dos programas de melhoramento genético. De maneira geral, não houve uma tendência clara entre os modelos de árvores individuais de sub ou superestimar os valores genéticos dos genótipos selecionados, e entre os modelos mais precisos na seleção dos indivíduos, o modelo IND 11 foi o que estimou o GS% mais próximo do estimado pelo método BLUP (Tabela 4) sendo,

portanto, escolhido como o melhor na estimação de valor genético de árvores individuais de eucalipto entre os modelos testados. A superioridade deste modelo fica evidenciada na análise dos gráficos de dispersão (Figura 1), onde nota-se também uma tendência geral dos modelos de árvores individuais de subestimar os valores genéticos dos indivíduos abaixo da média. Entretanto, este fato não foi utilizado como critério na escolha dos modelos por não influenciar na eficiência da seleção.

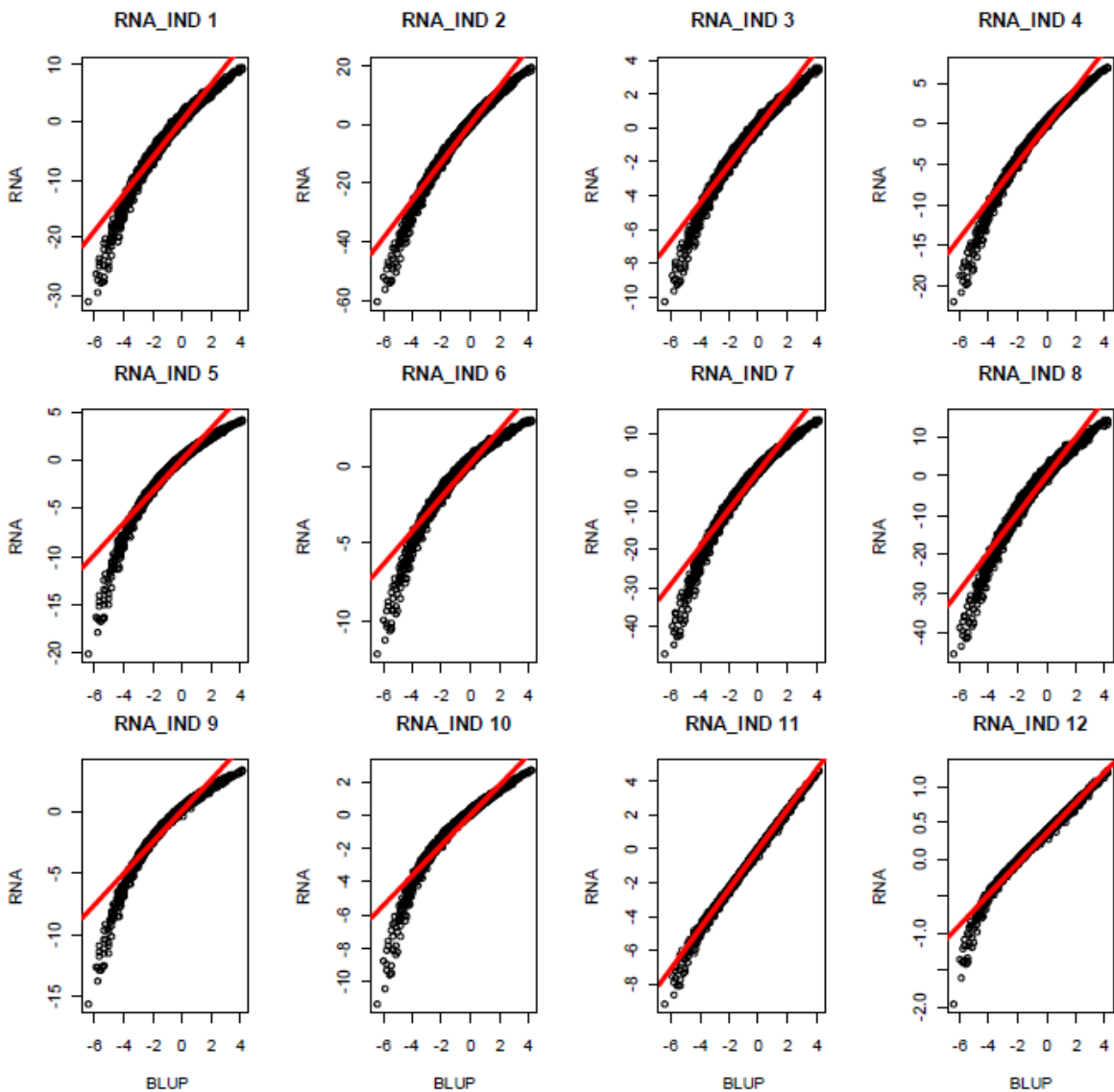


Figura 1. Análise gráfica do valor genético de árvores individuais de eucalipto predito pelo método BLUP vs. o valor genético predito por RNA na fase de validação das redes neurais artificiais.

Entre os modelos de RNA para predição do valor genético de famílias, dois alcançaram 100% de coincidência entre os selecionados em relação ao método REML/BLUP para os três níveis de seleção (FAM 8 e FAM 12). Porém, nos modelos de famílias, houve uma tendência geral de subestimar os valores genéticos, e estes dois modelos estão entre os que estimaram valores mais abaixo dos estimados pelo método BLUP (Tabela 5 e Figura 2).

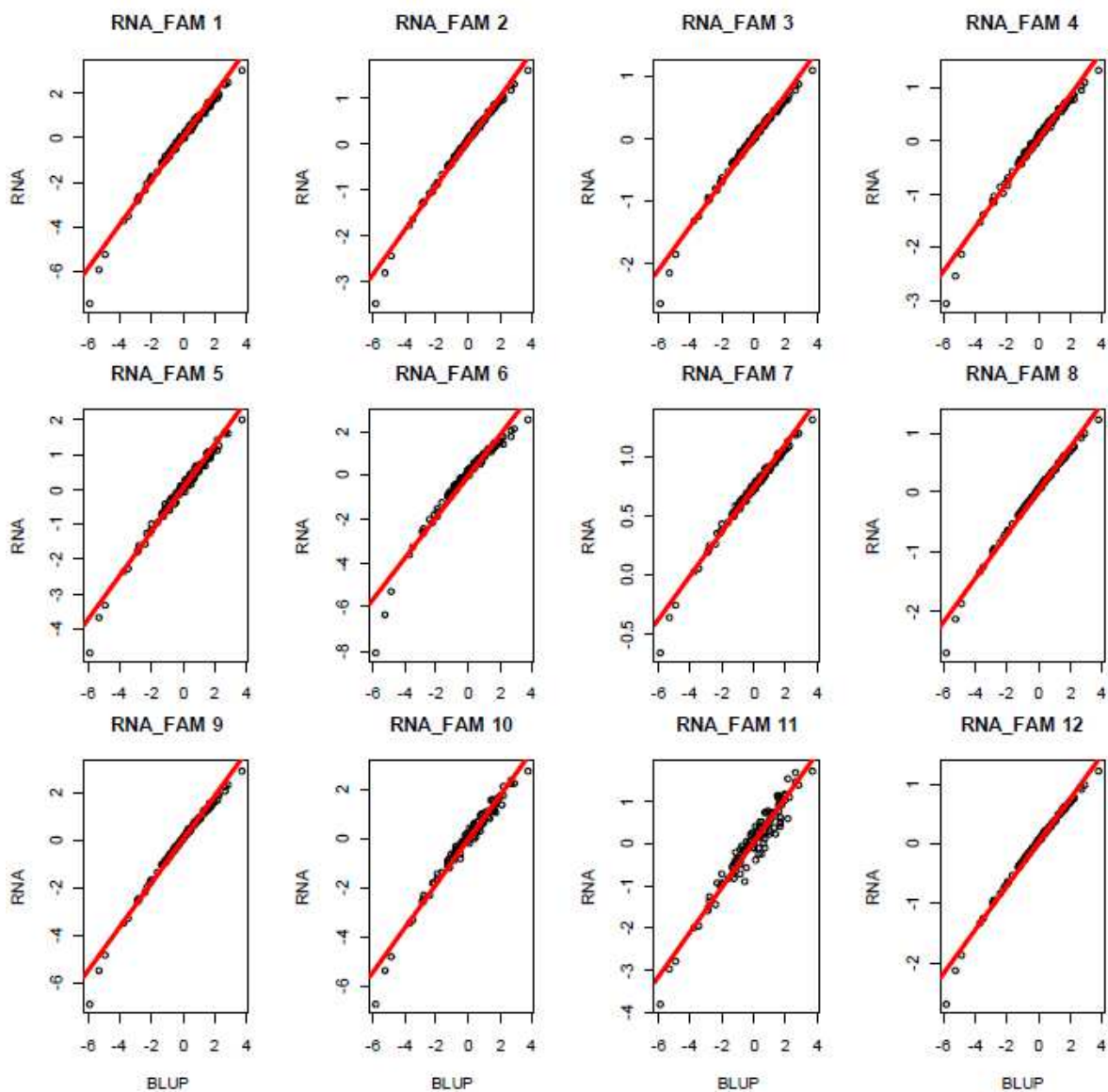


Figura 2. Análise gráfica do valor genético de famílias de eucalipto predito pelo método BLUP vs. o valor genético predito por RNA na fase de validação das redes neurais artificiais.

Essa tendência pode ter ocorrido porque a amplitude das progênes dentro do experimento, utilizada como variável de entrada das RNA com o intuito de facilitar a identificação da variação dentro das famílias, pode não ter sido suficiente para discriminação precisa dos efeitos de blocos ou de parcelas na etapa de validação das RNA. Desta forma, esses efeitos podem ter sido superestimados em razão das maiores amplitudes encontradas nos dados de validação, causando, conseqüentemente, a subestimação dos valores genéticos aditivos nesta etapa e comprometendo então a capacidade de generalização dos modelos. Assim, considerando as conseqüências da subestimação dos valores genéticos, optou-se pela escolha do modelo FAM 9 como o melhor na predição do valor genético de famílias, uma vez que este modelo proporcionou um dos menores desvios em relação à estimação do método REML/BLUP, e apresentou apenas uma troca de posição para o nível mais intenso de seleção (Tabela 5), excluindo apenas a pior família entre as selecionadas por este método.

Na aplicação conjunta dos melhores modelos visando à seleção entre e dentro de famílias com base no valor genético, as RNA também apresentaram ótimo desempenho, alcançando valores para porcentagem de coincidência das árvores selecionadas pelos métodos REML/BLUP e RNA e para correlação de postos de Spearman entre as árvores coincidentes, próximos de 100% (Tabela 6). Com base nestes dois critérios, as RNA apresentaram melhor desempenho na seleção entre e dentro de famílias quando comparada com a seleção individual realizada com a mesma intensidade de seleção (4,56%).

Tabela 6. Desempenho da aplicação conjunta dos modelos de redes neurais artificiais para seleção entre e dentro de famílias.

RNA	% COINCIDENTES	ρ -c(%)	GS (%)	
			BLUP	RNA
FAM 9- IND 11	99.38 (160 DE 161)	99.87	22.95	26.54
IND 11	96.89 (156 DE 161)	98.62	25.63	29.39

FAM 9-IND 11 = Aplicação conjunta dos melhores modelos para predição do valor genético aditivo de famílias e individuais; % COINCIDENTES = porcentagem coincidência na seleção pelo método BLUP e pela RNA em; ρ -c(%) = correlação de postos de Spearman entre as árvores coincidentes; GS = Ganho com a seleção.

Em relação ao GS(%), as RNA para seleção entre e dentro de famílias apresentaram a mesma tendência de superestimação dos valores genéticos encontrada no modelo de árvores individuais, isso já era esperado uma vez que os valores genéticos foram estimados pelo modelo de árvores individuais (IND 11), enquanto o melhor modelo de famílias (FAM 9) foi utilizado apenas para indicar as melhores famílias. O GS(%) estimado na seleção entre e dentro de famílias pela aplicação conjunta dos melhores modelos, foi inferior ao obtido com a seleção individual para mesma intensidade de seleção (Tabela 6), entretanto, esse fato não tem relação com o desempenho das RNA, e sim com a estrutura genética da população selecionada, uma vez que, enquanto a seleção individual é feita com base apenas no valor genético de cada árvore individualmente, na seleção entre e dentro de famílias, apesar do aumento da possibilidade de manutenção da variabilidade genética para futuros ciclos de seleção, pode ocorrer a eliminação de bons genótipos pela exclusão da família, por não estarem entre as melhores árvores da família, ou ainda devido à seleção de genótipos de menor valor em virtude do elevado valor da família (Azevedo et al., 2015; Costa et al., 2016).

4. Conclusão

Redes Neurais Artificiais do tipo múltiplas camadas apresentam bom desempenho na estimação do valor genético para a seleção individual, entre famílias e entre e dentro de famílias de *Eucalyptus grandis* utilizando apenas dados fenotípicos como variáveis de entrada. Sendo assim, as RNA têm grande potencial para serem utilizadas como ferramenta alternativa no melhoramento florestal. Entretanto, é importante ressaltar que os modelos de RNA foram validados com dados de progênies de uma única espécie de eucalipto, e que trabalhos subsequentes serão necessários para comprovação do bom desempenho dos modelos quando aplicado a dados de outras espécies florestais.

5. Referências

- Azevedo LPA, Costal RB, Martinez DT, Tsukamoto Filho AA, Brondani GE, Baretta MC, Ajala WV (2015) Seleção genética em progênies de *Eucalyptus camaldulensis* em área de cerrado matogrossense. **Ciência Rural** 45:2001-2006.
- Costa RML, Estopa RA, Biernaski FA, Mori ES (2016) Predição de ganhos genéticos em progênies de *Eucalyptus benthamii* Maiden & Cabbage por diferentes métodos de seleção. **Scientia Forestalis** 44:105-113.
- Cruz CD (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum Agronomy** 35:271-276.
- Dias PC, Xavier A, Resende MDV, Barbosa MHP, Biernaski FA, Estopa RA (2018) Genetic evaluation of *Pinus taeda* clones from somatic embryogenesis and their genotype x environment interaction. **Crop Breeding and Applied Biotechnology** 18:55-64.
- Ferreira RT, Viana AP, Barroso DG, Resende MDV, Amaral Júnior AT (2012) *Toona ciliata* genotype selection with the use of individual BLUP with repeated measures. **Scientia Agricola** 69:210-216.
- Hamilton MG, Acuna M, Wiedemann JC, Mitchell R, Pilbeam DJ, Brown MW, Potts BM (2015) Genetic control of *Eucalyptus globulus* harvest traits. **Canadian Journal of Forest Research** 45:615–624.
- Haykin S (2009) **Neural networks and learning machines**. 3 ed. Prentice Hall, 906p.
- Henriques EP, Moraes CB, Sebbenn AM, Tomazello Filho M, Moraes MLT, Mori ES (2017) Estimativa de parâmetros genéticos para caracteres silviculturais e densidade do lenho em teste de progênies de *Eucalyptus urophylla*. **Scientia Forestalis** 45: 119-128.
- Kainer D, Stone EA, Padovan A, Foley WJ, Külheim C (2018) Accuracy of Genomic Prediction for Foliar Terpene Traits in *Eucalyptus polybractea*. **G3: Genes|Genomes|Genetics** 8:2573-2583.
- Laviola BG, Silva DAS, Juhás ACP, Rocha RB, Oliveira RJP, Albrecht JC, Alves AA, Rosado TB (2014) Desempenho agrônômico e ganho genético pela seleção de pinhão-mansão em três regiões do Brasil. **Pesquisa Agropecuária Brasileira** 49:356-363.
- Leal AJF, Miguel EP, Baio FHR, Neves DC, leal UAS (2015) Redes neurais artificiais na predição da produtividade de milho e definição de sítios de manejo diferenciado por meio de atributos do solo. **Bragantia** 74:436-444.

Macedo HR, Freitas MLM, Moraes MLT, Zanata M, Sebben AM (2013) Variação, herdabilidade e ganhos genéticos em progênies de *Eucalyptus tereticornis* aos 25 anos de idade em Batatais-SP. **Scientia Forestalis** 41:533-540.

Miranda AC, Moraes MLT, Silva PHM, Sebbenn AM (2015) Ganhos genéticos na seleção pelo método do índice multifeitos em progênies polinização livre de *Eucalyptus grandis* Hill ex Maiden. **Scientia Forestalis** 43:203-209.

Mora F, Rubilar R, Emhart VI, Saavedra J (2013) Predicción bayesiana de parámetros genéticos en clones de *Eucalyptus globulus* bajo condiciones de suplemento hídrico. **Ciência Florestal** 23:529-536.

Moraes CB, Carvalho EV, Zimback L, Luz OSL, Pieroni GB, Mori ES, Leal TCAB (2015) Variabilidade genética em progênies de meios-irmãos de eucaliptos para tolerância ao frio. **Revista Árvore** 39:1047-1054.

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. **Biometrika** 58:545-554.

Resende MDV (2007) **SELEGEN-REML/BLUP: Sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Brasília, Colombo: Embrapa Florestas, 359 p.

Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília, Colombo: Embrapa Florestas, 975p.

Rosado AM, Rosado TB, Resende Júnior MFR, Bhering LL, Cruz CD (2009) Ganhos genéticos preditos por diferentes métodos de seleção em progênies de *Eucalyptus urophylla*. **Pesquisa Agropecuária Brasileira** 44:1653-1659.

Santos AM, Seixas JM, Pereira BB, Medronho RA (2005) Usando Redes Neurais Artificiais e Regressão Logística na Predição da Hepatite A. **Revista Brasileira de Epidemiologia** 8:117-26.

Silva AM, Canuto DSO, Moraes MLT (2017) Genetic parameters in riparian reforestation without structure of progenies by mixed models (reml/blup). **CERNE**, 23:501-506.

Silva GN, Tomaz RS, Sant'anna IC, Nascimento M, Bhering LL, Cruz CD (2014) Neural networks for predicting breeding values and genetic gains. **Scientia Agricola** 71:494-498.

Silva IN, Spattl DH, Flauzino RA (2010) **Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas**. São Paulo, ARTLIBER, 399p.

Silva WM, Moraes MLT, Pupin S, Silva JM, Sebbenn AM, Pavan BE (2016) Influência do tamanho de parcelas experimentais na seleção de progênies de *E. camaldulensis* Dehnh. **Scientia Forestalis** 44: 979-986.

Soares FC, Robaina AD, Peiter MX, Russi JL, Vivan GA (2014) Redes neurais artificiais na estimativa da retenção de água do solo. **Ciência Rural** v.44:293-300.

Souza CS, Freitas MLM, Moraes MLT, Sebbenn AM (2011) Estimativas de parâmetros genéticos para caracteres quantitativos em progênies de polinização aberta de *Eucalyptus urophylla*. **Floresta** 41: 847 - 856.

STATSOFT. **Statistica** (data analysis software system), version 7. Tulsa: StatSoft Inc., 2004.

Ventura RV, Silva MA, Medeiros TH, Dionello NL, Madalena FE, Fridrich AB, Valente BD, Santos GG, Freitas LS, Wenceslau RR, Felipe VPS, Corrêa GSS (2012) Uso de redes neurais artificiais na predição de valores genéticos para peso aos 205 dias em bovinos da raça Tabapuã. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia** 64:411-418.

Vieira GC, Mendonça AR, Silva GF, Zanetti SS, Silva MM, Santos AR (2018) Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. **Science of the Total Environment** 619–620:1473–1481.

Zanuncio AJV, Carvalho AG, Silva LF, Silva MG, Carneiro ACO, Colodette JL (2017) Prediction of the physical, mechanical and colorimetric properties of *Eucalyptus grandis* heat-treated wood using artificial neural networks. **Scientia Forestalis** 45: 109-118.

Zaruma DUG, Canuto DSO, Pupin S, Cambuim J, Silva AM, Mori ES, Sebbenn AM, Moraes MLT (2015) Variabilidade genética em procedências e progênies de *Dipteryx alata* vogel para fins de conservação genética e produção de sementes. **Scientia Forestalis** 43: 609-615.

Capítulo 4 - Redes Neurais Artificiais para seleção simultânea de progênies de eucalipto quanto à produtividade, estabilidade e adaptabilidade

RESUMO - A expansão do setor florestal para áreas com condições climáticas limitantes e a instabilidade climática atual têm comprometido a produtividade do eucalipto no Brasil, e as redes neurais artificiais (RNA) surgem como ferramentas promissoras na análise do comportamento de genótipos frente às variações ambientais. Este trabalho objetivou avaliar o desempenho das RNA como ferramenta de seleção simultânea de progênies de *Eucalyptus sp* quanto à produtividade, estabilidade e adaptabilidade. Para o treinamento das RNA foram utilizados dados simulados de 27 conjuntos de experimentos, e para validação dos modelos foram utilizados dados de testes de progênies de *Eucalyptus grandis* instalados em quatro locais, em delineamento de blocos casualizados, com várias plantas por parcela e número de repetições e progênies diferentes em cada local. Foram testadas diferentes arquiteturas de RNA de múltiplas camadas com algoritmo de aprendizado por retropropagação do erro. Como balizadores dos resultados obtidos com as RNA, foram usados os métodos de adaptabilidade e estabilidade da Média Harmônica da Performance Relativa dos Valores Genéticos (MHPRVG) e GGE-Biplot. A rede com três camadas ocultas e quatro neurônios por camada oculta apresentou melhor desempenho na seleção das progênies quanto à produtividade, estabilidade e adaptabilidade, com resultados condizentes com os encontrados por meio dos métodos MHPRVG e GGE Biplot e, portanto, as RNA de múltiplas camadas demonstram bom desempenho na seleção simultânea quanto à produtividade, estabilidade e adaptabilidade de progênies de *Eucalyptus grandis*.

Palavras-chave: *Eucalyptus*, Inteligência artificial, interação G x A, melhoramento florestal, MHPRVG, RNA

CHAPTER 4 - Artificial Neural Networks for simultaneous selection of eucalyptus progenies for productivity, stability and adaptability

ABSTRACT - The expansion of the forest sector to areas with limiting climatic conditions and current climate instability have compromised eucalyptus productivity in Brazil, and artificial neural networks (ANN) appear as promising tools in the analysis of genotype behavior in the face of environmental variations. This work aimed to evaluate the performance of ANN as a tool for simultaneous selection of *Eucalyptus* sp progenies for productivity, stability and adaptability. For the training of the ANN, simulated data of 27 sets of experiments were used, and for validation of the models data of *Eucalyptus grandis* progenies were used in four locations, in a randomized block design, with several plants per plot and number of replications and different progenies at each site. Different architectures of multiple layer ANN were tested with algorithm of learning by back propagation of the error. As a marker of the results obtained with RNA, the methods of adaptability and stability of the Harmonic Mean of Relative Performance of Genetic Values (HMRPGV) e GGE-Biplot. The network with three hidden layers and four neurons per hidden layer showed better performance in selection of progenies for productivity, stability and adaptability, with results consistent with those found through the HMRPGV and GGE Biplot methods and, therefore, the multilayer ANN demonstrate good performance in the simultaneous selection of productivity, stability and adaptability of *Eucalyptus grandis* progenies.

Keywords: *Eucalyptus*, Artificial intelligence, G x E interaction, forest improvement, HMRPGV, ANN

1. Introdução

O crescimento da área com florestas plantadas nos últimos anos no Brasil se deve exclusivamente ao aumento da área cultivada com eucalipto no país, que chegou a uma taxa de 2,4% a.a. entre 2011 e 2016 (IBÁ, 2017). Entretanto, esta expansão vem ocorrendo em áreas com condições climáticas cada vez mais específicas e limitantes ao desenvolvimento da eucaliptocultura e, além disso, a instabilidade climática atual tem comprometido a produtividade do eucalipto em áreas tradicionalmente cultivadas com a cultura, em função do aumento da intensidade e de maiores períodos de ocorrência de fatores estressantes (Silva et al., 2016).

A resposta de um genótipo para um determinado caráter é resultado da combinação dos efeitos genotípicos, ambientais e dos efeitos da interação genótipo x ambiente (G x A), ou seja, das diferentes respostas do genótipo frente às mudanças ambientais. Nesse sentido, estudos que forneçam informações sobre a capacidade dos genótipos apresentarem comportamento previsível frente às variações ambientais (estabilidade), e de responderem positivamente quando essas variações forem favoráveis (adaptabilidade), são essenciais para os programas de melhoramento genético (Santos et al., 2016).

Existem diversos métodos que possibilitam detalhar informações sobre as respostas dos genótipos diante das variações ambientais por meio de análise de estabilidade e adaptabilidade. Dentre estes, têm se destacado no melhoramento florestal o método GGE Biplot, que consiste em um sistema de análises que permite responder questões sobre o comportamento dos genótipos ao longo do gradiente ambiental por meio de abordagem gráfica (Correia et al., 2010; Santos et al., 2015; Yan e Tinker, 2006), e o método da Média Harmônica da Performance Relativa dos Valores Genéticos (MHPRVG), preditos pelo Método REML/BULP (Resende, 2002), que permite a seleção simultânea quanto à produtividade, estabilidade e adaptabilidade dos genótipos (Dias et al., 2018; Engel et al., 2016; Oliveira et al., 2018).

Os métodos citados, assim como outros utilizados com a mesma finalidade, têm em comum o fato de serem baseados em análises estatísticas convencionais,

portanto, apresentam certas exigências quanto a pressuposições em relação às variáveis estudadas. Entretanto, nem sempre é fácil atender tais pressuposições, que quando violadas podem comprometer o resultado e a confiabilidade das análises. Nesse contexto, as Redes Neurais Artificiais (RNA) surgem como alternativa para seleção simultânea de genótipos quanto à produtividade, estabilidade e adaptabilidade. As RNA são sistemas paralelos distribuídos compostos por unidades de processamento simples (neurônios artificiais) dispostas em uma ou mais camadas e interligadas por um grande número de conexões (Braga et al., 2012). A RNA tem a capacidade de aprender por meio de um conjunto de exemplos e generalizar a informação dando respostas coerentes para dados não conhecidos, e diferentemente das análises estatísticas comumente utilizadas nas diversas áreas do melhoramento genético, não possuem pressuposições quanto ao modelo, uma vez que seus resultados dependem do aprendizado e não da distribuição dos dados em si (Cruz e Nascimento, 2018). Assim, o presente trabalho objetivou avaliar o desempenho das redes neurais artificiais na seleção de progênies de *Eucalyptus sp* quanto à produtividade, estabilidade e adaptabilidade simultaneamente.

2. Material e métodos

Na fase de treinamento das Redes Neurais Artificiais (RNA) foram utilizados dados simulados de 27 conjuntos de experimentos, onde cada conjunto contou com quatro experimentos em Delineamento de Blocos Casualizados (DBC) com 100 progênies, quatro repetições (blocos) e seis plantas por parcela. Para cada experimento foram estabelecidos os valores da média (10, 15 e 20 cm), da herdabilidade (30, 40 e 50%) e do coeficiente de variação experimental (10, 20 e 30%), com base em valores comumente encontrados na literatura para o diâmetro à altura do peito (DAP) em testes de progênies de *Eucalyptus sp*. (Henriques et al., 2017; Moraes et al., 2015; Santos et al., 2015). Para a análise de produtividade, estabilidade e adaptabilidade das progênies simuladas, assumiu-se que os experimentos do mesmo conjunto continham as mesmas 100 progênies. A

simulação dos dados foi realizada com o auxílio do programa computacional GENES (Cruz, 2013).

A validação dos modelos de RNA foi feita com dados de testes de progênies de *Eucalyptus grandis* instalados em quatro locais, em DBC, conforme Tabela 1.

Tabela 1. Número de progênies, de blocos e de plantas por parcela de quatro testes de progênies de *Eucalyptus grandis*.

Local	Número de progênies	Blocos	Plantas por parcela
1	115	6	6
2	127	5	5
3	131	4	6
4	125	8	4

O modelo do neurônio artificial, expresso na equação abaixo, forma a base para o projeto das redes neurais artificiais, e conta com pesos sinápticos ($W_{k,j}$), sinais de entrada (X_j) que são multiplicados, ou seja, ponderados pelos respectivos pesos sinápticos, um somador (Σ) para os sinais de entrada já ponderados, uma função de ativação $f(x)$ que restringe a amplitude da saída do neurônio, e um viés (b_k) que aumenta ou diminui a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo, respectivamente (Haykin, 2009):

$$Y = f\left(\sum_{j=1}^m W_{k,j} \cdot X_j + b_k\right)$$

Neste trabalho foram utilizadas redes neurais artificiais do tipo múltiplas camadas com o algoritmo de aprendizado supervisionado por retropropagação do erro, função de ativação do tipo tangente hiperbólica, e número máximo de iterações (ou épocas) para cada rede igual a 1.000. As quatro variáveis utilizadas como entradas das RNA foram a média geral da progênie considerando todos os locais em que ela ocorre, a diferença entre a média geral da progênie e a média geral dos experimentos, a melhor e a pior média da progênie entre os locais em que ela ocorre. A variável utilizada como saída das redes foi a Média Harmônica da Performance Relativa dos Valores Genéticos (MHPRVG) preditos da progênie.

As análises realizadas com base no método MHPRVG foram feitas empregando o software genético-estatístico SELEGEN desenvolvido por Resende

(2007), assumindo-se os dados como sendo de experimentos em DBC, com várias observações por parcela, em vários locais, seguindo o modelo proposto por Resende (2007):

$$y = Xr + Zg + Wp + Ti + e$$

em que y é o vetor de dados, r é o vetor dos efeitos de repetição (assumidos como fixos) somados à média geral, g é o vetor dos efeitos genotípicos (assumidos como aleatórios), p é o vetor dos efeitos de parcela (aleatórios), i é vetor dos efeitos da interação genótipo x ambiente (aleatórios) e e é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

Para que os resultados das RNA não fossem influenciados pelas diferentes magnitudes das variáveis, os dados foram normalizados no intervalo de 0 a 1 por meio da seguinte equação:

$$X_{\text{norm}} = \left(\frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \right)$$

onde X_{norm} é o valor normalizado, X_i é o valor real a ser normalizado, X_{min} e X_{max} são os valores mínimo e máximo da variável, respectivamente.

Os 15 modelos de RNA avaliados variaram de 1 a 3 quanto ao número de camadas ocultas. O número de neurônios foi estabelecido conforme o método de Fletcher-Gloss (Silva et al., 2010) de acordo com a seguinte condição:

$$(2 \cdot \sqrt{n} + n2) \leq n1 \leq (2 \cdot n + 1)$$

em que n é o número de entradas da rede, $n1$ é o número de neurônios da camada oculta e $n2$ é a quantidade de neurônios na camada de saída (Tabela 2).

Tabela 2. Arquitetura das redes neurais artificiais (RNA) avaliadas.

RNA	ARQUITETURA	RNA	ARQUITETURA	RNA	ARQUITETURA
1	4-5-1	6	4-5-5-1	11	4-5-5-5-1
2	4-6-1	7	4-6-6-1	12	4-6-6-6-1
3	4-7-1	8	4-7-7-1	13	4-7-7-7-1
4	4-8-1	9	4-8-8-1	14	4-8-8-8-1
5	4-9-1	10	4-9-9-1	15	4-9-9-9-1

Para cada um dos modelos foram treinadas 100 redes neurais, sendo selecionada a melhor de cada modelo com base no desempenho avaliado pela correlação entre os valores estimados pelo método MHPRVG e os estimados pela RNA ($r_{y\hat{y}}$), pela tendência dos modelos em sub ou superestimar os valores do método MHPRVG, indicada pelo erro médio (ME), e pela magnitude dos erros dada pela raiz do erro quadrático médio (RMSE), conforme as expressões a seguir:

$$r_{y\hat{y}} = \frac{\text{Cov}(y, \hat{y})}{\sqrt{S^2(y) \cdot S^2(\hat{y})}}$$

em que Cov é a covariância, S^2 a variância, y é o valor estimado pelo método MHPRVG e \hat{y} o valor estimado pela RNA.

$$\text{ME} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}$$

em que: y_i é o valor estimado da i ésima variável pelo método MHPRVG; \hat{y}_i é valor estimado da i ésima variável pela RNA e n é o tamanho da amostra.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

em que: y_i , \hat{y}_i e n , conforme apresentado anteriormente.

As redes selecionadas foram avaliadas na etapa de validação, quanto à capacidade de generalização de acordo com os mesmos critérios utilizados na etapa de treinamento, além da porcentagem de coincidência entre os métodos MHPRVG e RNA na seleção das 30 melhores progênies, e da correlação de postos de Spearman (ρ), conforme a seguinte expressão:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)}$$

em que: d é a diferença entre os postos das progênies de acordo com os métodos MHPRVG e RNA, e n é o número de progênies.

Nesta etapa, os modelos foram avaliados primeiramente quanto à capacidade de lidar com dados balanceados em relação ao número de progênies, portanto, aplicados apenas às progênies comuns a todos os locais. Em seguida os modelos foram avaliados, de acordo com os mesmos critérios, quanto ao desempenho no

caso de dados desbalanceados, sendo incluídas progênes que não estavam presentes em todos os locais. Todo o processo das análises de RNA foi feito com o software Statistica (Statsoft, 2004).

A fim de analisar as progênes não coincidentes pelo método MHPRVG e pelas RNA, foi realizada a análise gráfica do método GGE Biplot com base nas médias fenotípicas, considerando o seguinte modelo:

$$\bar{Y}_{ij} - \mu = G_i + A_j + GA_{ij}$$

em que: \bar{Y}_{ij} representa a média fenotípica da progênie i no ambiente j, μ é a média geral, G_i é o efeito da progênie i, A_j é o efeito do ambiente j, e GA_{ij} é o efeito da interação entre a progênie i e o ambiente j.

O GGE Biplot não separa o efeito do genótipo do efeito da interação genótipo x ambiente, mantendo-os juntos em dois termos multiplicativos, assim a construção do gráfico GGE Biplot se deu pela dispersão simples desses termos por meio da decomposição em valores singulares (Yan e Kang, 2003). As análises de GGE Biplot foram realizadas no pacote *GGEbiplotGUI*, disponível no software estatístico R (R Core Team, 2018).

3. Resultados e discussão

Todos os modelos selecionados na etapa de treinamento, independentemente da arquitetura, apresentaram bom desempenho no reconhecimento de padrões para determinação da produtividade, estabilidade e adaptabilidade das progênes simuladas, alcançando correlações (r_{yy}) positivas e de alta magnitude, entretanto, notou-se uma pequena superioridade dos modelos com três camadas ocultas quanto a este critério de avaliação (Tabela 3). Nesta etapa os valores do erro médio (ME) mantiveram-se próximos de zero para todos os casos, indicando que não houve tendência clara de sub ou superestimação dos valores estimados pelo método Média Harmônica da Performance Relativa dos Valores Genéticos (MHPRVG). Da mesma forma, os valores da raiz quadrada do erro quadrático médio foram baixos em todos os casos, indicando boa precisão dos modelos na etapa de treinamento (Tabela 3).

Tabela 3. Desempenho das redes neurais artificiais (RNA) selecionadas nas fases de treinamento.

RNA ²	$r_{y\hat{y}}$ (%)	ME	RMSE	RNA ²	$r_{y\hat{y}}$ (%)	ME	RMSE	RNA ²	$r_{y\hat{y}}$ (%)	ME	RMSE
1	95.84	0.00004	0.00343	6	96.92	-0.00003	0.00296	11	97.00	-0.00022	0.00293
2	96.06	0.00003	0.00335	7	96.24	0.00002	0.00328	12	96.95	0.00012	0.00303
3	96.37	-0.00010	0.00321	8	95.95	-0.00010	0.00337	13	97.14	-0.00024	0.00286
4	96.44	-0.00007	0.00318	9	96.89	-0.00002	0.00300	14	97.14	0.00006	0.00285
5	95.91	0.00010	0.00341	10	96.67	-0.00005	0.00308	15	97.34	0.00006	0.00275

² - RNA conforme Tabela 2. $r_{y\hat{y}}$ (%), ME e RSME - respectivamente, correlação, erro médio e raiz quadrada do erro quadrático médio entre os valores estimados pelo método MHPRVG e os estimados pela RNA

As correlações entre os valores estimados pelo método MHPRVG e pelas RNA na etapa de validação, realizada com as progênes comuns a todos os locais, variaram de 86,64 a 97,03 (Tabela 4). Nesta etapa nota-se que os três maiores valores para $r_{y\hat{y}}$, assim como para correlação de postos de Spearman, foram alcançados por modelos com três camadas ocultas, e que, em média, os menores valores para esses critérios foram obtidos pelas RNA com apenas uma camada oculta (Tabela 4). Da mesma forma, os valores mais distantes de zero para ME assim como os maiores valores de RMSE foram obtidos pelos modelos com apenas uma camada oculta, de maneira que a tendência e a precisão dos modelos, em média, melhoraram com o aumento gradativo do número de camadas ocultas (Tabela 4). Nas redes de múltiplas camadas que usam apenas uma camada oculta, os neurônios nesta camada tendem a interagir entre si globalmente, de forma que, em situações complexas, essa interação dificulta a melhoria do ajuste de um determinado ponto sem que outro ponto seja prejudicado, comprometendo desta forma a capacidade de generalização da rede (Haykin, 2001). Geralmente, duas camadas ocultas são suficientes para resolver a maioria dos problemas, entretanto, para dados que apresentam descontinuidade e quando a precisão é tida como um critério essencial, a utilização de uma terceira camada oculta pode melhorar o desempenho da rede (Karsoliya, 2012).

Tabela 4. Desempenho das redes neurais artificiais (RNA) para dados balanceados na etapa de validação.

RNA ²	$r_{\hat{y}y}$ (%)	ME	RMSE	ρ (%)	% COINCIDENTES
1	89.86	-3.7530	3.7691	91.87	80.00
2	88.97	-3.2108	3.2257	89.66	80.00
3	91.45	-3.0622	3.0787	90.10	80.00
4	87.79	-3.3266	3.3506	82.04	73.33
5	88.84	-3.1155	3.2187	83.94	76.66
6	94.36	-1.4204	1.4285	92.88	80.00
7	93.08	-1.3360	1.3372	90.20	83.33
8	91.68	0.7100	0.7609	93.93	80.00
9	94.37	-0.3202	0.3284	93.25	73.33
10	93.34	-0.1539	0.3167	88.58	73.33
11	95.26	0.0853	0.3065	95.66	80.00
12	95.15	0.0086	0.0435	94.77	83.33
13	97.03	-0.0022	0.0210	97.74	86.66
14	93.90	-0.9644	0.9697	90.51	76.66
15	86.64	-0.8904	0.8935	84.50	70.00

² - RNA conforme Tabela 2. $r_{\hat{y}y}$ (%), ME e RSME - respectivamente, correlação, erro médio e raiz quadrada do erro quadrático médio entre os valores estimados pelo método MHPRVG e os estimados pela RNA. % Coincidentes - porcentagem, dentre as 30 famílias superiores pelo método da MHPRVG, de coincidentes com a seleção praticada sobre as RNA.

Entre os modelos com três camadas ocultas, apesar de serem encontrados os melhores valores para todos os critérios de avaliação, nota-se que houve queda no desempenho dos modelos com mais neurônios por camada oculta (Tabela 4). Este fato pode estar relacionado com o problema conhecido como *overfitting*, que ocorre quando, em função do aumento desnecessário do número de neurônios das camadas ocultas, a rede memoriza os dados de treinamento perdendo assim a capacidade de generalização (Karsoliya, 2012; Reis et al., 2016).

Considerando que as análises em múltiplos ambientes têm como principal objetivo a seleção de genótipos que, além de produtivos, tenham a capacidade de aproveitar positivamente os estímulos ambientais e sejam capazes de apresentar um comportamento previsível nos diferentes ambientes, uma seleção equivocada

pode comprometer o avanço dos programas de melhoramento genético ou mesmo a produtividade de uma empresa. Neste sentido, a porcentagem de coincidência entre os métodos MHPRVG e RNA na seleção das 30 melhores progênes foi vista como um critério importante na escolha dos melhores modelos de RNA. Entre as progênes comuns a todos os locais esta porcentagem variou de 70% (21 árvores coincidentes) a 86,66% (26 árvores coincidentes), sendo o modelo RNA 13 o de melhor desempenho quanto a este e aos demais critérios de avaliação (Tabela 4).

Tendo em vista a capacidade do método MHPRVG em lidar com dados desbalanceados, e que esta é uma situação comum no melhoramento florestal, os modelos de RNA também foram testados quanto ao desempenho ao estimar a produtividade, estabilidade e adaptabilidade para o caso de dados desbalanceados quanto ao número de progênes, ou seja, com a inclusão de 22 progênes que não estavam presentes em todos os locais, e neste caso nota-se que, de maneira geral, houve uma pequena queda no desempenho dos modelos (Tabela 5). Essa redução no desempenho pode ser notada principalmente na porcentagem de progênes coincidentes entre as 30 selecionadas pelos dois métodos, na qual alguns modelos chegaram a apresentar uma redução de mais de 16% (RNA 12 e RNA 14). O modelo RNA 13 reduziu de 26 para 24 o número de árvores coincidentes entre as 30 selecionadas, e apesar de também apresentar uma pequena queda de desempenho no caso dos experimentos desbalanceados, continuou sendo o melhor de acordo com todos os critérios de avaliação (Tabela 5).

Sabendo-se da importância do acerto na seleção das melhores progênes, buscou-se analisar as progênes não coincidentes pelo método MHPRVG e pelo melhor modelo de rede neural artificial (RNA 13), por meio da análise gráfica do método GGE Biplot com base nas médias fenotípicas das progênes comuns a todos os locais.

Tabela 5. Desempenho das redes neurais artificiais (RNA) para dados desbalanceados na etapa de validação.

RNA ²	ryŷ(%)	ME	RMSE	ρ (%)	% COINCIDENTES
1	89.14	-3.7373	3.7522	90.50	73.33
2	89.19	-3.1610	3.1746	89.09	76.66
3	91.97	-3.0280	3.0429	90.85	73.33
4	87.62	-3.2723	3.2943	83.04	66.66
5	88.42	-3.0675	3.1538	84.54	70.00
6	90.11	-1.4245	1.4325	89.47	70.00
7	90.90	-1.3267	1.3279	88.07	70.00
8	91.16	0.7401	0.7837	93.45	66.66
9	93.36	-0.3140	0.3210	91.97	66.66
10	93.48	-0.1333	0.2862	89.47	66.66
11	93.53	0.0964	0.3029	93.48	73.33
12	94.25	0.0152	0.0433	93.02	66.66
13	95.46	0.0040	0.0235	95.35	80,00
14	91.20	-0.9662	0.9714	86.59	60.00
15	86.94	-0.8685	0.8710	85.83	70.00

² - RNA conforme Tabela 2. ryŷ(%), ME e RSME - respectivamente, correlação, erro médio e raiz quadrada do erro quadrático médio entre os valores estimados pelo método MHPRVG e os estimados pela RNA. % Coincidentes - porcentagem, dentre as 30 famílias superiores pelo método da MHPRVG, de coincidentes com a seleção praticada sobre as RNA.

Os gráficos da análise GGE Biplot apresentam as informações do percentual de variação explicada por cada componente principal (CP), sendo que juntos o CP1 e o CP2 explicaram 81% da variação total, indicando alta confiabilidade nos resultados. No gráfico de relação entre ambientes (Figura 1), as linhas que ligam os locais (ambientes) à origem do biplot são os vetores dos ambientes, e o cosseno do ângulo formado pelos vetores de dois ambientes indica a correlação entre eles (Yan e Tinker, 2006).

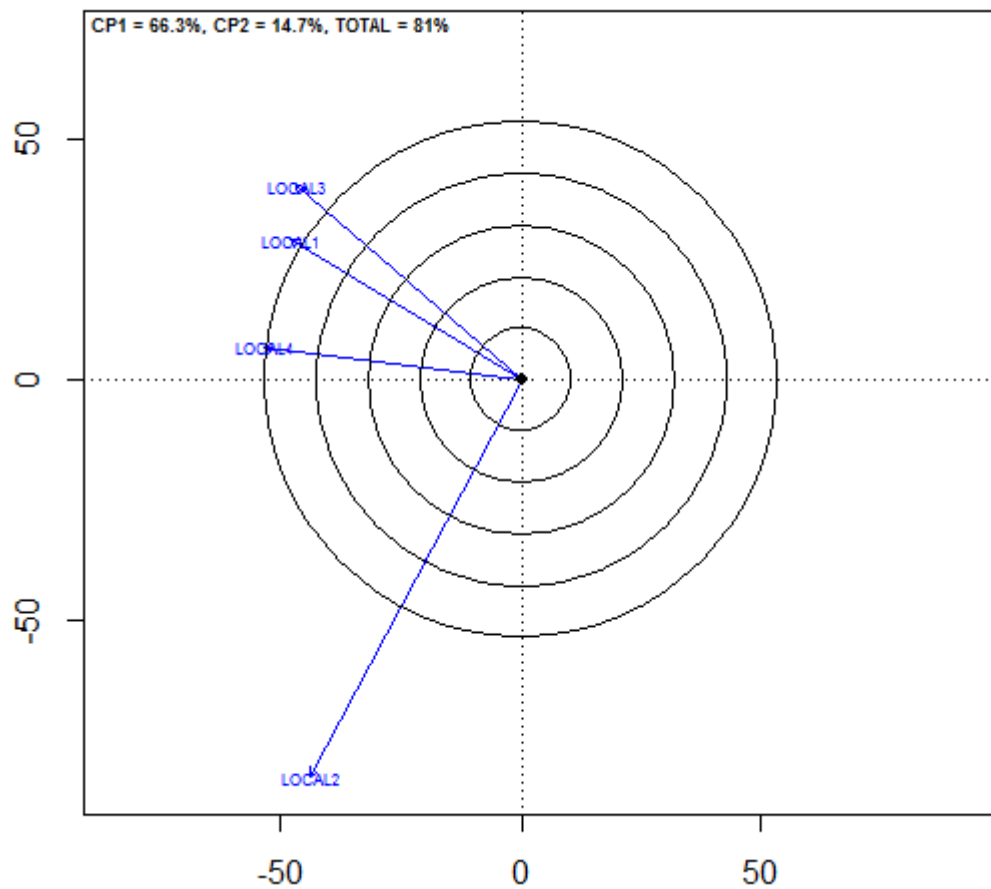


Figura 1. GGE Biplot “Relação entre ambientes” com base na média fenotípica do DAP de progênies de *Eucalyptus grandis* em quatro locais

Nesse sentido, é possível observar que o local 2 apresenta correlação negativa com os locais 1 e 3 devido ao ângulo obtuso formado pelos vetores desses locais (Figura 1), indicando que há interação genótipo x ambiente do tipo complexa (Yan e Tinker, 2006). Sendo assim, a indicação de um conjunto de genótipos que permita ganhos com a seleção nos diferentes locais deve ser feita com cuidado.

Na Figura 2, o eixo das abscissas passa através da média dos ambientes e a origem do biplot e é conhecido como Eixo de Ambiente Médio (EAM). A seta sobre este eixo aponta a direção do maior desempenho médio entre as progênies nos ambientes, e o eixo que passa pela origem do biplot e é perpendicular ao EAM indica a variabilidade das progênies, ou seja, infere sobre a estabilidade, sendo que quanto mais afastada de zero, em qualquer direção, maior a instabilidade da progênie (Bhartiya et al., 2017).

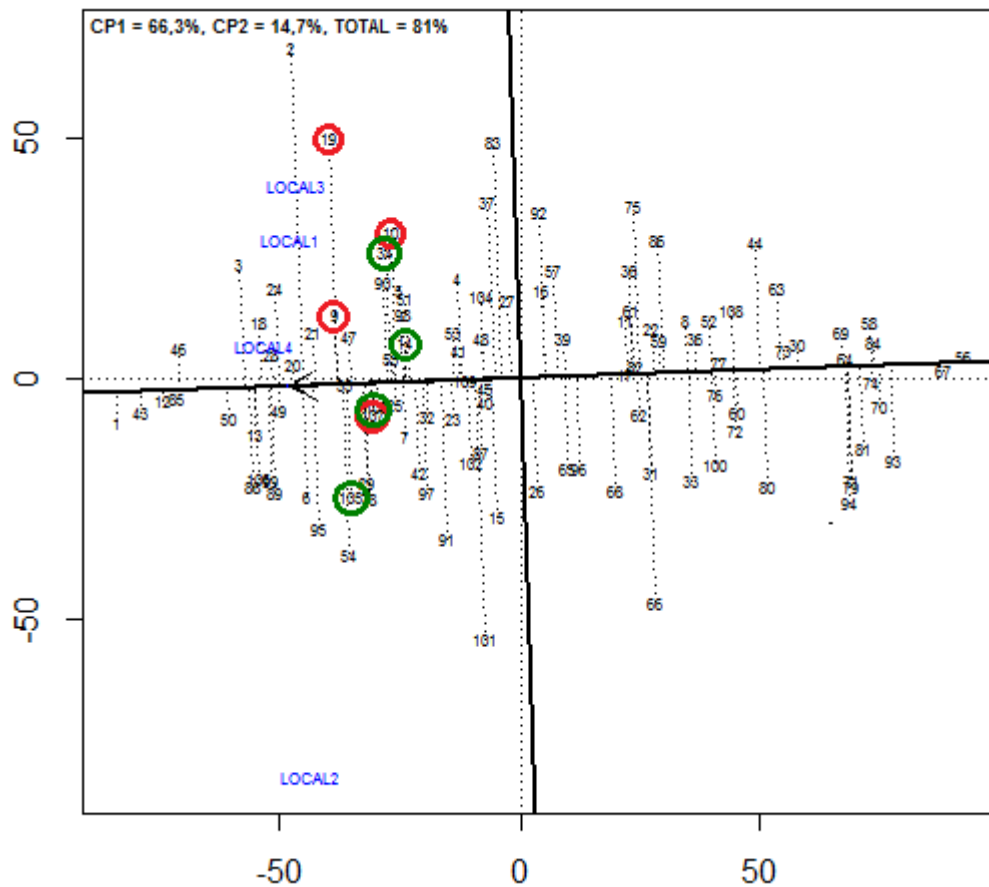


Figura 2. GGE Biplot “Média versus estabilidade” com base na média fenotípica do DAP de progênies de *Eucalyptus grandis* em quatro locais

As progênies marcadas com círculos vermelhos representam as que foram selecionadas pelo método MHPRVG e não foram selecionadas pela RNA, enquanto as progênies marcadas com círculos verdes representam as que foram selecionadas apenas pela rede neural artificial (Figura 2). Nota-se que as progênies não coincidentes, independentemente do método de seleção, ficaram acima da média, indicada pelo eixo das ordenadas, e não apresentaram grandes diferenças entre si quanto ao desempenho médio (Figura 2). Isso fica evidenciado no gráfico que simula o “genótipo ideal”, ou seja, um genótipo hipotético com maior desempenho médio e elevada estabilidade e adaptabilidade, representado pelo círculo menor no EAM marcado com uma seta (Figura 3), onde é possível selecionar as progênies utilizando as suas distâncias relativas a esse genótipo ideal como referência (Yan e Tinker, 2006). Desta forma, observa-se que as progênies não coincidentes estão distribuídas basicamente da mesma forma no gráfico, estando duas selecionadas

exclusivamente em cada método posicionadas no quinto círculo concêntrico e duas de cada método no sexto círculo, indicando que a seleção realizada com a RNA, apesar das progênes não coincidentes, não proporcionaria grandes diferenças no ganho com a seleção em relação à seleção realizada com o método MHPRVG (Figura 3).

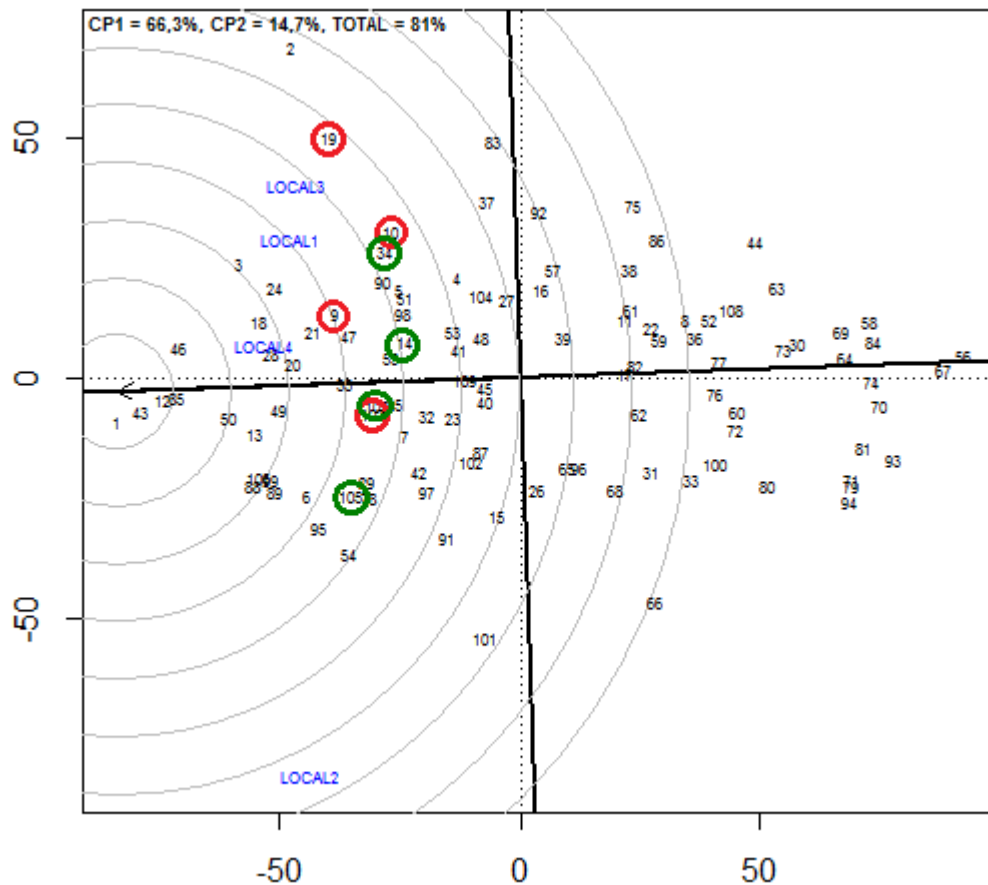


Figura 3. GGE Biplot “Classificação dos classificações dos genótipos com referência ao genótipo ideal” com base na média fenotípica do DAP de progênes de *Eucalyptus grandis* em quatro locais

4. Conclusão

As redes neurais artificiais de múltiplas camadas demonstram bom desempenho na seleção simultânea de progênes de *Eucalyptus grandis* quanto à produtividade, estabilidade e adaptabilidade, com resultados condizentes com os apresentados pelos métodos da Média Harmônica da Performance Relativa dos

Valores Genéticos (MHPRVG) e GGE Biplot. Entretanto, faz-se necessário a realização de trabalhos subsequentes para testar o desempenho do modelo quando aplicado a outras espécies florestais, assim como a diferentes níveis de interação genótipo x ambiente.

5. Referências

Bhartiya A, Aditya JP, Kumari V, Kishore N, Purwar JP, Agrawal A, Kant L (2017) Gge biplot & ammi analysis of yield stability in multi-environment trial of soybean [*Glycine max* (L.) merrill] genotypes under rainfed condition of north western Himalayan hills. **The Journal of Animal & Plant Sciences** 27:227-238.

Braga AP, Carvalho ACPLF, Ludermir TB (2012) **Redes Neurais Artificiais: Teoria e Aplicações**, 2 ed. LTC, 226p.

Correia I, Alía R, Yan W, David T, Aguiar A, Almeida MH (2010) Genotype x Environment interactions in *Pinus pinaster* at age 10 in a multi-environment trial in Portugal: a maximum likelihood approach. **Annals of Forest Science** 67:612–612.

Cruz CD (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum Agronomy** 35: 271-276.

Cruz CD, Nascimento M (2018) **Inteligência Computacional Aplicada ao Melhoramento Genético**. UFV 414p.

Dias PC, Xavier A, Resende MDV, Barbosa MHP, Biernaski FA, Estopa RA (2018) Genetic evaluation of *Pinus taeda* clones from somatic embryogenesis and their genotype x environment interaction. **Crop Breeding and Applied Biotechnology** 18:55-64.

Engel ML, Higa AR, Andrejow GP, Flôres Junior PC, Soares ID (2016) Genetic gain from different selection methods in *Eucalyptus macarthurii* progenies in different environments. **CERNE** 22: 299-308.

Haykin S (2009) **Neural networks and learning machines**. 3 ed. Prentice Hall, 906p.

Haykin S (2001) **Redes Neurais: princípios e práticas**. 2 ed. Bookman, 899p.

Henriques EP, Moraes CB, Sebbenn AM, Tomazello Filho M, Moraes MLT, Morl ES (2017) Estimativa de parâmetros genéticos para caracteres silviculturais e densidade do lenho em teste de progênies de *Eucalyptus urophylla*. **Scientia Forestalis** 45:119-128.

IBA - Indústria Brasileira de Árvores (2017): Anuário estatístico da indústria brasileira de árvores: ano base 2016. Brasília, 80p. Disponível em: http://www.iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2017.pdf 10 de jun. 2018.

Karsoliya S (2012) Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. **International Journal of Engineering Trends and Technology** 3:714-717.

Moraes CB, Carvalho EV, Zimback L, Luz OSL, Pieroni GB, Mori ES, Leal TCAB (2015) Variabilidade genética em progênies de meios-irmãos de eucaliptos para tolerância ao frio. **Revista Árvore** 39:1047-1054.

Oliveira TWG, Paula RC, Moraes MLT, Alvares CA, Miranda AC, Silva PHM (2018) Stability and adaptability for wood volume in the selection of *Eucalyptus saligna* in three environments. **Pesquisa agropecuária brasileira** 53:611-619.

R CORE TEAM (2018). **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing. Vienna. Disponível em: <https://www.R-project.org>.

Reis LP, Souza AL, Mazzel L, Reis PCM, Leite HG, Soares CPB, Torres CMME, Silva LF, Ruschel AR (2016) Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. **Forest Ecology and Management** 382:161 – 167.

Resende MDV (2007) **SELEGEN-REML/BLUP: Sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Brasília, Colombo: Embrapa Florestas, 359 p.

Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília, Colombo: Embrapa Florestas, 975p.

Santos GA, Nunes ACP, Resende MDV, Silva LD, Higa A, Assis TF (2016) An index combining volume and Pilodyn penetration to study stability and adaptability of *Eucalyptus* multi-species hybrids in Rio Grande do Sul, Brazil. **Australian Forestry** 79:.248–255.

Santos GA, Resende MDV, Silva LD, Higa A, Assis TF (2015) Interação genótipos x ambientes para produtividade de clones de *Eucalyptus* L'Hér. no estado do Rio Grande do Sul. **Revista Árvore** 39:81-91.

Silva PHM, Campoe OC, Paula, RC, Lee DJ (2016) Seedling Growth and Physiological Responses of Sixteen Eucalypt Taxa under Controlled Water Regim. **Forests** 7:1-13.

Silva IN, Spattl DH, Flauzino RA (2010) **Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas**. São Paulo, ARTLIBER, 399p.

STATSOFT. **Statistica** (data analysis software system), version 7. Tulsa: StatSoft Inc., 2004.

Yan W, Kang MS (2003) **GGE biplot analysis**: a graphical tool for breeders, geneticists, and agronomists. Flórida: Boca Raton 286p.

Yan W, Tinker NA (2006) Biplot analysis of multi-environment trial data: Principles and applications. **Canadian Journal of Plant Science** 86: 623-645.

Capítulo 5 – Considerações Finais

As Redes Neurais Artificiais (RNA) de múltiplas camadas apresentam bom desempenho na prognose do diâmetro à altura do peito (DAP), na predição de valor genético e na seleção quanto à produtividade, estabilidade e adaptabilidade em testes com eucalipto, sendo, portanto, ferramentas promissoras na prognose de crescimento e no melhoramento florestal.

No capítulo 2 ficou evidenciada a capacidade das RNA de extraírem informações detalhadas dos dados de treinamento, e utilizá-las para resolver problemas mesmo em condições distintas. No entanto, a pequena queda no desempenho das redes quanto à prognose em locais distantes e/ou distintos daquele para qual o foi modelo criado, evidencia a importância da utilização de genótipos adaptados a cada local e, como um próximo passo, seria importante a realização de estudos que permitam a caracterização dos diferentes locais de maneira a identificar os mais semelhantes quanto às condições ambientais, e assim testar o desempenho de modelos criados para cada grupo de ambientes.

Considerando o grande número de fatores que podem influenciar a estimativa do valor genético das árvores, nos capítulos 3 e 4 foi possível notar, mais uma vez, a capacidade das redes de captarem informações mesmo quando estas não são apresentadas de forma explícita. Além da vantagem da não exigência quanto às pressuposições em relação aos dados, as redes neurais, após o ajuste dos modelos, também se mostram mais simples de serem aplicadas em relação aos métodos mais comuns utilizados para estimativa de valor genético e seleção quanto à produtividade, estabilidade e adaptabilidade, podendo tornar o processo de seleção menos oneroso. Entretanto, é importante que em trabalhos futuros, os modelos selecionados nesta pesquisa sejam testados em outras condições como diferentes delineamentos experimentais, populações com estruturas genéticas distintas, diferentes espécies, diferentes níveis de contrastes entre ambientes, entre outras.

As aplicações das RNA no setor florestal, sobretudo no melhoramento genético, ainda são limitadas, no entanto, em função das suas vantagens, acredita-se que as redes neurais tenham maior visibilidade e procura por parte dos

pesquisadores deste setor nos próximos anos. Assim, a importância deste trabalho se dá, também, pela ampliação da base teórica das RNA aplicadas sob os conceitos de prognose de crescimento e melhoramento genético florestal.