

IARA GONÇALVES DOS SANTOS

**SELEÇÃO DE CARACTERES COMPLEXOS EM ALFAFA POR MEIO DE
INTELIGÊNCIA COMPUTACIONAL**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2017

**Ficha catalográfica preparada pela Biblioteca Central da
Universidade Federal de Viçosa - Campus Viçosa**

T

S237s
2017 Santos, Iara Gonçalves dos, 1992-
Seleção de caracteres complexos em alfafa por meio de
inteligência computacional / Iara Gonçalves dos Santos. - Viçosa, MG,
2017.
xi, 91f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.

Orientador: Cosme Damião Cruz.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Alfafa - Melhoramento genético. 2. *Medicago sativa*. 3. Redes neurais (Computação). 4. Inteligência artificial. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22 ed. 633.312

IARA GONÇALVES DOS SANTOS

**SELEÇÃO DE CARACTERES COMPLEXOS EM ALFAFA POR MEIO DE
INTELIGÊNCIA COMPUTACIONAL**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADA: 18 de julho de 2017.


Moysés Nascimento


Reinaldo de Paula Ferreira


Cosme Damião Cruz
(Orientador)

*Aos meus familiares e amigos,
por ajudarem a transformar meus
sonhos em objetivos de vida.*

DEDICO

“Crê em ti mesmo, age e verás os resultados. Quando te esforças, a vida também se esforça para te ajudar”

Chico Xavier

" O ser humano é um conquistador insuperável, fadado às estrelas que lhe estão ao alcance, na medida em que se empenha por alcançá-las"

Joana de Ângelis

“Age, enquanto é hoje. Ajuda, além do teu limite. Cresce pelo desprendimento de ti mesmo, e auxilia os que te retêm no dédalo das aflições. Não marches a sós, sem companhias com as quais sintonizas em razão do pretérito, tanto quanto dos objetivos que te fascinam a mente e o sentimento. Eleva o padrão das tuas aspirações, e trabalha o solo dos teus desejos, semeando a luz do amor, a fim de que o amor te responda com paz por cada lance de sacrifício e luta”

Joana de Ângelis

AGRADECIMENTOS

A Deus, onipotente, onisciente e onipresente e a Jesus Cristo, espírito de luz que guia todos os meus passos.

À Universidade Federal de Viçosa pela formação acadêmica.

Ao Conselho Nacional de desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

Ao professor Cosme Damião Cruz, pela orientação e amizade. Muito obrigada por tornar essa jornada mais leve e divertida.

Ao professor Leonardo Lopes Bhering pela coorientação, ensinamentos e sugestões.

À Embrapa Pecuária Sudeste em especial ao pesquisador Reinaldo de Paula Ferreira pela parceria.

Ao professor Moyses Nascimento, por participar da banca e contribuir para aprimorar o trabalho.

Aos colegas do Programa de Pós-Graduação em Genética e Melhoramento, pela troca de conhecimentos.

Aos amigos do Laboratório de Bioinformática por terem me recebido tão bem, compartilhando ensinamentos e bons momentos.

A todos os meus professores, desde o maternal. Todos foram igualmente responsáveis pelo meu desenvolvimento humano, intelectual e profissional.

Aos meus grandes amigos, por compartilharem comigo seus sonhos, torcida e carinho.

A Painho (meu anjo da guarda) pelo amor dedicado e imensa torcida, sei que me olha, guarda e ilumina.

À Mainha, pelo amor incondicional, pelas orações e torcida.

Às minhas irmãs Juliane e Gabriela, por acreditarem em mim e por tornarem a minha vida mais alegre.

À minha grande família! Avós, tios e primos que mesmo distantes fisicamente, mantêm fortes os elos de afeto e carinho.

Ao Arthur, parceiro da vida, pelo amor e paciência.

A todos que de alguma forma contribuíram para essa conquista, a minha imensa gratidão!

BIOGRAFIA

IARA GONÇALVES DOS SANTOS, filha de Marlene Gonçalves dos Santos e Jonilson Pereira dos Santos, nasceu em 04 de fevereiro de 1992, em Pavão, Minas Gerais.

Estudou na Escola Municipal Professora Davina Santos e concluiu o ensino médio no ano de 2009, na Escola Estadual Caio Nelson de Sena, Pavão, Minas Gerais.

Em março de 2011, ingressou no curso de Agronomia da Universidade Federal de Viçosa, colando grau em janeiro de 2016.

Em março de 2016, iniciou o curso de Mestrado em Genética e Melhoramento pela Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em julho de 2017.

SUMÁRIO

RESUMO	viii
ABSTRACT	x
1. INTRODUÇÃO GERAL	1
2. REVISÃO DE LITERATURA.....	3
2.1. Seleção baseada em métodos tradicionais.....	3
2.2. Redes Neurais Artificiais	7
2.2.1. Aplicações das Redes Neurais Artificiais no melhoramento	7
2.2.2. Considerações gerais sobre as Redes Neurais Artificiais	9
3. REFERÊNCIAS BIBLIOGRÁFICAS	16
CAPÍTULO 1	22
1. INTRODUÇÃO	24
2. MATERIAL E MÉTODOS	26
2.1. Material vegetal e caracteres mensurados	26
2.2. Métodos de seleção direta, indireta e simultânea para seleção de genótipos de alfafa com base em caracteres complexos.....	28
2.2.1. Ganhos por seleções direta e indireta	28
2.2.2. Seleção simultânea: Índice de Taí (1997).....	29
3. RESULTADOS E DISCUSSÃO	30
3.1. Desempenho dos genótipos e estimativas dos parâmetros genéticos e ambientais.....	30
3.2. Seleções diretas e indiretas entre médias.....	36
3.3. Seleção pelo método de Taí (1977)	41
4. CONCLUSÕES	44
5. REFERÊNCIAS BIBLIOGRÁFICAS	46
CAPÍTULO 2	49
1. INTRODUÇÃO	51

2. MATERIAL E MÉTODOS	53
2.1. Material vegetal e conjunto de dados.....	53
2.2. Modelagem da Rede Neural Artificial.....	59
3. RESULTADOS E CONCLUSÕES	60
5. REFERÊNCIAS BIBLIOGRÁFICAS	68
CAPÍTULO 3	71
1. INTRODUÇÃO	73
2. MATERIAL E MÉTODOS	74
2.1. Material vegetal e conjunto de dados.....	74
2.2. Modelagem da Rede Neural Artificial.....	78
3. RESULTADOS E DISCUSSÃO	79
4. CONCLUSÕES.....	84
5. REFERÊNCIAS BIBLIOGRÁFICAS	85
6. CONCLUSÕES GERAIS.....	89
APÊNDICE	90

RESUMO

SANTOS, Iara Gonçalves dos, M.Sc., Universidade Federal de Viçosa, julho de 2017. **Seleção de caracteres complexos em alfafa por meio de inteligência computacional.** Orientador: Cosme Damião Cruz. Coorientador: Leonardo Lopes Bhering.

O desenvolvimento de novos cultivares de alfafa está relacionado, além do potencial produtivo, com o valor nutritivo da forragem, ambos determinados por uma série de caracteres de controle genético complexo. Por isso, o melhoramento da cultura exige adoção de técnicas que maximizem a acurácia seletiva. Nesse contexto, a utilização das Redes Neurais Artificiais (RNAs) pode ser útil para otimizar a seleção simultânea de caracteres, que tem sido a estratégia mais utilizada no melhoramento da alfafa. Assim, o objetivo deste trabalho foi aplicar a abordagem de índices de seleção no contexto das redes neurais artificiais para auxiliar o melhoramento da cultura da alfafa, além de avaliar a eficiência das redes para realizar a seleção de caracteres complexos. Para isso, foram utilizados dados de 77 genótipos de alfafa avaliados em quatro cortes realizados ao longo das diferentes estações no ano, mensurados quanto a caracteres de produção e valor nutritivo. Primeiramente, foram estimados os ganhos por seleção direta, indireta e simultânea pelo índice de Taí para cada um dos grupos de caracteres. Para obtenção das redes para cada grupo, os genótipos foram subdivididos em classes segundo os escores do índice. A fim de garantir boa capacidade de generalização das redes, nas etapas de treinamento e validação foram utilizados dados ampliados de cada grupo de caracteres e, só então, as redes foram utilizadas para classificar os dados reais dos quatro cortes. Posteriormente, utilizou-se dados ampliados de três cortes para treinamento e validação e a rede, então estabelecida, foi empregada em dados reais, relativos a um dos cortes que não participou do processo de ampliação. De modo geral, as seleções diretas proporcionaram ganhos *per se* satisfatórios para os propósitos do melhoramento, mas sua utilização deve ser bastante criteriosa tendo em vista a existência de correlações desfavoráveis entre alguns caracteres estudados. Já as seleções indiretas apresentaram dificuldades para manter ganhos desejados em todos os caracteres no conjunto de genótipos. Os resultados da seleção simultânea utilizando o índice de Taí proporcionaram distribuição de ganhos mais equilibrada para os conjuntos de caracteres em todos os cortes. No contexto das redes neurais artificiais, as taxas de

erros das redes de produção e valor nutritivo foram baixas, evidenciando a grande concordância destas com o índice. Entretanto, a limitação do conjunto de treinamento e teste das RNAs para ampliação de três cortes, levou tanto a altas taxas de erro ou a níveis muito baixos, indicativo de perda da habilidade de generalização. Quando se utilizou as redes no novo corte, as taxas de erro foram altas em todos os cenários. O baixo desempenho obtido pode ser explicado pela discrepância observada entre o conjunto de treinamento e o conjunto de teste. Dessa forma, percebe-se a importância da qualidade do conjunto de dados para treinamento de uma rede eficiente, pois a utilização desta é diretamente relacionada com a disponibilidade de bons exemplos e em grande número. O uso das RNAs será mais difundido à medida que haja mudança no paradigma da experimentação em que as decisões passarão a ser tomadas em bancos de dados, construídos ao longo do tempo de um programa de melhoramento, que constituirão de base para o processamento de dados.

ABSTRACT

SANTOS, Iara Gonçalves dos, M.Sc., Universidade Federal de Viçosa, July, 2017. **Complex traits selection in alfalfa by means of computational intelligence.** Adviser: Cosme Damião Cruz. Co-Adviser: Leonardo Lopes Bhering.

Developing new alfalfa cultivars is related with productive potential besides to the nutritive value of the forage, which are determined by traits of complex genetic control. That is why alfalfa breeding requires adoption of techniques that maximize selective accuracy. In this context, using Artificial Neural Networks (ANNs) can be useful to optimize the simultaneous selection, which has been the strategy most used in alfalfa breeding. Thus, the aim of this work was to apply the selection index approach in the context of artificial neural networks to help in alfalfa breeding, as well as to evaluate the efficiency of the ANNs to perform the complex trait selection. The data from 77 alfalfa genotypes evaluated in four cuts were collected considering different seasons in the year. In each cut were measured production and nutritive value traits. Firstly, we estimated the gains by direct, indirect and simultaneous selection by the Taí index for each trait groups. To obtain the ANN for each group, the genotypes were subdivided into classes according to the index scores. In order to ensure good network generalization, in the training and validation stages we used amplified data from each group of characters, later we used the ANNs to classify the real data of the four cuts. Subsequently, we used amplified data from three cuts for training and validation, and the established ANNs were used in a new cut that did not participate in the amplification process. In general, the direct selections provided satisfactory individual gains for the breeding purposes, but their use must be very judicious since the existence of unfavorable correlations between some studied traits. Indirect selections, presented difficulties in maintaining desired gains in all traits in the genotype set. The results of simultaneous selection using the Taí index provided a more balanced distribution gains for the trait sets in all cuts. In the context of the artificial neural networks, the error rates of the production and nutritional value groups were low, evidencing the great agreement of these with the index. However, the limitation of the training set and test of ANNs for data of three cut amplification, led either to high error rates or to very low levels, indicating of loss of the generalization ability. When the networks were used in the new cut, error rates were high in all

scenarios. The low performance obtained can be explained by the discrepancy between the training set and the test one. In this way, the importance of the quality of data set for training an efficient network is perceived, since its use is directly related to the availability of good examples and in large numbers. The use of ANNs will be more widespread with a change in the experimentation paradigm in which decisions will be made into databases built over time from a breeding program that will form the basis for data processing.

1. INTRODUÇÃO GERAL

A alfafa cultivada constitui um grupo de oito subespécies próximas sendo o conjunto dessas subespécies denominado complexo *Medicago sativa*, que é pertencente ao gênero *Medicago*, subgênero *Medicago*, seção *Falcago* e à subseção *Falcatae* (Kopp, 2011). Atualmente, os cultivares de alfafa mais comuns em todo o mundo são originadas de *M. sativa* ssp. *sativa* e *M. sativa* spp. *falcata*, em menor número (CFIA, 2005). *Medicago sativa* ssp. *sativa* é autotetraploide ($2n = 4x = 32$), perene e caracterizada por flores púrpuras e vagens espiraladas. Por apresentar autoincompatibilidade e autoesterilidade, a polinização na alfafa é cruzada e realizada pela ação de abelhas e besouros (Ferreira et al., 2011; Kopp, 2011).

A alfafa é uma das principais leguminosas forrageiras nos países temperados (Annicchiarico et al., 2015; Ferreira et al., 2008), seja pela abrangência de área explorada, seja pelas suas características como alta produtividade e teor proteico, palatabilidade, digestibilidade, capacidade de fixação de nitrogênio atmosférico no solo, baixa sazonalidade, além de altos teores de vitaminas A, E e K e de minerais tais como cálcio, potássio magnésio e fósforo. Essa série de atributos garante à cultura a possibilidade de utilização na alimentação de rebanhos leiteiros especializados, com ótimos resultados na produção de leite (Comeron et al., 2015).

Nos sistemas intensivos de produção leiteira, o uso da alfafa como parte da alimentação animal pode reduzir os custos de produção, já que a quantidade requerida de concentrado, que representa grande parte do custo de produção do leite, é reduzida (Pompeu et al., 2003; Ferreira et al., 2008). Na Argentina, a alfafa é utilizada majoritariamente em pastejo direto e nos Estados Unidos, na forma de feno. No Brasil a forma mais utilizada até o momento tem sido o feno, no entanto a utilização na forma verde picada e sob pastejo é crescente uma vez que o feno tem preço elevado (Ferreira e Vilela, 2015).

Embora a alfafa apresente potencial de cultivo para regiões com diferentes condições edafoclimáticas (Viana et al., 2004), no Brasil dos 40 mil hectares cultivados, cerca de 90% estão concentrados na região sul, principalmente nos estados do Paraná e Rio Grande do Sul (Ferreira e Vilela, 2015). Os fatores que limitam a utilização da alfafa nos sistemas produtivos estão relacionados ao desconhecimento das técnicas de cultivo, à baixa fertilidade dos solos e a pouca

disponibilidade de cultivares adaptadas às condições tropicais, atualmente a única cultivar com boa adaptabilidade e estabilidade no Brasil é a Crioula (Botrel et al., 2001; Ferreira e Vilela, 2015).

O desenvolvimento de cultivares de alfafa está diretamente relacionado com o potencial produtivo e o valor nutritivo, que são determinados por uma série de caracteres tais como produção de matéria seca, proteína bruta, digestibilidade “*in vitro*” da matéria seca, fibra em detergente neutro, fibra em detergente ácido e teor de lignina. Uma vez que o controle genético dos principais caracteres da alfafa é complexo e sujeito a forte ação do ambiente, há necessidade de utilização de métodos específicos de melhoramento para a cultura (Basigalup e Odorizzi, 2011; Ferreira et al., 2008).

Os principais objetivos dos programas de melhoramento da alfafa no mundo todo são o aumento da produtividade, melhoria do valor nutritivo e tolerância aos estresses bióticos e abióticos. A seleção em alfafa é baseada principalmente no método de seleção recorrente fenotípica, por isso a seleção pode ser dificultada, já que os caracteres de interesse geralmente são muito afetados pelo ambiente e conseqüentemente apresentam baixa herdabilidade (Li and Brummer, 2012), além disso, o melhoramento é feito visando melhorar uma série de caracteres importantes, requerendo adoção de técnicas que maximizem a acurácia seletiva.

Atualmente, as estratégias de seleção para mais de um caráter em genótipos de alfafa segundo Basigalup e Odorizzi (2011) têm sido: i) a seleção em *tandem* em que se melhora independentemente cada caráter por vez até se atingir os objetivos. ii) níveis independentes de seleção, onde são selecionados a cada ciclo, genótipos que satisfaçam os níveis determinados de cada caráter. iii) e a seleção por índices, onde as unidades de seleção são avaliadas conjuntamente e a cada ciclo em relação a uma série de caracteres de interesse. De todas as estratégias, a seleção por índices tem se mostrado mais eficiente (Hazel e Lush, 1942).

A acurácia da seleção simultânea é determinada pela qualidade do conjunto de dados experimentais. Para que sejam eficientes, determinados índices de seleção necessitam além do conhecimento do pesquisador, de matrizes de variância e covariância bem estimadas, fato que só é possível se os experimentos forem muito bem conduzidos. Por isso, apesar da maior eficiência se comparada às demais

estratégias de seleção, a utilização de índices pode ser dificultada pelo grande volume de dados gerados nos vários experimentos nos programas de melhoramento da alfafa.

Dessa forma, a abordagem da inteligência computacional, que se baseia no aprendizado a partir de exemplos, poderia ser utilizada para facilitar a tomada de decisão do melhorista. A partir de um banco de dados que fosse constituído de experimentos bem conduzidos e que caracterizasse acessos em diferentes condições ambientais, genótipos de alfafa poderiam ser selecionados com maior eficiência pelas Redes Neurais Artificiais (RNAs). Uma RNA treinada a partir de informações obtidas ao longo do tempo poderia ser capaz de identificar padrões e ponderar os caracteres tal qual é feito em um índice de seleção com a vantagem de trabalhar muito bem com informações incompletas e com ruído, tornando dispensável todo o rigor necessário para a condução de todos os experimentos e das análises estatísticas de novos cortes realizados.

Uma vez que a teoria dos índices de seleção é pouco explorada no ramo da inteligência computacional e que as Redes Neurais Artificiais (RNAs) têm se mostrado bastante eficazes para solucionar problemas nas diversas áreas da ciência, o desenvolvimento de procedimentos que otimizem o processo de melhoramento de alfafa baseado em RNAs pode garantir maior eficiência nos programas de melhoramento genético da cultura.

Dessa forma, esse trabalho objetiva aplicar técnicas biométricas e de inteligência computacional para auxiliar o melhoramento da cultura da alfafa nas fases intermediárias dos programas de melhoramento. Pretende-se comprovar a eficiência e potencialidade do uso das RNAs como ferramenta de seleção simultânea, propor redes para caracteres complexos de produção e valor nutritivo dos genótipos de alfafa da Embrapa Pecuária Sudeste, além de avaliar a importância do tamanho e diversidade dos dados de treinamento da rede para generalização de informações.

2. REVISÃO DE LITERATURA

2.1. Seleção baseada em métodos tradicionais

Como a maioria dos caracteres de importância da cultura da alfafa é de natureza poligênica, realizar a seleção de forma acurada não é uma tarefa simples (Cruz, 2006).

Por isso, é extremamente interessante que o melhorista determine a variabilidade disponível em suas populações, em relação aos principais caracteres, e escolha os métodos mais adequados para seleção de plantas, reduzindo o tempo gasto para condução dos ciclos de melhoramento (Costa et al., 2000). No caso da alfafa, que é altamente sensível à endogamia, a base genética deve ser ampla. Os métodos de melhoramento utilizados são tanto o melhoramento interpopulacional quanto o intrapopulacional (Ferreira et al., 2008).

O melhoramento interpopulacional é realizado a partir de populações de polinização aberta, onde o intercâmbio de pólen é realizado tanto manualmente quanto por insetos e visa a obtenção de variedades sintéticas, por exemplo. Já o melhoramento intrapopulacional utiliza diversos métodos de seleção para aumentar a frequência de alelos favoráveis dentro da população (Ferreira et al., 2008; Basigalup e Odorizzi, 2011; Saraiva, 2011).

Durante todo o processo de desenvolvimento de cultivares, é fundamental que se faça seleção de forma acurada. Dentre as estratégias utilizadas nas etapas iniciais de programas de melhoramento encontra-se a seleção direta, praticada quando se objetiva obter ganhos em um único caráter (Cruz et al, 2014). A seleção direta proporciona ganhos superiores nos caracteres sob seleção, no entanto, se o caráter sofrer grande influência do meio a precisão da seleção será baixa. Além disso, a existência das correlações genéticas desfavoráveis pode conduzir a respostas indesejáveis para os caracteres que não são alvos da seleção, tornando o uso da seleção direta limitado.

Segundo Li e Brummer (2012), o fato de o ganho para produtividade em alfafa nos Estados Unidos ser baixo se justifica uma vez que a seleção para produção tem que considerar entre outros fatores, o valor nutritivo da forragem. O foco do melhoramento não é o rendimento *per se*, mas sim a avaliação de uma série de componentes essenciais para o estabelecimento da cultura no campo e suplementação eficiente dos animais. Vasconcelos et al. (2010) encontraram maiores ganhos para alguns caracteres de alfafa quando se aplicou a seleção direta, todavia essa seleção levou ao aumento da suscetibilidade dos cultivares em relação a incidência de doenças. Esses autores destacaram que nenhuma das seleções diretas para alfafa propiciaram ganhos satisfatórios para todos os caracteres avaliados. Este fato também tem sido relatado em outras culturas. Assim, Freitas et al. (2012),

estudando a formação de população base para seleção recorrente em maracujazeiro-amarelo, concluíram que os indivíduos selecionados com foco apenas em produtividade apresentaram estimativas de ganho indesejáveis para uma série de caracteres sob avaliação.

A avaliação da resposta indireta ou correlacionada pode ser uma estratégia útil para realizar seleção. A determinação da magnitude da resposta correlacionada é interessante principalmente para seleção de caracteres que apresentam baixa herdabilidade ou que são de difícil mensuração (Cruz et al., 2014), nesses casos, a seleção indireta com base em caracteres secundários de baixa influência ambiental, fácil mensuração e correlacionados com a característica de interesse é uma alternativa bastante interessante para maximizar a acurácia preditiva do material sob avaliação.

O sucesso da seleção indireta já é relatado na área agrária. No melhoramento da cana-de-açúcar por exemplo, uma das estratégias iniciais nas fases de seleção é a aplicação da seleção indireta, que facilita a obtenção de ganhos para teor de sólidos solúveis totais. Como este é um caráter de difícil mensuração, a partir de caracteres de fácil mensuração e que apresentem alta correlação com ele, é possível obter ganhos para essa característica (Pedrozo et al., 2009). Todavia, esse tipo de seleção precisa ser criteriosamente aplicado, uma vez que, correlações desfavoráveis podem levar a sérios problemas na população melhorada, podendo esta apresentar desempenho para características secundárias que inviabilizaria sua utilização (Cruz et al., 2014).

Uma outra estratégia é a realização de seleção entre e dentro de populações baseada nos valores fenotípicos. Essa prática engloba uma maior fração da variância aditiva, mas apresenta a desvantagem de basear-se em valores fenotípicos individuais. Assim, materiais de desempenho medianos, porém interessantes, podem ser excluídos por não se destacarem nas seleções iniciais (Cedillo et al., 2008; Costa et al., 2000).

Falconer e Mackay (1996) afirmam que a seleção combinada (baseada na obtenção de informações fenotípicas de parentes dos indivíduos sob seleção) permite uma inferência mais acurada sobre o valor genético destes. O conceito de seleção combinada pode ser entendido como a identificação de genótipos superiores a partir da informação do próprio indivíduo e de seus parentes a partir dos valores fenotípicos

dos indivíduos, medidos como o desvio da média da população (Cruz et al., 2014). Na alfafa os tipos de progênies avaliadas incluem meios irmãos (polinização aberta e policruzamentos), irmãos completos (*topcrosses* e cruzamentos dialélicos) ou até mesmo progênies endogâmicas (S_1) em menor número devido à fatores intrínsecos à cultura (Ferreira et al., 2008).

O sucesso no emprego da seleção entre e dentro ou seleção combinada tem sido relatado em várias espécies. Neves et al. (2011), a fim de predizer os ganhos genéticos em populações de maracujazeiro utilizando o Delineamento I de Comstock e Robinson, encontraram maiores ganhos de seleção em todas as características avaliadas quando se aplicou a seleção combinada. Matias et al. (2011), visando efetuar seleção para baixa necessidade de frio hibernar em genótipos de pessegueiro, também concluíram que a seleção combinada proporcionou estimativas maiores de ganhos quando comparada à seleção entre e dentro.

Todavia, a seleção baseada em uma ou poucas características quase sempre leva a resultados pouco satisfatórios. Por isso, a seleção simultânea de caracteres é uma prática que maximiza as chances de êxito no processo de melhoramento (Cruz et al., 2014). As vantagens da seleção simultânea já foram relatadas em muitas culturas vegetais. Destacando-se principalmente pelo ganho equilibrado para todos os caracteres e a possibilidade de obtenção de ganho até mesmo para caracteres desfavoravelmente correlacionados, como é o caso da produtividade e capacidade de expansão em milho-pipoca (Granate et al., 2002). A utilização de índices de seleção na alfafa é eficiente desde que as amostras avaliadas sejam suficientemente grandes para minimizar erros na estimação dos ganhos (Basigalup e Odorizzi, 2011).

A teoria dos índices de seleção foi proposta primeiramente por Smith (1936) e Hazel (1943). O Índice Clássico de Seleção, derivado da teoria proposta por Smith e Hazel, consiste em uma função linear dos valores fenotípicos dos caracteres sob os quais se deseja praticar seleção, cujos coeficientes de ponderação são estimados visando maximizar a correlação existente entre o próprio índice e o agregado genotípico, esse índice é dito não-restrito, uma vez que podem ocorrer respostas indesejáveis em alguns caracteres dentro do agregado genotípico (Cruz et al., 2014).

Outro critério de seleção é o índice proposto por Pesek & Baker (1969), no qual dispensa-se o cálculo dos pesos econômicos. Os coeficientes são definidos a partir dos ganhos desejados pelo melhorista para cada caráter, no entanto, os ganhos

desejados estão limitados pela estrutura genética da população em estudo (Cruz et al., 2014; Martins et al. 2003). Tai (1977) propôs uma generalização para o Índice de Pesek & Baker onde a inclusão de caracteres secundários no índice auxilia no melhoramento dos caracteres principais. O índice de seleção proposto por Taí (1977), permite que caracteres secundários, para os quais não se deseja ganho máximo, sejam incluídos no índice para auxiliar no ganho dos caracteres principais. No entanto, os valores econômicos dos caracteres secundários devem ser nulos no agregado genotípico e o vetor de ganhos desejados deve ser conhecido (Cruz et al., 2014).

Quando se objetiva fixar ou manter o ganho em um nível pré-determinado em um conjunto de caracteres e obter ganho em outro, devem-se adotar índices de seleção restritos. Kempthorne e Nordskog (1959) propuseram que a correlação entre a equação linear do índice e o agregado genotípico fosse máxima, mas que a covariância entre o índice e o valor genotípico dos caracteres que o melhorista não quer alterar fosse nula.

A partir do Índice de Kempthorne e Nordskog, outros autores desenvolveram índices com diferentes restrições como por exemplo Tallis (1962) e James (1968). Cunningham et al. (1970) propuseram uma nova forma de cálculo do índice restrito, simplificando o conjunto de equações do índice, a manipulação dos pesos econômicos e possibilitando o cálculo do custo genético das restrições impostas (Cruz et al., 2014).

Ademais, índices fenotípicos mais simplificados e de extensa utilização pelos melhoristas foram desenvolvidos ao longo do tempo. Williams (1962) propôs o índice-base, que dispensa as estimativas de variâncias e covariâncias fenotípicas e genotípicas e estabelece uma combinação linear entre pesos econômicos e valores fenotípicos médios. Mulamba e Mock (1978) propuseram o índice baseado na soma de postos, que dispensa além das estimativas de variância e covariância, os pesos econômicos. Apesar de bastante simplificados, estes índices demonstram alta eficiência para seleção de genótipos superiores nas mais diversas culturas (Ferreira et al., 2008).

2.2. Redes Neurais Artificiais

2.2.1. Aplicações das Redes Neurais Artificiais no melhoramento

A utilização de técnicas convencionais de seleção pode tornar todo o processo de melhoramento moroso, principalmente para espécies como a alfafa. Por isso o uso de técnicas de inteligência computacional, como por exemplo, as Redes Neurais Artificiais (RNAs), pode ser útil para maximizar a acurácia preditiva e otimizar a seleção simultânea de caracteres. Isso porque as RNAs se baseiam no aprendizado a partir de exemplos para generalizar informações. Assim, se a RNA fosse treinada com dados vindos de experimentos bem conduzidos e que caracterizassem acessos em diferentes condições ambientais, genótipos de alfafa poderiam ser selecionados com maior eficiência pelas RNAs nos experimentos realizados posteriormente, tal qual é feito em um índice de seleção com a vantagem de trabalhar muito bem com informações incompletas e com ruído, tornando dispensável todo o rigor necessário para a condução de todos os experimentos e das análises estatísticas de novos cortes realizados.

Devido a sua estrutura não linear as RNAs conseguem alta eficiência trabalhando com caracteres complexos (Galvão et al., 1999), sendo capazes de capturar relações entre variáveis que não são captadas pelos modelos estocásticos, podendo ainda generalizar as informações para novos casos (Mackay, 1994). Além disso, não requerem informações muito detalhadas a respeito dos processos físicos dos sistemas modelados (Sudheer et al., 2003).

Procedimentos de Redes Neurais Artificiais têm sido amplamente empregados em diversos ramos da ciência. A utilização dessa abordagem no melhoramento vegetal tem sido crescente e inclui, por exemplo, predição de produção de grãos de arroz (Ji et al., 2007), avaliação de adaptabilidade e estabilidade em genótipos de alfafa (Nascimento et al., 2013), seleção em cana-de-açúcar (Zhou et al., 2011) e previsão de características complexas em vacas Jersey e trigo (Gianola et al., 2011).

Sant'anna (2014), avaliando a eficiência das RNAs em estudos classificatórios constatou a eficiência destas para classificar populações derivadas de retrocruzamento com alta similaridade com o genitor recorrente, sendo o desempenho da rede muito superior ao desempenho de técnicas multivariadas como as análises discriminantes de Anderson e Fisher. Silva (2014), avaliando a eficiência das RNAs na predição de valores genéticos, concluiu que as redes demonstraram eficiência na predição dos valores e ganhos genéticos em ensaios simulados, no entanto a autora destaca a importância de estudos com populações reais para comprovação da

superioridade encontrada. Resultado semelhante já havia sido relatado por Peixoto (2013), avaliando a eficiência das RNAs para predição do valor genético em experimentos balanceados para características quantitativas.

Barbosa et al. (2011) realizaram estudo de diversidade genética de acessos de mamão papaia utilizando oito características quantitativas e compararam os resultados obtidos na RNA com aqueles obtidos na análise discriminante de Anderson. Os autores concluíram que as RNAs classificam de forma eficiente os acessos em estudos de diversidade genética.

Teodoro et al. (2015) estudando a utilização das RNAs para identificação de genótipos de feijão-caupi semiprostrado com alta adaptabilidade e estabilidade fenotípica, encontraram alta concordância entre as análises de adaptabilidade e estabilidade de Eberhart e Russell com as RNAs. Os autores concluíram que as RNAs são uma alternativa eficaz nos estudos de adaptabilidade e estabilidade fenotípica em programas de melhoramento genético.

Brasileiro et al. (2015) avaliou o uso das RNAs na seleção dentro de famílias de cana de açúcar e ajustou um modelo que não apresentou erros de classificação quando comparado ao BLUP individual simulado. Os autores confirmaram o grande potencial das RNAs para o processo de seleção e identificação de genótipos nos programas de melhoramento de cana-de-açúcar.

2.2.2. Considerações gerais sobre as Redes Neurais Artificiais

O neurônio biológico

Redes neurais artificiais são técnicas de aprendizado baseadas em modelos de natureza biológica inspirados no funcionamento do cérebro humano, que utiliza neurônios para transmissão de informações (Norvig e Russel, 2008). O cérebro humano apresenta capacidade de processar informação complexas e não-lineares de forma muito superior à de qualquer máquina e, devido à plasticidade do sistema nervoso, a rede de neurônios biológicos pode se adaptar às novas situações a que é exposta (Haykin, 2001). Segundo Silva (2014), todas as ações do corpo humano são governadas pelo sistema nervoso, que tem os neurônios como estrutura básica, estes são células especializadas em processar e transmitir informações. Na Figura 1 está

representado a estrutura de um neurônio biológico, estruturado em soma (corpo celular), dendritos e axônio.

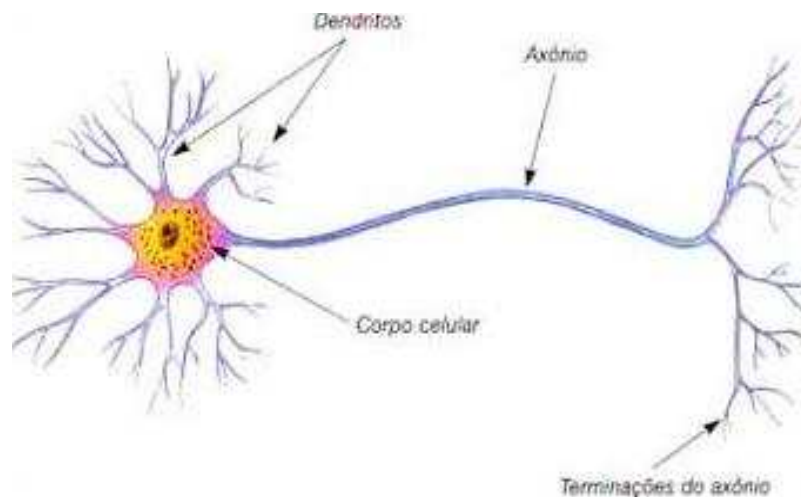


Figura 1: Modelo de neurônio biológico. Fonte: (<http://www.sogab.com.br/anatomia/neuronio.jpg>)

Os dendritos são responsáveis por captar as informações iniciais e conduzi-las para o corpo celular. No corpo celular ocorre o processamento dos estímulos recebidos. A combinação dos sinais no corpo celular pode formar sinais inibitórios ou excitantes, governados por um potencial de ação (Haykin, 2001; Kovacs, 2006). Uma vez que o sinal atinge determinado limiar, ele é transmitido através do axônio de uma célula para os dendritos de outra. O ponto de contato do axônio de um neurônio e o dendrito de um próximo, é chamado de sinapse. (Braga et al., 2007).

O neurônio artificial: Modelo de McCulloch e Pitts (1943)

O primeiro modelo de neurônio artificial data de 1943 e foi proposto por Warren McCulloch e Walter Pitts (McCulloch & Pitts, 1943). O neurônio artificial de McCulloch e Pitts (Figura 2) representa uma simplificação do neurônio biológico (Braga et al., 2007). São constituídos basicamente pelo conjunto de entradas associadas a pesos sinápticos, por um combinador linear que sumariza em uma única medida os valores de entradas e seus respectivos pesos e por uma função de ativação que controla a ação do neurônio através de um potencial de ativação constituído pelo combinador linear (Carneiro, 2015).

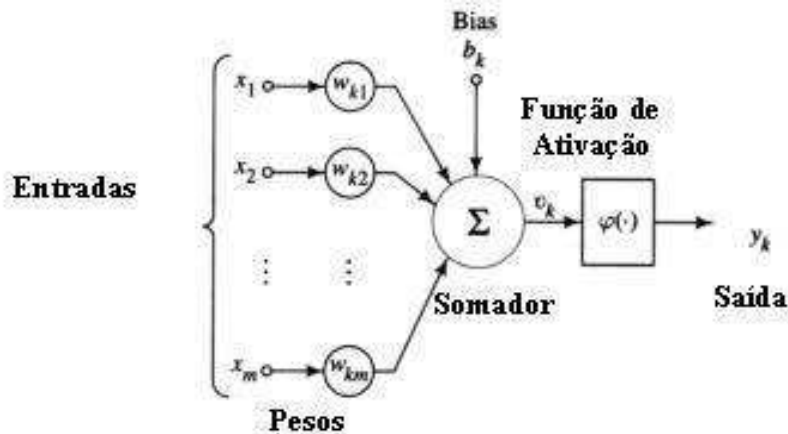


Figura 2: Modelo de neurônio artificial proposto por McCulloch & Pitts (1943). Onde x_1, x_2, \dots, x_m são as entradas; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos; b_k é o viés; v_k é a combinação linear das entradas; $\varphi(\cdot)$ é a função de ativação e y_k é a saída do neurônio. (Fonte: Haykin, 2001)

O neurônio biológico transmite a informação quando a soma dos impulsos que lhe são dados atinge certo limiar. No modelo de neurônio artificial, toda a transmissão da informação é controlada pela função de ativação, que permite ou não a ativação da saída. Dessa forma, a sinapse é representada pelas m entradas ponderadas por pesos (w_k) positivos ou negativos, um limiar conhecido como bias que se mantém em associação com uma função de ativação levando a uma saída y (Silva, 2014).

Devido às particularidades do modelo, existem algumas limitações que devem ser consideradas, como por exemplo, as redes de McCulloch e Pitts só conseguem solucionar problemas onde as classes são separadas por uma reta, ou seja, são linearmente separáveis, pesos negativos são melhores para representar sinais inibitórios e os pesos são não ajustáveis (Braga et al., 2007).

Funções de ativação

O modelo de McCulloch e Pitts foi fundamental para o desenvolvimento de vários outros modelos mais dinâmicos, que não necessariamente apresentam saída 0 ou 1 e que utilizam diferentes funções de ativação (Braga et al., 2007). As funções de ativação são responsáveis por fornecer a saída de um neurônio baseadas nas informações de entrada e pesos atribuídos e a escolha da função pode exercer forte influência na convergência dos algoritmos de aprendizado (Gomes, 2010). A escolha

de uma função deve ser feita de acordo com o problema a ser solucionado (Sant'anna, 2014), além disso, deve ser considerado o gasto computacional e as características do próprio conjunto de dados (Gomes, 2010). As funções de ativação comumente utilizadas para solucionar problemas complexos em redes multicamadas, são: função sigmoidal e função tangente hiperbólica. Já a função limiar é uma das mais comuns para os neurônios da camada de saída de rede utilizada por exemplo, na saída de redes neurais de regressão.

i. Função Sigmoidal

Segundo Haykin (2001) a função Sigmoidal é a mais utilizada na construção das RNAs, podendo assumir valores entre 0 e 1. Além disso, apresenta gráfico com forma de "S", é monótona crescente com propriedades assintóticas e de suavidade como exemplificado na Figura 3:

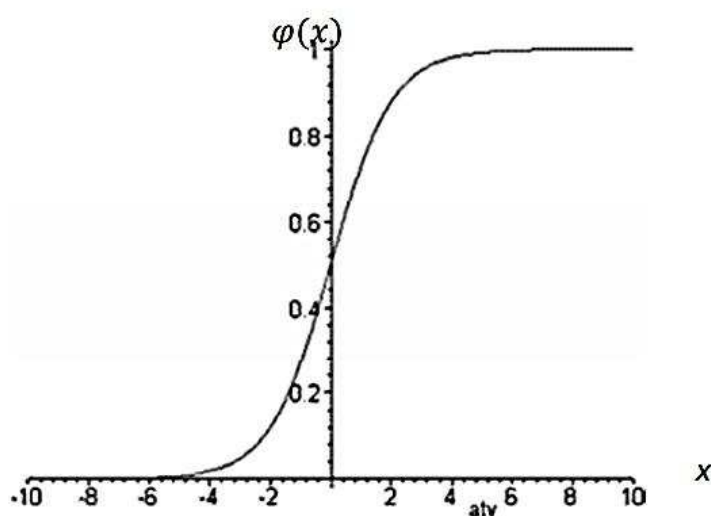


Figura 3: Gráfico da função sigmoidal

ii. Função Tangente Hiperbólica

Há certas situações onde é necessário que a função de ativação se estenda de -1 a 1, assumindo forma antissimétrica em relação à origem (Haykin, 2001), conforme mostrado na Figura 4:

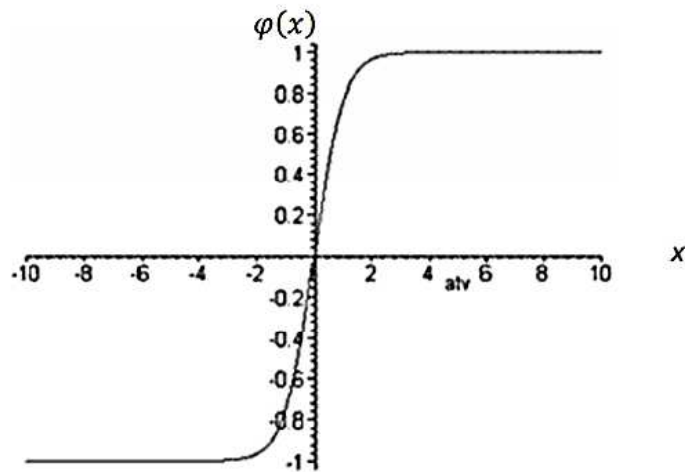


Figura 4: Gráfico da função tangente hiperbólica

iii. Função Limiar

Os neurônios que contêm essa função terão saída 1, quando o valor de ativação for negativo ou 0, quando o valor de ativação for positivo (Figura 5). É a função utilizada no modelo de neurônio de McCulloch & Pitts.

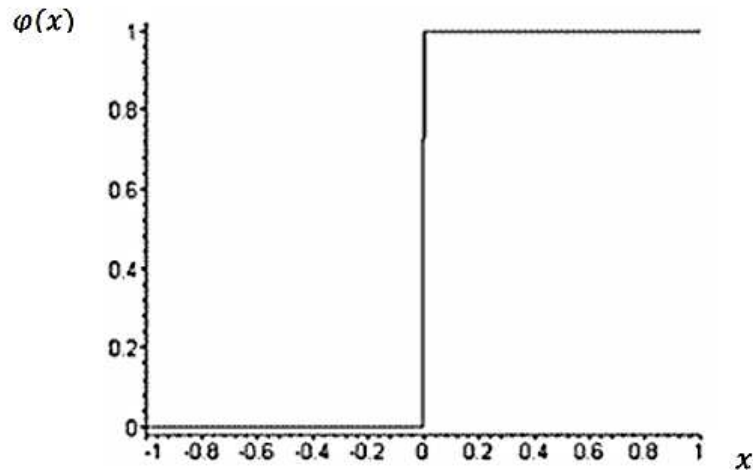


Figura 5: Gráfico da função Limiar

Arquiteturas e topologias das Redes Neurais Artificiais

O conceito de arquitetura de rede pode ser entendido como o arranjo ou disposição dos neurônios em uma rede (Haykin, 2001), alguns exemplos de

arquitetura são as redes feedforward de camada simples (Rosenblatt, 1958; Widrow e Hoff, 1960) e feedforward de camadas múltiplas (Rumelhart et al., 1986). Já o conceito de topologia de rede pode ser definido como as diferentes formas que uma determinada arquitetura de rede pode assumir (Carneiro, 2015).

As RNAs podem ser divididas em três partes principais: a camada de entrada, que recebe os dados iniciais, as camadas intermediárias, que realizam a maior parte do processamento das informações, e as camadas de saída responsáveis pelo processamento das informações das camadas anteriores, resultando na resposta final da rede. Cada rede neural apresenta uma estrutura bem definida para resolver os mais diversos problemas (Carneiro, 2015).

De maneira geral, as RNAs diferem entre si pelas características que apresentam, e suas estruturas são diretamente relacionadas ao algoritmo de aprendizado utilizado no treinamento (Silva, 2014). As redes podem ser divididas em dois grupos, considerando o número de camadas: i) Redes de única camada de neurônios e ii) Redes que apresentam várias camadas intermediárias ou ocultas. As redes de única camada foram primeiramente adotadas, mas apresentam uma série de limitações quanto à resolução de problemas não linearmente separáveis (Haykin, 2001).

Posteriormente, frente à necessidade de maior eficiência para resolução de problemas complexos, foram desenvolvidas redes com uma ou várias camadas intermediárias, como por exemplo a Perceptron Multicamadas, que utiliza o algoritmo de treinamento chamado de “algoritmo de retropropagação de erro” (Haykin, 2001). As redes Perceptron Multicamadas são amplamente utilizadas e apresentam ótima eficiência para solucionar problemas complexos, graças à inclusão das camadas ocultas no modelo.

Os demais parâmetros de uma rede são definidos empiricamente e ajustados conforme o desempenho da rede durante a etapa de treinamento. O número de neurônios por camada oculta determina a precisão da solução do problema, mas ainda não existe uma indicação da distribuição destes ao longo das camadas. No geral, o que se pode afirmar é que um número excessivo de neurônios leva à memorização do conjunto de dados de treinamento e um número baixo aumenta o tempo de aprendizagem da rede, que pode não alcançar resultado satisfatório (Sant’anna, 2014).

Modelo Perceptron Multicamadas (Multilayer Perceptron)

No ano de 1958 Frank Rosenblatt propôs um modelo de rede neural composto por neurônios de McCulloch e Pitts e uma regra de aprendizado (Braga et al., 2007). Assim como o neurônio de McCulloch e Pitts, esse modelo, denominado Perceptron, era capaz de solucionar somente problemas linearmente separáveis. No entanto, em 1986 tornou-se possível a utilização das redes Perceptron devido a aplicação do algoritmo de treinamento *back propagation* nas redes de múltiplas camadas (Braga et al., 2007; Haykin, 2001). Esse novo modelo foi denominado Perceptron Multicamadas (MLP) e é mostrado na Figura 6:

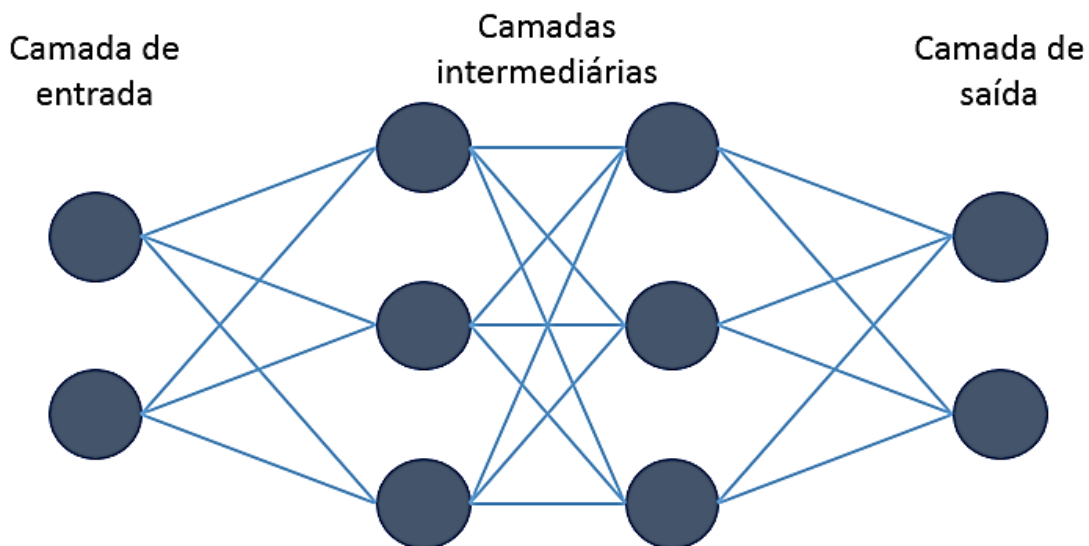


Figura 6: Representação do modelo de rede Perceptron Multicamadas

A rede MLP é caracterizada pela presença de uma ou mais camadas ocultas, alto grau de conectividade e pela presença de diferentes funções de ativação não-lineares (Peixoto, 2013). O conjunto dessas características garantem às redes MLP alto poder computacional (Silva, 2014).

3. REFERÊNCIAS BIBLIOGRÁFICAS

- Annicchiarico, P.; Barrett, B.; Brummer, E. C.; Julier, B.; Marshall, A. H. (2015) Achievements and challenges in improving temperate perennial forage legumes. *Critical Reviews in Plant Sciences*, 34:327-380.
- Barbosa, C. D.; Viana, A. P.; Quintal, S. S. R.; Pereira, M. G. (2011) Artificial neural network analysis of genetic diversity in *Cariacica papaya* L. *Crop Breeding and Applied Biotechnology*. 11:224-231.
- Basigalup, D.H.; Odorizzi, A. S. Melhoria genética da alfafa. In: Ferreira, R. P.; Basigalup, D. H.; Gioco, J. O. (Ed.). *Melhoramento Genético da Alfafa*. São Carlos: Embrapa Pecuária Sudeste, 2011. p. 225-260.
- Botrel, M. A.; Ferreira, R. P.; Alvim, M. J.; Xavier, D. F. (2001) Cultivares de alfafa em área de influência da Mata Atlântica no Estado de Minas Gerais. *Pesq. agropec. bras.*, 36:1437-1442.
- Braga, A. P.; Ferreira, A. C. P. L.; Ludermir, T. B. (2007) *Redes Neurais Artificiais: teoria e aplicações*. Rio de Janeiro, LTC Editora, 262p.
- Brasileiro, B. P. et al. (2015) Selection in sugarcane families with artificial neural networks. *Crop Breed. Appl. Biotechnol.*, 5:72-78.
- Canadian Food Inspection Agency (2005) The Biology of *Medicago sativa* L. (Alfafa). Disponível em: <http://www.inspection.gc.ca/plants/plants-with-novel-traits/applicants/directive-94-08/biology-documents/medicago-sativa-l-/eng/1330981151254/1330981232360#a1>. Acesso em: 26/05/2017.
- Carneiro, V. Q. (2015) Rede neural e lógica Fuzzy aplicadas no melhoramento do feijoeiro. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 108p, 2015.
- Cedillo, D. S. O.; Ferreira, F. M.; Barros, W. S.; Cruz, C. D.; et al. (2008) Selection among and within and combined selection in oil palm families derived from Dura x Dura. *Ciência Rural*, 38:65-71.

- Comeron, E. A.; Ferreira, R. P.; Vilela, D.; Kuwahara, F. A.; Tupy, O. Utilização da alfafa em pastejos para alimentação de vacas leiteiras. In: Ferreira RP, Vilela D, Comeron EA, Bernardi ACC, Karam D (Ed.). *Cultivo e utilização da alfafa em pastejo para alimentação de vacas leiteiras*. Brasília: Embrapa, 2015. p.13-16.
- Costa, R. B.; Resende, M. D. V.; Araujo, A. J.; Gonçalves, P. S.; et al. (2000) Seleção combinada univariada e multivariada aplicada ao melhoramento genético da seringueira. *Pesq. Agropec. bras.*, 35:381-388.
- Cruz, C. D. (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum.*, 35:271-276.
- Cruz, C. D. (2006) Genes: biometria. Viçosa, Editora UFV. 302p.
- Cruz, C. D.; Carneiro, P. C. S.; Regazzi, A. J. (2014) Modelos biométricos aplicados ao melhoramento genético. Viçosa, Editora UFV. 668p.
- Cunningham, E. P.; Moen, R. A.; Gjedrem, T. (1970) Restriction of selection indexes. *Biometrics*, 26:67-74.
- Falconer, D. S.; Mackay, T. F. (1996) Introduction to quantitative genetics. [S.l. : s.n.] 464p.
- Ferreira, R. P.; Vilela, D. Potencial de utilização da alfafa. In: Ferreira RP, Vilela D, Comeron EA, Bernardi ACC, Karam D (Ed.). *Cultivo e utilização da alfafa em pastejo para alimentação de vacas leiteiras*. Brasília: Embrapa, 2015. p.13-16.
- Ferreira, R. P.; Basigalup, D.H.; Vasconcelos, E. S.; Cruz, C. D.; Pereira, A. V. Genética quantitativa e métodos de melhoramento em alfafa. In: Ferreira, R. P.; Rassini, J. B.; Rodrigues, A. A.; Freitas, A. R.; Camargo, A. C.; Mendonça, F. C. (Ed.). *Cultivo e Utilização da Alfafa nos Trópicos*. São Carlos: Embrapa Pecuária Sudeste, 2008. p. 171-205.
- Freitas, J. P. X.; Oliveira, E. J.; Jesus, O. N.; Neto, A. J. C. et al. (2012) Formação de população base para a seleção recorrente em maracujazeiro-amarelo com uso de índices de seleção. *Pesq. Agropec. bras.*, 47:393-401.

- Galvão, C. O.; Valença, M. J. S.; Vieira, V. P. P. B.; Diniz, L. S. et al. (1999) Sistemas inteligentes: Aplicações a recursos hídricos e ciências ambientais. UFRGS: ABRH.
- Gomes, G. S. S. (2010) Novas funções de ativação em redes neurais artificiais multilayer perceptron (Doutorado em Ciência da Computação) - Universidade Federal de Pernambuco, 137p, 2010.
- Granate, M. J.; Cruz, C. D.; Pacheco, C. A. P. (2002) Predição de ganho genético com diferentes índices de seleção no milho pipoca CMS-43. *Pesq. Agropec. bras.*, 37:1001-1008.
- Gianola, D.; Okut, H.; Kent A. W.; Rosa, J. M. R. (2011) Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genetics*, 45:34.
- Hazel, L. N. (1943) The genetics basis for constructing selection index. *Genetics*, 28:476-490.
- Hazel, L. N. and Lush, J. L. (1942) The efficiency of three methods of selection. *Journal Heredity*, Oxford, 33:393-399.
- Haykin, S. (2008) *Neural Networks and Learning Machines* (3rd ed.). Hamilton: Pearson – Prentice Hall, 906p.
- James, J. W. (1968) Index selection with restrictions. *Biometrics*, 24:1015-1018.
- Ji, B.; Sun, Y.; Yang, S. and Wan, J. (2007) Artificial neural networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science*, 145:249-261.
- Kempthorne, O.; Nordskog, A. W. (1959) Restricted selection índices. *Biometrics*, 15:0-19.
- Kopp, M. M. Origem, evolução e domesticação da alfafa. In: Ferreira, R. P.; Basigalup, D. H.; Gioco, J. O. (Ed.). *Melhoramento Genético da Alfafa*. São Carlos: Embrapa Pecuária Sudeste, 2011. p. 225-260.

- Kovács, Z. L. (2006) Redes Neurais Artificiais: fundamentos e aplicações. 4ª ed. São Paulo, Editora Livraria da Física, 174p.
- Li, X.; Brummer, E. C. (2012) Applied Genetics and Genomics in Alfalfa Breeding. *Agronomy*, 2:40-61.
- Mackay, D. J. C. (1994) Bayesian non-linear modelling for the prediction competition. In: ASHRAE Transactions, ASHRAE, Atlanta Georgia. 100:1053-1062.
- Martins, I. S.; Cruz, C. D.; Regazzi, A. J.; Pires, I. E. (2003) Eficiência da seleção univariada direta e indireta e de índices de seleção em *Eucalyptus grandis*. *R Árvore*, 27:327-333.
- Matias, R. G. P.; Bruckner, C. H.; Santos, C. E. M.; Dias, D. C. F. S. D.; et al. (2011) Seleção de genótipos de pessegueiro com base na necessidade de frio para germinação das sementes. *Ver. Bras. Frutic.*, 33:571-576.
- McCulloch, W. S.; Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol.* 5:115-133.
- Mulamba, N. N.; Mock, J. J. (1978). Improvement of yield potential of the Eto Blanco maize (*Zea mays* L) population by breeding for plant traits. *Egypt. J. Gen. Cytol.*, 7:40-51.
- Nascimento, M.; Peternelli, L. A.; Crus, C. D.; Nascimento, A. C. C.; Ferreira, R. P.; Bhering, L. L. and Sangado, C. C. (2013) Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. *Crop Breeding and Applied Biotechnology*, 13:152-156.
- Neves, L. G.; Bruckner, C. H.; Cruz, C. D.; Duarte, L. P.; et al. (2011) Predição de ganhos genéticos utilizando o Delineamento I em populações de maracujazeiro. *Revista Ciência Agronômica*, 42:495-501.
- Norvig, P. and Russell, S. (2013) Inteligência Artificial. 3rd ed. CAMPOS, Rio de Janeiro, Elsevier, 958p.

- Pedrozo, C. A.; Benites, F. R. G.; Barbosa, M. H. P.; Resende, M. D. V.; et al. (2009) Eficiência de índices de seleção utilizando a metodologia REML/BLUP no melhoramento da cana-de-açúcar. *Scientia Agraria*, 10:031-036.
- Peixoto, L. A. Redes Neurais Artificiais na predição do valor genético. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 97p, 2013.
- Pompeu, R. C. F. F.; Uchoa, F. C.; Neiva, J. N. M.; Oliveira Filho, G. S.; et al., (2003). Produção de matéria seca e qualidade de quatorze cultivares de alfafa (*Medicago sativa* L.) sob irrigação no Estado do Ceará. *Revista Ciência Agronômica*, 34:153-160.
- Rosenblatt, F. (1958) The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Rev.* 65:386–408
- Rumelhart, D.E., Hinton, G.E. e Williams, R.J. (1986). Learning representations by backpropagating errors. *Nature*. 323:533–536.
- Sant'anna, I. C. Redes Neurais Artificiais na discriminação de populações de retrocruzamento com diferentes graus de similaridade. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 115p, 2014.
- Silva, G. N. Redes Neurais Artificiais: Novo paradigma para a predição de valores genéticos. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, 92p, 2014.
- Smith, H. F. (1936) A discriminant function for plant selection. *Annals of Eugenics*, 7:240-250.
- Sudheer, K. P; Gosain, A. K.; Ramasastry, K. S. (2003) Estimating actual evapotranspiration from limited climatic data using neural computing technique. *Journal of Irrigation and Drainage Engineering*, 129:214-218.
- Tai, G. C. C. (1977) Index selection with desired gains. *Crop Sci.*, 17:182-183.
- Tallis, G. M. (1962) A selection index for optimum genotype. *Biometrics*, 18:120-122.

- Teodoro, P. E.; Barroso, L. M. A.; Nascimento, M.; Torres, E. S.; Santos, A.; Ribeiro, L. P. (2015) Redes neurais artificiais para identificar genótipos de feijão-caupi semiprostrado com alta adaptabilidade e estabilidade fenotípicas. *Pesq. agropec. bras.*, 50:1054-1060.
- Vasconcelos, E. S.; Ferreira, R. P.; Cruz, C. D.; Moreira, A.; et al. (2010) Estimativas de ganho genético por diferentes critérios de seleção em genótipos de alfafa. *Rev. Ceres*, 57:205-210.
- Viana, M. C. M.; Purcino, H. M. A.; Konzen, E. A.; Botrel, M. A.; et al. (2004) Avaliação de cultivares de alfafa nas condições de Cerrado no estado de Minas Gerais. *Pesq. Agropec. bras.*, 39:289-292.
- Widrow, B. e Hoff, M.E. (1960) Adaptive switching circuits. IRE West. Electr. Show Conv. Rec.: 96–104.
- Williams, J. S. (1962) The evaluation of a selection index. *Biometrics*, 18:375-393.
- Zhou, M. M.; Collins, A.; Kimbeng, C. A.; Tew, T. L.; Gravois, K. A. and Pontif, M. J. (2011) Artificial neural networks models as a decision support tool for selection in sugarcane: a case study using seedling populations. *Crop Science*, 51:21-31.

CAPÍTULO 1

TÉCNICAS BIOMÉTRICAS APLICADAS À SELEÇÃO EM GENÓTIPOS DE ALFAFA

RESUMO

O melhoramento da cultura da alfafa considera tanto a produtividade como o valor nutritivo da forragem. Como a maioria dos caracteres de importância da cultura são quantitativos, realizar a seleção acurada não é uma tarefa simples. O objetivo deste trabalho foi utilizar métodos de seleção direta, indireta e simultânea para seleção de genótipos de alfafa com base em caracteres de produção e valor nutritivo. Foram utilizados dados de quatro cortes de alfafa, com 77 genótipos cada, provenientes da Embrapa Pecuária Sudeste. Os caracteres avaliados foram subdivididos em dois grupos: Produção e Valor Nutritivo. Foram estimados os ganhos por seleção direta, indireta e simultânea para cada um dos grupos, considerando a seleção de 25% das melhores famílias. De modo geral, as seleções diretas proporcionaram ganhos *per se* interessantes para os propósitos do melhoramento e as seleções indiretas apresentaram dificuldades para manter ganhos desejados em todos os caracteres no conjunto de genótipos. O quarto corte, colhido sob déficit hídrico, apresentou resultado satisfatório para seleção direta e indireta quando se selecionou diretamente o caractere proteína bruta. Todavia, os resultados de seleções diretas e indiretas devem ser avaliados criteriosamente para evitar interpretações equivocadas, já que os caracteres são muito influenciados pelo ambiente. Os resultados da seleção simultânea utilizando o índice de Taí proporcionaram uma distribuição de ganhos mais equilibrada para os conjuntos de caracteres em todos os cortes. Entretanto a reunião de genótipos que proporcione ganhos equilibrados dentro de cada grupo de caracteres, e que posteriormente sejam favoráveis em outro grupo de caracteres nem sempre é possível. Por isso, outras estratégias podem ser utilizadas para seleção, como por exemplo a seleção em *tandem*. Para o presente trabalho, também deve ser considerado que o conjunto de genótipos estudados já apresentava alto valor nutritivo e por isso, a seleção para esse grupo de caracteres poderia ser mais flexível. Neste caso, um maior número de genótipos poderia ser selecionado atendendo aos objetivos de alta produção e alto valor nutritivo.

1. INTRODUÇÃO

O grande desafio do melhorista é desenvolver materiais superiores e de ampla aceitação no mercado. Para isso, é necessário o conhecimento aprofundado da estrutura genética da população que se quer melhorar (Borém e Miranda, 2013). No melhoramento da cultura da alfafa, os objetivos principais têm sido a produtividade, o valor nutritivo da forragem, além da tolerância aos estresses bióticos e abióticos (Li e Brummer, 2012).

A maioria dos caracteres de importância da cultura da alfafa tem controle poligênico e por isso, realizar a seleção de forma acurada não é uma tarefa simples (Cruz, 2006). Os fatores ambientais têm sido determinantes em todo o processo seletivo, somado a isso, as mudanças climáticas e o aquecimento global vêm motivando estudos mais elaborados e ações mais criteriosas no melhoramento genético (Sant'anna, 2014).

Melhorar geneticamente a alfafa inclui além de aumento de produção, melhoria no valor nutritivo da forragem (Li e Brummer, 2011). Entre os caracteres que determinam o valor nutritivo da forrageira, pode-se citar o teor de proteína bruta, digestibilidade “*in vitro*” da matéria seca, lignina, fibra em detergente neutro e fibra em detergente ácido. O teor de proteína bruta apresenta correlação positiva com o conteúdo de energia digestível das forragens e correlação negativa com o conteúdo de fibras presente nos alimentos (Almeida et al., 1999; Minson, 1982).

Segundo Basigalup e Odorizzi (2011), as estratégias de seleção para mais de um caráter em genótipos de alfafa têm sido a seleção em *tandem* em que se melhora independentemente cada caráter por vez até se atingir os objetivos, níveis independentes de seleção, onde são selecionados a cada ciclo, genótipos que satisfaçam os níveis determinados de cada caráter e a seleção por índices, onde as unidades de seleção são avaliadas conjuntamente em relação a uma série de caracteres de interesse. De todas as estratégias, a seleção por índices tem se mostrado mais eficiente (Hazel e Lush, 1942). As estratégias de seleção mais utilizadas em programas de melhoramento incluem a seleção direta, indireta, seleção entre e dentro, seleção combinada e a seleção simultânea:

A seleção direta é praticada quando se objetiva obter ganhos em um único caráter. Esse tipo de seleção pode proporcionar ganhos individuais superiores nos

caracteres sob seleção, no entanto, se o caráter sofrer grande influência do meio a precisão da seleção será baixa, isso devido à existência das correlações genéticas que levam à alteração do comportamento de diversos outros caracteres, tornando o seu uso limitado (Cruz et al, 2014). A resposta indireta ou correlacionada é utilizada principalmente para seleção de caracteres que apresentam baixa herdabilidade ou que são de difícil mensuração (Cruz et al., 2014), nesses casos, a seleção indireta com base em caracteres secundários de baixa influência ambiental, fácil mensuração e correlacionados com a característica de interesse é uma alternativa bastante interessante para maximizar a acurácia preditiva do material sob avaliação.

Para populações estruturadas em famílias recomenda-se o uso da seleção entre e dentro baseada nos valores fenotípicos que engloba uma maior fração da variância aditiva, mas apresenta a desvantagem de basear-se em valores fenotípicos individuais. Assim, materiais de desempenho medianos, porém interessantes, podem ser excluídos por não se destacarem nas seleções iniciais (Cedillo et al., 2008; Costa et al., 2000). Para contornar este problema, recomenda-se o emprego da seleção combinada que tem a particularidade de identificar de genótipos superiores a partir da informação do próprio indivíduo e de seus parentes a partir dos valores fenotípicos dos indivíduos, medidos como o desvio da média da população (Cruz et al., 2014). Esse tipo de seleção permite uma inferência mais acurada sobre o valor genético dos indivíduos sob seleção (Falconer e Mackay, 1996). Por fim, destaca-se a seleção simultânea que maximiza as chances de êxito no processo de melhoramento (Cruz et al., 2014), pois proporciona ganho equilibrado para todos os caracteres e obtenção de ganho até mesmo para caracteres desfavoravelmente correlacionados.

Na cultura da alfafa, informações sobre a melhor estratégia de seleção permite maximizar ganhos e orientar a escolha de genótipos superiores. Vasconcelos et al. (2010) encontraram maiores ganhos para alguns caracteres de alfafa quando se aplicou a seleção direta, todavia essa seleção levou ao aumento da suscetibilidade dos cultivares em relação a incidência de doenças. Esses autores destacaram que nenhuma das seleções diretas para alfafa propiciaram ganhos satisfatórios para todos os caracteres avaliados.

O presente trabalho teve como objetivo testar a eficiência dos métodos de seleção direta, indireta e simultânea para seleção de genótipos de alfafa com base em caracteres de produção e valor nutritivo da forragem.

2. MATERIAL E MÉTODOS

2.1. Material vegetal e caracteres mensurados

Para realização deste estudo, foram utilizadas informações de genótipos de alfafa envolvendo um período de cultivo com quatro cortes, provenientes de ensaios experimentais realizados na Embrapa Pecuária Sudeste, localizada no município de São Carlos, latitude 22° 01' 03" S e longitude 47° 53' 27" W". Os cortes foram realizados nos meses de novembro de 2015, fevereiro, maio e agosto de 2016. O delineamento utilizado em cada experimento foi o de blocos casualizados, com três repetições, de acordo com a Equação 1:

$$Y_{ij} = \mu + G_i + B_j + e_{ij} \quad (\text{Equação 1})$$

Em que Y_{ij} é a observação referente ao tratamento i avaliado no bloco j ; μ é a média geral; G_i é o efeito do i -ésimo tratamento, com $i = 1, 2, \dots, I$; B_j é o efeito do j -ésimo bloco, com $j = 1, 2, \dots, J$; e e_{ij} é o erro experimental, tal que $e_{ij} \sim \text{NID}(0; \sigma^2)$.

Cada unidade experimental era constituída por quatro de 4 m de comprimento, com espaçamento de 20 cm entre fileiras. A área útil da parcela era as duas fileiras centrais, eliminando-se 0,5 m de cada extremidade. Todos os experimentos foram irrigados, com exceção do corte quatro, que estava sob déficit hídrico. Todos os demais tratamentos culturais foram realizados de acordo com o recomendado para a cultura. Em cada corte foram coletados dados de 76 genótipos mais uma testemunha, a cultivar Crioula, conforme mostrado na Tabela 1:

Tabela 1 – Genótipos de alfafa avaliados no município de São Carlos, SP.

	Acessos		Acessos
1	5681	39	Maitena
2	ACA 900	40	Mecha
3	ACA 901	41	MEDINA
4	ACTIVA	42	Milonga II
5	Bacana	43	MONARCA

6	BACANA	44	MONARCA SP INTA
7	Bar Pal 5	45	P 30
8	Bar Pal 10	46	P 5715
9	Baralfa 85	47	Patriarca
10	BÁRBARA SP INTA	48	PATRICIA
11	CALIFORNIA 50	49	Pintado
12	CUF 101	50	PINTO
13	CW 1010	51	PRIMAVERA
14	CW 194	52	ProINTA Carmina
15	CW 620	53	ProINTA Luján
16	CW 830	54	ProINTA Mora
17	DIAMOND	55	ProINTA Patricia
18	DK 166	56	PROINTA PATRICIA
19	DK 181	57	ProINTA Super Monarca
20	DK 187 R	58	Queen 910
21	DK 192	59	RIO GRANDE
22	DK 194	60	Ruano
23	Don Enrique	61	Ruano
24	F 708	62	SEQUEL
25	FLÓRIDA 77	63	SEQUEL 2
26	G 909	64	SIRIVER 2
27	GAPP 969	65	SPS 6550
28	Gateado	66	TRINIDAD 87
29	KERN	67	Verdor
30	LE N 1	68	Verzi
31	LE N 2	69	VICTORIA SP INTA
32	LE N 3	70	Villa
33	LE N 4	71	WINTER
34	LPS 8500	72	WL 1058
35	Magna 601	73	WL 516
36	Magna 804	74	WL 525
37	Magna 860	75	WL 818
38	Magna 868	76	WL 903
		77	Crioula

Os caracteres avaliados foram: Altura de plantas dada pela média de cinco plantas em centímetros na área útil da parcela quando as plantas atingissem 10% de florescimento; Produção de matéria seca (kg/ha) medida pelos cortes feitos a uma altura de 8 a 10 cm do nível do solo, sempre que cada cultivar atingia 10% da floração ou quando as brotações basais atingissem altura média de 3 a 5 cm; Relação caule/folha: feita na matéria verde, durante os cortes de produção;

Além disso foram feitas análises de proteína bruta (PB); digestibilidade “*in vitro*” da matéria seca (DIVMS); fibra em detergente neutro (FDN); fibra em detergente ácido (FDA); lignina e suscetibilidade a doenças – estimada visualmente sendo a severidade determinada de acordo com a porcentagem de área foliar atacada em cada parcela, classificando as cultivares da seguinte maneira: a) nota 0 - susceptível (>50% de severidade da doença); b) nota 1 - baixa resistência (49 a 30% de severidade da doença); c) nota 2 - moderadamente resistentes (29 a 11% de severidade da doença); d) nota 3 - resistentes (< 11% de severidade da doença).

2.2. Métodos de seleção direta, indireta e simultânea para seleção de genótipos de alfafa com base em caracteres complexos

Os caracteres avaliados foram subdivididos em dois grandes grupos: Produção e Valor Nutritivo. As variáveis do grupo “Produção” foram: altura de plantas, produção de matéria seca, relação caule/folha e suscetibilidade a doenças, já o grupo “Valor nutritivo” foi composto pelos caracteres: proteína bruta, lignina, digestibilidade “*in vitro*” de matéria seca, fibra em detergente neutro e fibra em detergente ácido.

2.2.1. Ganhos por seleções direta e indireta

Foram estimados os ganhos por seleção direta e indireta entre os 77 genótipos de alfafa para cada um dos grupos de características anteriormente mencionados, em cada corte avaliado, considerando a seleção de 25% dos melhores genótipos. Considerando o grupo “produção”, a seleção para os caracteres altura de plantas, produção e doenças foi no sentido de acréscimo enquanto que para relação caule/folha foi desejado o decréscimo. Considerando o grupo “valor nutritivo”, estabeleceu-se acréscimo para proteína bruta e digestibilidade “*in vitro*” de matéria seca e decréscimo para os demais caracteres.

As estimativas de ganho direto foram obtidas segundo a expressão:

$$\Delta G\% = \frac{100(DS \times h^2)}{\bar{x}}$$

em que:

$\Delta G\%$ é o ganho de seleção expresso em porcentagem

DS é o diferencial de seleção

h^2 é o coeficiente de herdabilidade

\bar{x} é a média original do caractere X

Os ganhos indiretos foram estimados por:

$$\Delta G_{X(Y)}\% = \frac{DS_{X(Y)} \times h_X^2}{\bar{x}}$$

Onde:

$\Delta G_{X(Y)}\%$ é o ganho de seleção indireto em X selecionando em Y

$DS_{X(Y)}$ é o diferencial de seleção indireto no caráter X selecionando em Y

h_X^2 é o coeficiente de herdabilidade do caráter X

\bar{x} é a média original do caráter X

2.2.2. Seleção simultânea: Índice de Taí (1997)

Para utilização da seleção simultânea, foi considerado o índice de seleção proposto por Taí (1977), pois ele permite que caracteres secundários, para os quais não se deseja ganho máximo, sejam incluídos no índice para auxiliar no ganho dos caracteres principais. Para construção deste índice, deve-se considerar que os valores econômicos dos caracteres secundários sejam nulos no agregado genotípico (Cruz et al., 2014).

O ganho desejado para os vários caracteres é expresso por:

$$\Delta G_d = \frac{G\hat{b}}{\hat{\sigma}_1}$$

Onde:

ΔG_d é o vetor de ganhos desejados

G é a matriz de variâncias e covariâncias genéticas entre os caracteres

\hat{b} é o vetor $n \times 1$ de coeficientes do índice

$\hat{\sigma}_I$ é a raiz quadrada da variância do índice

Como $\hat{\sigma}_I$ é um escalar, ele não afeta a proporcionalidade do índice e por isso, pode ser desprezado. Portanto, o índice seria estabelecido a partir apenas do vetor de ganhos desejados e da matriz de variâncias e co-variâncias genotípicas. Segundo Cruz et al. (2014), o vetor de ganhos desejados pode ser estabelecido com base em um desvio padrão genético para cada caráter.

Para o grupo de produção, o caráter produção de matéria seca foi considerado como principal e os demais (altura de plantas, doenças e relação caule/folha), secundários. Já para o grupo “valor nutritivo” os caracteres proteína bruta e digestibilidade “*in vitro*” de matéria seca foram considerados como principais para inclusão no índice. Todas as análises foram realizadas utilizando-se o aplicativo GENES (Cruz, 2013).

3. RESULTADOS E DISCUSSÃO

3.1. Desempenho dos genótipos e estimativas dos parâmetros genéticos e ambientais

Na Tabela 2 são apresentadas as estimativas dos parâmetros genéticos dos caracteres do grupo produção. De modo geral, os coeficientes de variação para os caracteres assumiram valores compatíveis para a cultura da alfafa. (Vasconcelos et al., 2010). No corte 1, os genótipos apresentaram as maiores médias para produção de matéria seca, altura de plantas e relação caule/folhas quando comparado aos demais cortes, entretanto a relação CVg/CVe foi baixa indicando um cenário desfavorável para a seleção e obtenção de ganhos genéticos, já para o caráter doenças, as estimativas de variâncias foram nulas, mas é importante destacar que a média dos genótipos para esse caráter os classifica como moderadamente resistentes. Para o caráter produção de matéria seca a variabilidade genotípica de baixa magnitude revelou um cenário desfavorável para seleção. A herdabilidade para esse caráter foi baixa, se comparada aos valores obtidos para os demais cortes, indicando a ocorrência de alta variância fenotípica. Neste caso é prudente inferir que a alta variação ambiental não permitiu a expressão da variância genotípica.

Tabela 2 – Estimativas dos parâmetros genéticos e ambientais associados aos genótipos de alfafa avaliados quanto aos caracteres do grupo produção.

Cortes	Parâmetros	PROD MS*	ALT	C/F	DOENÇAS
C1	Média Geral	2287,25	51,50	0,83	2,40
	CV (%)	24,43	6,17	10,99	23,57
	CVg (%)	6,44	4,73	4,34	0,00
	h ²	17,27	63,77	31,89	0,00
C2	Média Geral	1631,66	43,91	0,67	2,75
	CV (%)	21,16	6,70	13,35	12,13
	CVg (%)	12,59	7,25	9,70	9,78
	h ²	51,52	77,82	61,26	66,10
C3	Média Geral	1841,34	50,57	0,73	1,73
	CV (%)	18,57	9,47	13,81	31,69
	CVg (%)	11,47	9,34	9,15	12,77
	h ²	53,34	74,49	56,83	32,75
C4	Média Geral	1362,77	36,81	0,52	2,91
	CV (%)	18,96	16,24	18,53	9,89
	CVg (%)	29,04	26,24	14,22	0,82
	h ²	87,55	88,68	63,88	2,02

(*) Produção de matéria seca – PROD MS; Altura de planta – ALT; Relação caule/folhas – C/F; Suscetibilidade a doenças – DOENÇAS

Para os cortes 2 e 3 (Tabela 2), os valores de herdabilidade para altura de plantas foram altos (77, 82% e 74,49% respectivamente), indicando a alta confiabilidade do valor fenotípico como indicador do valor genotípico (Falconer, 1981). Para os caracteres produção de matéria seca e doenças, as estimativas de coeficiente de variação genotípico assumiram valores superiores que os obtidos no corte 1. Assume-se que baixas estimativas de CVg, como encontrado em C1, são resultado de baixos valores de variância genotípica (Silva et al., 2012). Nos cortes 2 e 3, a melhor qualidade ambiental permitiu o aparecimento da variância genotípica, mesmo que em menor magnitude do que o corte 4. No corte 4 (Tabela 2), em que os dados foram obtidos sob condições de déficit hídrico, o caráter doenças apresentou CVg e herdabilidade baixos, indicando a baixa variabilidade genotípica. Para os caracteres produção de matéria seca e altura de plantas, os valores de CVg e herdabilidade foram altos, demonstrando um cenário favorável para a seleção, indicando que as diferenças detectadas são de natureza predominantemente genética, porém deve-se chamar a atenção para o fato de que esta condição experimental impactou o desempenho dos

genótipos proporcionando redução considerável em suas médias. Como a alfafa deste corte foi mantida sob déficit hídrico, as estimativas dos parâmetros genéticos permitem selecionar genótipos mais tolerantes à seca, fato bastante interessante para o programa de melhoramento da cultura.

Para fins de seleção indireta procura-se características auxiliares de alta herdabilidade e pouco influenciadas pelo ambiente. Assim, tendo em vista os resultados encontrados (Tabela 2), destaca-se a característica altura de plantas cujo valor da herdabilidade sempre foi superior à da característica produção de matéria seca. A característica relação caule/folha (C/F) tem potencial de uso para fins de resposta correlacionada apenas nos cortes sem déficit hídrico pois nestas condições se torna fortemente influenciada pelo ambiente elevando o coeficiente de variação experimental e reduzindo a herdabilidade a níveis menos favoráveis do que a característica principal que é a produção de matéria seca. A característica suscetibilidade a doenças, tem comportamento imprevisível considerando os fatores incluídos no modelo estatístico. Seus efeitos dependem de vários outros fatores tais como ocorrência de patógenos, populações hospedeiras, derivas genéticas e combinações de fatores ambientais favoráveis.

Na Tabela 3 são apresentadas as estimativas dos parâmetros genéticos dos caracteres do grupo valor nutritivo. De modo geral, os coeficientes de variação para os caracteres assumiram valores compatíveis para a cultura da alfafa (Vasconcelos, et al., 2010). À semelhança do obtido para o grupo de caracteres de produção, para os caracteres de valor nutritivo no corte 1, a relação CV_g/CV_e foi baixa indicando um cenário desfavorável para o ganho com a seleção. Para o caractere DIVMS a estimativa de variância foi pequena, este resultado também foi observado nos demais cortes ao longo do ano. Resultado semelhante foi relatado por Rassini et al. (2007), os autores ainda destacam que valores entre 65% e 70% para este caráter indicam alta digestibilidade, dessa forma, a baixa variabilidade apresentada pelos genótipos para esse particular caráter assume grande importância. Os caracteres lignina, FDN e FDA apresentaram valores médios de 8,51; 44,37 e 30,99 respectivamente, compatíveis com os valores ótimos para a cultura (Rassini et al., 2007).

Tabela 3 – Estimativas dos parâmetros genéticos e ambientais associados aos genótipos de alfafa avaliados quanto aos caracteres do grupo valor nutritivo.

Cortes	Parâmetros	LIGNINA*	PB	DIVMS	FDN	FDA
--------	------------	----------	----	-------	-----	-----

C1	Média Geral	8,51	25,70	69,27	44,37	30,99
	CV (%)	17,43	5,46	3,74	11,95	10,21
	CVg (%)	10,05	2,94	1,15	5,90	4,94
	h ²	49,93	46,45	22,02	42,26	41,28
C2	Média Geral	7,83	26,76	69,28	47,04	29,62
	CV (%)	5,38	3,82	1,75	4,31	4,18
	CVg (%)	0,00	2,87	0,37	0,00	1,67
	h ²	0,00	62,97	11,85	0,00	32,36
C3	Média Geral	7,18	26,99	69,17	42,85	31,51
	CV (%)	5,84	5,85	1,89	3,84	4,01
	CVg (%)	1,42	5,82	0,65	1,82	2,31
	h ²	15,05	74,79	26,17	40,44	49,84
C4	Média Geral	7,18	26,48	71,18	39,49	30,57
	CV (%)	8,24	3,54	1,57	6,96	4,78
	CVg (%)	3,30	3,61	0,70	2,25	0,00
	h ²	32,50	75,66	37,17	23,82	0,00

(*) Teor de lignina – LIGNINA; Proteína bruta – PB; Digestibilidade “*in vitro*” de matéria seca – DIVMS; Fibra em detergente neutro – FDN; Fibra em detergente ácido – FDA.

A média para PB em todos os cortes variou entre 25,70 e 26,48 (Tabela 3). Esses valores foram superiores aos encontrados por Costa et al. (2006) e Vasconcelos (2010), e representam valores ótimos para o caráter. De modo geral, excluindo-se o primeiro corte, as estimativas de herdabilidade para este caráter foram altas, indicando alta relação entre valores fenotípicos e genotípicos. No corte 2, os caracteres lignina e FDN apresentaram variância nula e por isso não foi possível obter estimativas de CVg e herdabilidade. Já no corte 4, o mesmo foi observado para o caráter FDA (Tabela 3). Destaque deve ser dado para os valores de PB e DIVMS (26,48 e 71,18) que se mantiveram altos mesmo no ambiente desfavorável, este fato pode ser explicado pela maior produção observada nos cortes irrigados que leva à diluição da PB na forragem produzida (Ruggieri et al., 2001). É interessante perceber que os valores dos caracteres de valor nutritivo para este corte se mantiveram no mesmo patamar dos demais cortes. Botrel et al. (2001) relataram que na época da seca os valores dos caracteres de valor nutritivo caem acentuadamente.

Para fins de predição de ganhos por seleção simultânea ou indireta é indispensável o conhecimento da relação entre os caracteres. De início procura-se calcular a magnitude, o sinal e a natureza das correlações entre caracteres verificando

se são determinadas geneticamente ou resultantes de causas ambientais (Cruz et al., 2012), pois tal fato tem papel fundamental no entendimento dos ganhos de seleção nos programas de melhoramento. As estimativas de correlações genóticas e fenotípicas entre os caracteres para o grupo produção são apresentadas na Tabela 4. Para o corte 1 realizou-se a análise de correlação entre altura de plantas e relação caule/folha, uma vez que os demais caracteres haviam apresentado variância genética não significativa, sendo a correlação entre eles significativa (Tabela 2). Observou-se que a estimativa da correlação genotípica superou o valor da correlação fenotípica, indicando que os fatores genéticos que controlam estes caracteres têm influência superior na determinação das correlações em relação aos fatores ambientais.

Nos demais cortes as correlações entre produção de matéria seca e altura de plantas e produção de matéria seca e relação caule/folhas foram positivas e de médias a elevadas magnitudes, indicando uma relação linear entre estes caracteres. Dessa forma, apesar de os valores dos caracteres terem sofrido alterações ao longo dos cortes, sugere-se que a seleção possa ser praticada para produção de matéria seca, por exemplo, sem que haja mudanças na posição relativa dos genótipos em relação a relação caule/folhas ou altura de plantas (Ferreira et al., 1999).

No corte 3, as estimativas de correlação genotípica entre os pares de caracteres doenças e altura, e doenças e relação C/F foram não significativas, enquanto as correlações fenotípicas entre estes mesmos pares foram significativas. Esse resultado ocorreu possivelmente, pela diferença entre os testes utilizados para cada tipo de correlação (teste t e bootstrap), já que as magnitudes não foram muito discrepantes entre si. No corte 4 o caráter doenças também foi retirado da análise de correlações por ter apresentado variância nula (Tabela 2).

Tabela 4 – Estimativas de correlação genotípica (diagonal superior) e fenotípica (diagonal inferior) dos caracteres produção de matéria seca (PROD MS), altura de plantas (ALT), relação caule/folhas (C/F) e doenças em alfafa.

CORTE 1	ALT	PROD MS	C/F	DOENÇAS
ALT		--	0,9369++	--
PROD MS	--		--	--
C/F	0,5622**	--		--
DOENÇAS	--	--	--	

CORTE 2	ALT	PROD MS	C/F	DOENÇAS
ALT		0,8886 ⁺⁺	0,7146 ⁺⁺	0,8367 ⁺⁺
PROD MS	0,6489 ^{**}		0,7114 ⁺⁺	0,6760 ⁺⁺
C/F	0,5769 ^{**}	0,5299 ^{**}		0,5850 ⁺⁺
DOENÇAS	0,6654 ^{**}	0,4352 ^{**}	0,4322 ^{**}	

CORTE 3	ALT	PROD MS	C/F	DOENÇAS
ALT		1,0532 ⁺⁺	1,1368 ⁺⁺	0,1382 ^{ns}
PROD MS	0,8872 ^{**}		0,9116 ⁺⁺	-0,2342 ^{ns}
C/F	0,7259 ^{**}	0,7414 ^{**}		0,3988 ^{ns}
DOENÇAS	0,2376 [*]	0,1698 ^{ns}	0,3039 ^{**}	

CORTE 4	ALT	PROD MS	C/F	DOENÇAS
ALT		0,9405 ⁺⁺	0,9263 ⁺⁺	--
PROD MS	0,8501 ^{**}		0,9325 ⁺⁺	--
C/F	0,7259 ^{**}	0,7428 ^{**}		--
DOENÇAS	--	--	--	

^{*}, ^{**} Significativo a 5 e 1% de probabilidade, respectivamente, pelo teste t. ^{ns} Não significativo; ⁺, ⁺⁺ Significativo a 5 e 1% de probabilidade, respectivamente, pelo método de bootstrap com 1000 simulações.

As estimativas de correlações genotípicas e fenotípicas entre os caracteres para o grupo valor nutritivo são apresentadas na Tabela 5. Para o corte 1 retirou-se o caráter DIVMS da análise de correlação pois a variância genética deste foi não significativa (Tabela 2). As correlações genotípicas e fenotípicas entre lignina e proteína bruta e FDN e proteína bruta foram não significativas, no entanto ausência de relação linear entre caracteres não significa a independência entre os mesmos (Cruz et al., 2012). Foram detectadas diferenças na significância entre as correlações fenotípicas e genotípicas para alguns pares de caracteres tal como para o par FDN e Lignina, causadas pelos diferentes critérios adotados por cada teste utilizado (teste t e bootstrap), somado a isso, a qualidade ambiental pode influenciar nas estimativas das correlações.

Tabela 5 – Estimativas de correlação genotípica (diagonal superior) e fenotípica (diagonal inferior) dos caracteres lignina, proteína bruta (PB), digestibilidade “*in vitro*” de matéria seca (DIVMS), fibra em detergente neutro (FDN) e fibra em detergente ácido (FDA) em alfafa.

CORTE 1	LIGNINA	PB	DIVMS	FDN	FDA
LIGNINA		0,0620 ^{ns}	--	0,4866 ^{ns}	0,6110 ⁺
PB	-0,0232 ^{ns}		--	-0,3900 ^{ns}	-0,2011 ^{ns}

DIVMS	--	--	--	--	--
FDN	0.3627**	-0.1053 ^{ns}	--	--	0,7459 ⁺
FDA	0.5225**	-0.3267**	--	0.5267**	--
CORTE 2	LIGNINA	PB	DIVMS	FDN	FDA
LIGNINA	--	--	--	--	--
PB	--	--	--	--	-1,1634 ⁺⁺
DIVMS	--	--	--	--	--
FDN	--	--	--	--	--
FDA	--	-0.6079**	--	--	--
CORTE 3	LIGNINA	PB	DIVMS	FDN	FDA
LIGNINA	--	--	--	--	--
PB	--	--	--	-0,6868 ⁺⁺	-0,8255 ⁺⁺
DIVMS	--	--	--	--	--
FDN	--	-0.5790**	--	--	0,5586 ⁺
FDA	--	-0.6164**	--	0.4391**	--
CORTE 4	LIGNINA	PB	DIVMS	FDN	FDA
LIGNINA	--	0,3087 ^{ns}	0,7620 ⁺⁺	--	--
PB	0.1691 ^{ns}	--	1,0935 ⁺⁺	--	--
DIVMS	0.1390 ^{ns}	0.6289**	--	--	--
FDN	--	--	--	--	--
FDA	--	--	--	--	--

*, ** Significativo a 5 e 1% de probabilidade, respectivamente, pelo teste t. ^{ns} Não significativo; ⁺, ⁺⁺ Significativo a 5 e 1% de probabilidade, respectivamente, pelo método de bootstrap com 1000 simulações.

Nos demais cortes muitas variáveis foram excluídas da análise por apresentarem variâncias genéticas não significativas, mas ainda assim, foi possível detectar correlações genotípicas e fenotípicas baixas entre FDN e lignina e FDA e lignina, indicando um comportamento diferencial entre os genótipos, onde é necessário maior critério na seleção (Ferreira et al., 1999). As correlações negativas entre proteína bruta e FDN e proteína bruta e FDA encontradas nos cortes 2 e 3 sugerem uma associação linear decrescente entre esses caracteres. Em todos os cortes, as estimativas de correlações genotípicas significativas foram superiores do que as estimativas das correlações fenotípicas, indicando que os fatores ambientais influenciam as estimativas de correlações de forma menos significativa do que os fatores genéticos.

3.2. Seleções diretas e indiretas entre médias

Em um programa de melhoramento de alfafa, deseja-se aumento da produtividade, digestibilidade e que a forragem tenha mais folhas em detrimento de caules, pois estas garantem um alimento de alto valor nutritivo pela maior concentração de proteína bruta existente (Botrel et al., 2001). Somado a isso, procura-se redução nos valores dos componentes fibrosos e lignina, pois estes reduzem a digestibilidade da forragem (Oliveira, 2014).

De modo geral, pode-se considerar que as seleções diretas proporcionaram ganhos *per se* interessantes para os propósitos do melhoramento, tomando por exemplo a característica produção de matéria seca que atingiu patamar de ganho de 27,11% no último corte (Tabela 6). Considerando o grupo “produção”, para o corte 1 o caráter doenças apresentou variância genética nula e a produção de matéria seca foi não significativa, mas por não interferir no ganho para os demais, estes caracteres foram mantidos na análise, além disso a resposta correlacionada para produção foi inexpressiva. Para os demais cortes, destaca-se a resposta correlacionada da produção de matéria seca por meio da característica auxiliar altura de plantas evidenciando que, para os caracteres que são submetidos à seleção indireta, as relações existentes entre eles exercem grande influência na resposta.

Tabela 6 – Respostas preditas com as seleções diretas e indiretas entre os genótipos de alfafa para os caracteres produção de matéria seca (PROD MS), altura de plantas (ALT), relação caule/folhas (C/F) e doenças.

Corte	Seleção	Resposta esperada (%)			
		ALTURA	PROD MS	C/F	DOENÇAS
1	ALTURA	3,94	0,04	1,99	0,0
	PROD MS	0,55	3,48	0,52	0,0
	C/F	-3,12	-0,65	-2,89	0,0
	DOENÇAS	-0,46	-0,21	-0,47	0,0
2	ALTURA	6,60	4,77	4,16	5,01
	PROD MS	4,0	11,56	3,37	3,81
	C/F	-5,57	-6,92	-9,51	-5,37
	DOENÇAS	6,16	8,65	5,76	9,93
3	ALTURA	23,60	18,21	9,08	0,03
	PROD MS	18,58	24,14	7,93	0,02
	C/F	-24,5	-22,51	-14,28	-0,04

	DOENÇAS	4,54	4,07	3,73	0,06
	ALTURA	23,60	20,60	9,08	0,03
4	PROD MS	17,61	27,11	8,71	0,01
	C/F	-24,5	-28,60	-14,28	-0,04
	DOENÇAS	4,54	1,87	3,73	0,06

Segundo Cruz et al. (2014), a seleção direta pode proporcionar ganhos satisfatórios nos caracteres sob seleção, no entanto para caracteres que sofrem grande influência do meio, a precisão da seleção será baixa devido à existência das correlações genéticas que levam à alteração do comportamento de diversos outros caracteres. No corte 2, por exemplo, a seleção direta para produção de matéria seca, que apresentou herdabilidade de 51,52%, proporcionou ganhos positivos para todos os outros caracteres, inclusive para relação caule/folhas, onde se esperava ganho negativo. Segundo Falconer (1981) a resposta correlacionada pode proporcionar ganhos superiores à resposta direta, entretanto a expressão utilizada para o cálculo da resposta indireta, neste trabalho, possibilita que no máximo, a resposta correlacionada seja equivalente à resposta direta. Assim, a proximidade do ganho indireto por meio da seleção em altura de plantas em relação ao ganho direto com produção de matéria seca, em especial no corte 4, é indicativo do potencial desta característica para fins de seleção indireta ou simultânea no melhoramento da alfafa. De maneira geral os ganhos diretos em todos os caracteres foram condizentes com o objetivo pré-determinado (acréscimo nos valores de altura de plantas, produção, doenças e decréscimo para o caractere relação caule/folha).

Segundo Botrel et al. (2001) é fundamental que se mantenha a maior proporção de folhas na forragem colhida, pois isso resulta em uma forragem de alto valor nutritivo, característica que é importante principalmente no período de seca, quando o valor nutritivo das forragens cai acentuadamente. Utilizando as seleções univariadas, em nenhum corte foi possível reduzir a relação caule/folhas sem que houvesse redução em outros caracteres importantes para a cultura (Tabela 6). A correlação significativa entre este caráter e os caracteres produção de matéria seca e altura de plantas nos cortes 2, 3 e 4 já demonstrava a relação linear positiva entre eles (Tabela 2).

As respostas diretas e indiretas sobre as variáveis representativas do grupo valor nutritivo são apresentadas na Tabela 7. Verifica-se que para o corte 1, a seleção para

redução da lignina levou a resultado negativo para proteína bruta (Tabela 7). O ganho positivo para proteína bruta é importante pois valores abaixo do recomendado deste componente podem limitar a produção de leite (Oliveira, 2014).

Tabela 7 – Respostas previstas com as seleções diretas e indiretas nos genótipos de alfafa para os caracteres proteína bruta (PB), digestibilidade “*in vitro*” de matéria seca (DIVMS), fibra em detergente neutro (FDN) e fibra em detergente ácido (FDA).

Corte	Seleção	Resposta prevista (%)				
		LIGNINA	PB	DIVMS	FDN	FDA
1	LIGNINA	-8,22	-0,52	0,25	-2,30	-1,78
	PB	-1,31	2,65	0,37	-0,26	-1,68
	DIVMS	-4,60	1,66	0,70	-1,44	-2,14
	FDN	-3,40	0,71	0,25	-4,40	-2,68
	FDA	-3,18	1,21	0,32	-3,05	-3,45
2	LIGNINA	0,0	-0,01	0,0	0,0	0,24
	PB	0,0	2,86	0,13	0,0	-0,57
	DIVMS	0,0	2,25	0,17	0,0	-0,48
	FDN	0,0	0,88	0,03	0,0	-0,04
	FDA	0,0	1,86	0,07	0,0	-1,17
3	LIGNINA	-0,66	-0,49	-0,02	-0,80	-0,13
	PB	-0,06	7,07	0,38	-1,01	-1,40
	DIVMS	0,0	5,82	0,44	-0,72	-1,07
	FDN	-0,38	4,17	0,22	-1,51	-1,15
	FDA	-0,06	3,87	0,25	-0,68	-2,09
4	LIGNINA	-2,22	-0,26	0,02	-0,71	0,0
	PB	0,79	4,27	0,43	-0,41	0,0
	DIVMS	0,44	2,34	0,57	-0,37	0,0
	FDN	-1,25	1,92	0,24	-1,44	0,0
	FDA	0,44	1,72	0,14	0,02	0,0

Para o corte 2, mesmo que os caracteres lignina e fibra em detergente neutro tenham apresentado valores das estimativas de variância genotípica iguais a zero e DIVMS tenha apresentado variância genética não significativa, optou-se por mantê-los na análise uma vez que eles não irão influenciar no ganho dos demais caracteres (Tabela 7). De modo geral, as estimativas de ganho para esse corte foram baixas. Para o corte 3, os caracteres lignina e DIVMS tiveram variância genética não

significativa, não proporcionando ganhos para os genótipos. As respostas para a seleção direta foram bastante interessantes para os propósitos do melhoramento, principalmente para o caractere proteína bruta, que apresentou um ganho direto de 7,07%. Os ganhos indiretos obtidos com a seleção direta em proteína bruta representaram -1,01% e -1,40% para os caracteres fibra em detergente neutro e fibra em detergente ácido, respectivamente (Tabela 7), concordando com o sinal negativo das correlações entre esses caracteres (Tabela 5).

No corte 4, o valor da estimativa de variância genotípica para fibra em detergente ácido foi igual a zero e para FDN foi não significativa e, assim como nas situações anteriores, optou-se por manter esses caracteres na análise. Nesse caso, as estimativas de ganhos diretos e indiretos foram incompatíveis com o desejado, exceto para seleção direta em fibra em detergente neutro, que proporcionou aumento para proteína bruta e digestibilidade “*in vitro*” de matéria seca e redução para os demais caracteres (Tabela 3). Visto que a maioria dos caracteres da alfafa são muito influenciados pelo ambiente, mesmo resultados interessantes de seleção direta devem ser avaliados criteriosamente para evitar interpretações equivocadas.

No contexto de predição de ganhos por seleção direta e indireta, constata-se que resultados semelhantes foram encontrados por Vasconcelos et al. (2010) que, investigando as estimativas de ganho genético por diferentes critérios de seleção em alfafa, encontraram maiores ganhos para alguns caracteres com a seleção direta, mas obtiveram ganho indireto indesejado, levando ao aumento da suscetibilidade dos cultivares em relação a incidência de doenças. Esses autores destacaram que nenhuma das seleções diretas para alfafa propiciaram ganhos satisfatórios para todos os caracteres avaliados. Em comparação com outras espécies perenes, a obtenção de respostas favoráveis a um conjunto de caracteres também não tem sido tarefa fácil.

Em alguns trabalhos, como o apresentado por Martins et al. (2003), estudando a eficiência da seleção univariada direta e indireta e de índices de seleção em *Eucalyptus grandis*, há relatos de não conseguir eficiência tanto para as seleções diretas quanto para as indiretas. Os autores concluíram que os processos de seleção simultânea foram eficientes para proporcionar ganhos mais equilibrados no material sob avaliação. Outro exemplo encontra-se no trabalho realizado por Freitas et al. (2012), que estudando a formação de população base para seleção recorrente em maracujazeiro-amarelo, concluíram sobre a seleção de indivíduos com foco apenas

em produtividade que estes apresentaram estimativas de ganho indesejáveis para uma série de caracteres sob avaliação.

3.3. Seleção pelo método de Taí (1977)

Na busca de resposta à seleção mais equilibrada para um conjunto de característica de importância para a alfafa, foi adotada técnica fundamentada em seleção simultânea. Desta forma, para cada corte de alfafa foi construído um índice de seleção utilizando como vetor de ganhos desejados o equivalente a um desvio padrão genético. Assim como nas seleções anteriores, os caracteres foram divididos em dois grandes grupos, um de produção e o outro de valor nutritivo.

Para o grupo de caracteres de produção em todos os cortes houve ganhos positivos para relação caule/folha (Tabela 8). No corte 2 especialmente, o ganho para esse caractere foi alto (7,60%), resultado desfavorável uma vez que é desejável que a forragem tenha proporção de folhas superior à de caules (Botrel et al., 2001). É importante destacar que só se conseguiu ganhos negativos para relação caule/folha nas seleções diretas, mas estas levaram a respostas indesejáveis para os outros caracteres, isso devido às correlações positivas entre esse caráter e os demais (Tabela 4). Cruz et al. (2014) afirmam que a seleção simultânea de caracteres é uma prática que maximiza as chances de êxito no processo de melhoramento. De fato, a seleção simultânea foi mais eficiente que as seleções diretas e indiretas, pois proporcionaram na maioria dos cortes, ganhos mais equilibrados no conjunto de caracteres. Na literatura há relato de que os índices apresentam a possibilidade de obtenção de ganho até mesmo para caracteres desfavoravelmente correlacionados (Granate et al., 2002).

Tabela 8 – Seleção simultânea através do índice de seleção de Taí (1977) nos quatro cortes de alfafa para os caracteres produção de matéria seca (PROD MS), altura de plantas (ALT), relação caule/folhas (C/F) e doenças.

Cortes	Ganho de seleção (%)			
	ALTURA	PROD MS	C/F	DOENÇAS
1	1,52	0,64	2,96	0,00
2	5,51	5,43	7,60	5,81
3	18,54	18,12	3,79	0,06
4	20,54	23,55	5,23	0,03

Considerando os caracteres de valor nutritivo, os ganhos também foram mais equilibrados se comparados às seleções diretas e indiretas (Tabela 9). Para os cortes 1 e 2, as estimativas de ganhos atingiram valores aceitáveis para todos os caracteres envolvidos. Entretanto o ganho para os caracteres DIVMS no corte 1, lignina, DIVMS e FDN no corte 2, lignina e DIVMS no corte 3 e FDN e FDA no corte 4 não devem ser interpretados como ganhos reais, uma vez que estes tiveram variâncias genéticas não significativas (Tabela 3). Para o corte 4, houve ganho indesejado para lignina (1,04%). No entanto, como o ganho para os demais componentes foram satisfatórios e como o conjunto de genótipos estudados apresentam valores ótimos de valor nutritivo, este valor específico de ganho para esta característica não deve significar grandes problemas na seleção.

Tabela 9 – Seleção simultânea através do índice de seleção de Taí (1977) nos quatro cortes de alfafa para os caracteres lignina, proteína bruta (PB), digestibilidade “*in vitro*” de matéria seca (DIVMS), fibra em detergente neutro (FDN) e fibra em detergente ácido (FDA).

Cortes	Ganho de seleção (%)				
	LIGNINA	PB	DIVMS	FDN	FDA
1	-2,02	2,48	0,51	-1,08	-1,68
2	0,00	2,53	0,14	0,00	-0,93
3	0,05	6,54	0,42	-0,74	-1,19
4	1,04	4,18	0,42	-0,47	0,00

Apesar de este estudo ter considerado dois grupos separados de variáveis, é de interesse uma interpretação conjunta de forma a permitir identificar genótipos que demonstrem superioridade em relação a todos os caracteres mensurados que têm importância econômica e, ou, fitotécnica. Quando se utiliza a abordagem de índice de seleção obtém-se um “super caráter fenotípico” que contempla todas as variáveis de importância e é caracterizado por manter máxima a correlação com o agregado genotípico de forma que cada indivíduo passa a expressar um valor denominado score. Na Figura 7 é apresentada a dispersão dos genótipos de alfafa em relação aos grupos avaliados, onde os números em cada eixo representam os scores do índice de Taí e as retas que cortam os eixos representam os genótipos selecionados segundo cada grupo de caracteres. Considerando a seleção de 25% dos genótipos em cada um dos quatro cortes e em cada grupo de caracteres, a seleção simultânea

pele método de Taí foi mais eficiente do que as seleções univariadas pois permitiu ganhos equilibrados para cada conjunto de caracteres. Todavia, quando se confrontou os resultados obtidos nos cortes individualmente em relação aos grupos de produção e valor nutritivo, a seleção para ambos os grupos foi dificultada.

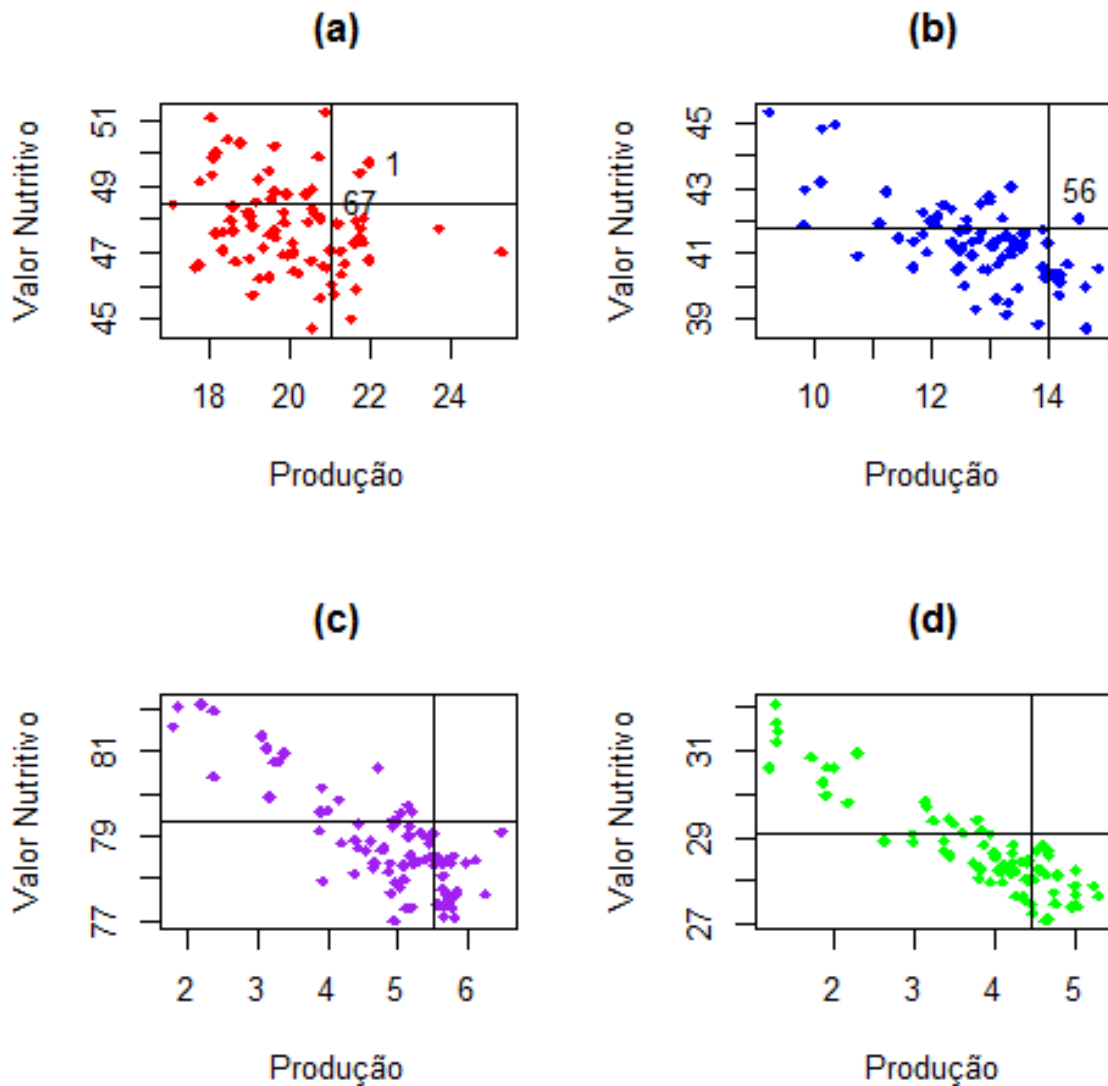


Figura 7 – Seleção dos genótipos selecionados pelo método de Taí (1977) para os grupos de caracteres de produção e valor nutritivo, em que: (a) = corte 1, (b) = corte 2, (c) = corte 3 e (d) = corte 4.

Somente nos cortes 1 (Figura 7a) e 2 (Figura 7b) foi possível identificar genótipos que seriam selecionados para ambos os grupos. No entanto, nota-se em todos os

cortes que existiram genótipos muito próximos do valor de seleção estabelecido nos dois grupos.

Atualmente, as estratégias de seleção para mais de um caráter em genótipos de alfafa segundo Basigalup e Odorizzi (2011) têm sido: i) a seleção em *tandem* em que se melhora independentemente cada caráter por vez até se atingir os objetivos. ii) níveis independentes de seleção, onde são selecionados a cada ciclo, genótipos que satisfaçam os níveis determinados de cada caráter. iii) e a seleção por índices, onde as unidades de seleção são avaliadas conjuntamente em relação a uma série de caracteres de interesse. A adoção da seleção simultânea tem suporte em dois relatos básicos. O primeiro refere-se ao fato de que, entre todas as estratégias, a seleção por índices tem se mostrado mais eficiente (Hazel e Lush, 1942). O segundo, no contexto agrônômico, fundamenta-se no relato de Li e Brummer (2012) de que o melhoramento da alfafa deva incluir além de aumento de produção, melhoria no valor nutritivo da forragem.

Entretanto, de maneira geral, pode-se concluir sobre as dificuldades encontradas em reunir genótipos que proporcione ganhos equilibrados dentro de cada grupo de caracteres, e que posteriormente sejam favoráveis em outro grupo de caracteres. Certamente a seleção simultânea não exclui a possibilidade de agregar estratégia de seleção em *tandem* para selecionar complexos ao invés de caracteres individuais. Para o presente trabalho, também deve ser considerado que, de acordo com os parâmetros de Rassini et al. (2007), o conjunto de genótipos estudados já apresenta alto valor nutritivo e por isso, a seleção para esse grupo de caracteres pode ser mais flexível. Neste caso, um maior número de genótipos poderia ser selecionado atendendo aos objetivos de alta produção e alto valor nutritivo.

4. CONCLUSÕES

O conjunto de genótipos apresenta valores ótimos para os caracteres de valor nutritivo, permitindo uma maior flexibilidade para realizar seleção;

Os métodos de seleção direta e indireta entre médias dos genótipos não são eficientes para proporcionar respostas desejáveis para todo o conjunto de caracteres de alfafa avaliados neste trabalho;

O índice de Taí (1977) foi que proporcionou uma distribuição de ganhos para o conjunto de caracteres de produção e valor nutritivo de alfafa mais equilibrada em todos os cortes avaliados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, M. I. V.; Ferreira, W. M.; Almeida, F. Q.; Gonçalves, L. C.; Rezende, A. S. C. (1999) Composição química e predição do valor nutritivo de dietas para equinos. *R. Bras. Zootec.* 28:1268-1278.
- Basigalup, D.H.; Odorizzi, A. S. (2011) Melhoramento genético da alfafa. In: Ferreira, R. P.; Basigalup, D. H.; Gioco, J. O. (Ed.). *Melhoramento Genético da Alfafa*. São Carlos: Embrapa Pecuária Sudeste, p. 225-260.
- Borém, A; Miranda, G. V. (2013) Melhoramento de plantas. Viçosa, Editora UFV. 523p.
- Botrel, M. A.; Ferreira, R. P.; Alvim, M. J.; Xavier, D. F. (2001) Cultivares de alfafa em área de influência da Mata Atlântica no estado de Minas Gerais. *Pesq. agropec. bras.*, 36:1437-1442.
- Cedillo, D. S. O.; Ferreira, F. M.; Barros, W. S.; Cruz, C. D.; et al. (2008) Selection among and within and combined selection in oil palm families derived from Dura x Dura. *Ciência Rural*, 38:65-71.
- Costa, M. M.; Meirelles, P. R.; Vieira, M. E. Q (2006) Produção de matéria seca e composição bromatológica de vinte cultivares de alfafa (*Medicago sativa* L.) em Botucatu-SP. *Veterinária e Zootecnia*, 12:42-51.
- Costa, R. B.; Resende, M. D. V.; Araujo, A. J.; Gonçalves, P. S.; et al. (2000) Seleção combinada univariada e multivariada aplicada ao melhoramento genético da seringueira. *Pesq. Agropec. bras.*, 35:381-388.
- Cruz, C. D. (2006) Genes: biometria. Viçosa, Editora UFV. 302p.
- Cruz, C. D.; Carneiro, P. C. S.; Regazzi, A. J. (2012) Modelos biométricos aplicados ao melhoramento genético. Viçosa, Editora UFV. 514p.
- Cruz, C. D. (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum.*, 35:271-276.
- Cruz, C. D.; Carneiro, P. C. S.; Regazzi, A. J. (2014) Modelos biométricos aplicados ao melhoramento genético. Viçosa, Editora UFV. 668p.

- Falconer, D. S. (1981) Introdução à genética quantitativa. Viçosa, Editora UFV, 279p.
- Falconer, D. S.; Mackay, T. F. (1996) Introduction to quantitative genetics. [S.l. : s.n.] 464p.
- Ferreira, R. P.; Botrel, M. A.; Pereira, A. V.; Cruz, C. D. (1999) Avaliação de cultivares de alfafa e estimativas de repetibilidade de caracteres forrageiros. *Pesq. agropec. bras.*, 34:995-1002.
- Freitas, J. P. X.; Oliveira, E. J.; Jesus, O. N.; Neto, A. J. C. et al. (2012) Formação de população base para a seleção recorrente em maracujazeiro-amarelo com uso de índices de seleção. *Pesq. Agropec. bras.*, 47:393-401.
- Granate, M. J.; Cruz, C. D.; Pacheco, C. A. P. (2002) Predição de ganho genético com diferentes índices de seleção no milho pipoca CMS-43. *Pesq. Agropec. bras.*, 37:1001-1008.
- Hazel, L. N. and Lush, J. L. (1942) The efficiency of three methods of selection. *Journal Heredity*, Oxford, 33:393-399.
- Li, X.; Brummer, E. C. (2012) Applied Genetics and Genomics in Alfalfa Breeding. *Agronomy*, 2:40-61.
- Minson, D.J. (1982) Effect of chemical composition on feed digestibility and metabolizable energy. *Nutr. Abstr. Rev.*, Series B, 52:591-615.
- Oliveira, A. (2014) Gado de leite – a importância das forrageiras no sistema de produção a pasto. Disponível em: <https://www.cpt.com.br/cursos-bovinos-gadodeleite/artigos/gado-de-leite-a-importancia-das-forrageiras-no-sistema-de-producao-a-pasto>. Acesso em: 28/05/2017.
- Rassini, J. B.; Ferreira, R. P.; Rodrigues, A. A.; Barioni Junior, W.; Moreira, A.; Machado, R. (2007) Qualidade da Forragem de Alfafa na Região Sudeste do Brasil. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 44, Jaboticabal, SP. Anais... Jaboticabal: SBZ: Unesp, p.1-3.
- Ruggieri, A. C.; Almeida, A. R. P.; Camargo, G. M. F.; Rassini, J. B.; Botrel, M. A. (2001) Avaliação de cultivares de alfafa em diferentes regiões do estado de são

Paulo. Disponível em: <http://www.iz.agricultura.sp.gov.br/pdfs/1308750822.pdf>.
Acesso em: 03/06/2017.

Sant'anna, I. C. Redes Neurais Artificiais na discriminação de populações de retrocruzamento com diferentes graus de similaridade. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 115p, 2014.

Silva, M. G. M.; Viana, A. P.; Amaral Júnior, A. T. A., Gonçalves, L. S. A.; Reis, R. V. (2012) Biometria aplicada ao melhoramento intrapopulacional do maracujazeiro amarelo. *Rev. Ciênc. Agron.*, 43:493-499.

Tai, G. C. C. (1977) Index selection with desired gains. *Crop Sci.*, 17:182-183.

Vasconcelos, E. S.; Ferreira, R. P.; Cruz, C. D.; Moreira, A.; et al. (2010) Estimativas de ganho genético por diferentes critérios de seleção em genótipos de alfafa. *Rev. Ceres*, 57:205-210.

CAPÍTULO 2

REDES NEURAS ARTIFICIAIS NO MELHORAMENTO GENÉTICO DA ALFAFA VISANDO SELECIONAR GENÓTIPOS POR MEIO DE CARACTERÍSTICAS COMPLEXAS

RESUMO

A seleção simultânea de caracteres na cultura da alfafa pode ser facilitada pela utilização de técnicas de inteligência computacional, como por exemplo, as Redes Neurais artificiais (RNAs). Estas podem ser treinadas para identificar e ponderar pesos dos caracteres, além de generalizar informações para novos conjuntos de dados. Assim, o objetivo desse trabalho foi propor RNAs para serem utilizadas na seleção de genótipos de alfafa. Foram utilizados dados de quatro cortes da forrageira, com 77 genótipos cada. Os caracteres avaliados quanto à produção e valor nutritivo, foram subdivididos em classes segundo os escores do índice de Taí. Para garantir a capacidade de generalização das RNAs, os dados dos grupos produção e valor nutritivo, foram ampliados de modo que os dados originais de cada classe do índice gerassem 50 novas observações. As redes foram treinadas e validadas com os dados ampliados e posteriormente foram utilizadas para classificar os dados reais dos quatro cortes. As taxas de erros das redes de produção e valor nutritivo atingiram valores de no máximo 15%. Além disso, a rede foi mais eficiente para classificar genótipos quanto ao valor nutritivo do que quanto à produção. De modo geral, houve altas taxas de concordância com o índice quanto a classificação dos genótipos em cada grupo de caracteres. Mesmo nos cortes onde a rede apresentou pior desempenho foi possível observar a potencialidade da mesma para classificação de genótipos de alfafa. Para as redes obtidas neste trabalho, o treinamento ocorreu a partir do fornecimento de exemplos de cortes obtidos em diferentes épocas do ano e isso foi suficiente para que a rede conseguisse ótimo desempenho em todos os cortes. Assim, estas poderão ser utilizadas em futuros cortes realizados pela Embrapa Pecuária Sudeste para classificação de genótipos do programa de melhoramento da cultura, quando à produção e valor nutritivo, auxiliando os pesquisadores na tomada de decisão.

1. INTRODUÇÃO

Os programas de melhoramento da alfafa visam desenvolver cultivares de alta produtividade e com alto valor nutritivo. Como a seleção na alfafa é predominantemente fenotípica e a maioria dos caracteres de importância têm baixa herdabilidade, a acurácia do processo seletivo pode ser dificultada (Li e Brummer, 2012). Além disso, o melhoramento é feito com base em uma série de caracteres importantes, requerendo adoção de técnicas que maximizem a acurácia seletiva.

Atualmente, a estratégia de seleção mais utilizada em alfafa é a seleção simultânea (Basigalup e Odorizzi, 2011), onde seleciona-se tanto para os caracteres de produção quanto para os componentes do valor nutritivo. Entretanto, a seleção simultânea pode ser dificultada pelo grande volume de informações geradas nos programas de melhoramento da forrageira, pela dificuldade do próprio pesquisador em estabelecer pesos adequados e pela qualidade requerida das matrizes de variância e covariância para estimação de um bom índice. Assim, a aplicação de técnicas convencionais de seleção pode tornar o processo de melhoramento da alfafa moroso.

A utilização de técnicas de inteligência computacional, como por exemplo, as Redes Neurais artificiais (RNAs), pode ser útil para otimizar a seleção simultânea de caracteres uma vez que a rede pode aprender a classificar genótipos da mesma forma que é feito no índice, não necessitando, porém, de informações muito detalhadas sobre os genótipos. As RNAs são técnicas de aprendizado baseadas em modelos de natureza biológica inspirados no funcionamento do cérebro humano (Norvig e Russel, 2008). O cérebro humano apresenta capacidade de processar informações complexas e não-lineares de forma muito superior à de qualquer máquina e, devido à plasticidade do sistema nervoso, a rede de neurônios biológicos pode se adaptar às novas situações a que é exposta (Haykin, 2001).

Devido a sua estrutura não linear as RNAs conseguem alta eficiência trabalhando com caracteres complexos (Galvão et al., 1999), sendo capazes de capturar relações entre variáveis que não são captadas pelos modelos estocásticos, podendo ainda generalizar as informações para novos casos (Mackay, 1994). Somado a isso, os avanços obtidos na área da informática têm possibilitado a realização de

trabalhos de RNAs cada vez mais abrangentes pela geração, *in silico*, de grandes volumes de dados, através da simulação (Sant'anna, 2014).

O primeiro modelo de neurônio artificial data de 1943 e foi proposto por Warren McCulloch e Walter Pitts (McCulloch & Pitts, 1943). O neurônio artificial de McCulloch e Pitts representa uma simplificação do neurônio biológico (Braga et al., 2007), onde toda a transmissão da informação é controlada pela função de ativação, que permite ou não a ativação da saída. Devido às particularidades do modelo, o neurônio artificial de McCulloch e Pitts só conseguia obter solução para problemas linearmente separáveis, além disso pesos negativos eram melhores para representar sinais inibitórios e os pesos eram não ajustáveis (Braga et al., 2007).

Desde a proposição de McCulloch e Pitts em 1943, vários outros modelos mais dinâmicos foram desenvolvidos (Braga et al., 2007), como por exemplo a Perceptron Multicamadas no ano de 1986, que utiliza o algoritmo de treinamento chamado de “algoritmo de retropropagação de erro” (Haykin, 2001). As redes Perceptron Multicamadas (MLP) são amplamente utilizadas e apresentam ótima eficiência para solucionar problemas complexos, graças à inclusão das camadas ocultas no modelo.

A rede MLP é caracterizada pela presença de uma ou mais camadas ocultas, alto grau de conectividade e pela presença de diferentes funções de ativação não-lineares (Peixoto, 2013). O conjunto dessas características garantem às redes MLP alto poder computacional (Silva, 2014). Os parâmetros da rede são definidos empiricamente e ajustados conforme o seu desempenho durante a etapa de treinamento.

A utilização de RNAs no melhoramento vegetal tem sido crescente e inclui, por exemplo, predição de produção de grãos de arroz (Ji et al., 2007), avaliação de adaptabilidade e estabilidade em genótipos de alfafa (Nascimento et al., 2013), seleção em cana-de-açúcar (Zhou et al., 2011; Brasileiro et al., 2015), previsão de características complexas em vacas Jersey e trigo (Gianola et al., 2011), dentre outros.

Durante o processo de seleção em alfafa são realizados muitos cortes para estimar a produção de matéria seca e o valor nutritivo da forragem nas diferentes estações de cultivo. Para cada corte são realizadas, além da coleta de dados, análises do conjunto de dados que demandam tempo do pesquisador. Para cada ciclo da

cultura, a seleção pode ser determinada por índices visando obter ganhos em todo o conjunto de caracteres.

Nesse contexto, a utilização de RNAs pode ser útil para classificar genótipos de alfafa baseando-se nas informações coletadas. A partir de um banco de dados que fosse constituído de experimentos bem conduzidos e que caracterizasse acessos em diferentes condições ambientais, um bom índice de seleção poderia ser ajustado e a partir dele, as RNAs poderiam aprender a identificar padrões e ponderar os caracteres dos genótipos tal qual é feito no índice, mas com a vantagem de trabalhar muito bem com informações incompletas e com ruído, tornando dispensável todo o rigor necessário para a condução de todos os experimentos e das análises estatísticas de novos cortes realizados, além de generalizar essas informações para os próximos cortes realizados.

Assim, o objetivo desse trabalho foi propor Redes Neurais Artificiais a partir de um índice de seleção estabelecido, para serem utilizadas na seleção simultânea de genótipos de alfafa do programa de melhoramento de alfafa da Embrapa Pecuária Sudeste, em relação aos caracteres de produção e valor nutritivo.

2. MATERIAL E MÉTODOS

2.1. Material vegetal e conjunto de dados

Foram utilizados dados de quatro cortes de alfafa realizados em 77 genótipos provenientes da Embrapa Pecuária Sudeste, localizada no município de São Carlos, latitude 22° 01' 03" S e longitude 47° 53' 27" W". Os cortes foram realizados nos meses de novembro de 2015, fevereiro, maio e agosto de 2016. Todos os cortes foram irrigados, exceto o corte 4, colhido sob déficit hídrico. O delineamento utilizado em cada experimento foi o de blocos casualizados, com três repetições, de acordo com a Equação 1:

$$Y_{ij} = \mu + G_i + B_j + e_{ij} \quad (\text{Equação 1})$$

Em que Y_{ij} é a observação referente ao tratamento i avaliado no bloco j ; μ é a média geral; G_i é o efeito do i -ésimo tratamento, com $i = 1, 2, \dots, I$; B_j é o efeito do j -ésimo bloco, com $j = 1, 2, \dots, J$; e e_{ij} é o erro experimental, tal que $e_{ij} \sim \text{NID}(0; \sigma^2)$.

Os caracteres avaliados foram: Altura de plantas dada pela média de cinco plantas em centímetros na área útil da parcela quando as plantas atingissem 10% de florescimento; Produção de matéria seca (kg/ha) medida pelos cortes feitos a uma altura de 8 a 10 cm do nível do solo, sempre que cada cultivar atingia 10% da floração ou quando as brotações basais atingissem altura média de 3 a 5 cm; Relação caule/folha: feita na matéria verde, durante os cortes de produção;

Além disso foram feitas análises de proteína bruta (PB); digestibilidade “*in vitro*” da matéria seca (DIVMS); fibra em detergente neutro (FDN); fibra em detergente ácido (FDA); lignina e suscetibilidade a doenças – estimada visualmente sendo a severidade determinada de acordo com a porcentagem de área foliar atacada em cada parcela, classificando as cultivares da seguinte maneira: a) nota 0 - susceptível (>50% de severidade da doença); b) nota 1 - baixa resistência (49 a 30% de severidade da doença); c) nota 2 - moderadamente resistentes (29 a 11% de severidade da doença); d) nota 3 - resistentes (< 11% de severidade da doença).

Primeiramente os caracteres avaliados foram subdivididos em dois grandes grupos: Produção e Valor Nutritivo. As variáveis do grupo “Produção” foram: altura de plantas, produção de matéria seca, relação caule/folha e suscetibilidade a doenças, já o grupo “Valor nutritivo” foi composto pelos caracteres: proteína bruta, lignina, digestibilidade “*in vitro*” de matéria seca, fibra em detergente neutro e fibra em detergente ácido.

Posteriormente os genótipos foram classificados conforme desempenho em produção e valor nutritivo segundo o índice de seleção de Taí (1977), que permite que caracteres secundários, para os quais não se deseja ganho máximo, sejam incluídos no índice para auxiliar no ganho dos caracteres principais.

Considerando que existam n caracteres avaliados e destes, m ($m \leq n$) são principais, o agregado genotípico (H) e o índice (I) são dados por:

$$H = a_1g_1 + a_2g_2 + \dots + a_mg_m$$
$$I = b_1x_1 + b_2x_2 + \dots + b_mx_m + b_{m+1}x_{m+1} + \dots + b_{m+1}x_{m+1}$$

em que:

a = pesos econômicos

g = valores genotípicos

x = valores fenotípicos

b = coeficiente dos caracteres no índice

O vetor de ganhos desejados pode ser estabelecido com base em um desvio padrão genético para cada caractere (Cruz et al., 2014).

O vetor b pode ser estimado a partir das equações 1 e 2:

$$P\hat{b} = Ga \quad (1)$$

$$G\hat{b} = \Delta g_d \quad (2)$$

em que:

G = matriz de covariâncias genotípicas entre as variáveis

P = matriz de covariâncias fenotípicas entre as variáveis

Δg_d = vetor de ganhos desejados

Substituindo 2 em 1:

$$a = GP^{-1}G \Delta g_d$$

Considerando que:

$$a = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad e \quad \Delta g_d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

em que:

c_1 e d_1 são vetores de dimensão $m \times 1$

c_2 e d_2 são vetores de dimensão $(n-m) \times 1$

Considerando ainda:

$$c_2 = \phi$$

$$\omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$$

em que:

ω_{11} , ω_{12} , ω_{21} e ω_{22} são matrizes de dimensão $m \times m$, $m \times (n-m)$, $(n-m) \times m$ e $(n-m) \times (n-m)$, respectivamente

Assim:

$$a = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} \omega_{11}d_1 + \omega_{12}d_2 \\ \omega_{21}d_1 + \omega_{22}d_2 \end{bmatrix}$$

e

$$\Delta g_d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ -\omega_{22}^{-1} \omega_{21}d_1 \end{bmatrix}$$

Assim, o vetor b pode ser estimado tanto por 1 quanto por 2.

Após a obtenção dos escores, os genótipos foram ordenados em ordem decrescente segundo o valor obtido no índice. A partir disso, foram determinados quatro subgrupos para produção e quatro para valor nutritivo, onde os genótipos foram classificados como ótimos, bons, medianos e ruins, conforme mostrado na Tabela 1:

Tabela 1 – Classificação dos genótipos de alfafa segundo Índice de Taí (1977), quanto ao desempenho em valor nutritivo e produção. Em que: 1 = ótimo, 2 = bom, 3 = mediano e 4 = ruim.

Genótipos	Valor Nutritivo				Produção			
	Corte 1	Corte 2	Corte 3	Corte 4	Corte 1	Corte 2	Corte 3	Corte 4
1	1	1	1	1	1	4	4	4
2	1	4	3	3	3	3	2	1
3	2	1	1	2	4	2	3	3
4	2	1	2	4	1	2	1	3
5	2	1	3	3	3	3	2	1
6	3	4	3	1	3	3	3	2
7	2	3	1	1	4	4	4	4
8	3	4	4	2	3	3	3	1
9	4	2	4	4	2	3	1	2

10	4	2	3	3	2	2	1	1
11	4	1	3	3	2	3	2	2
12	4	3	4	3	1	1	1	3
13	4	3	2	2	4	2	2	1
14	3	4	4	2	1	1	1	2
15	1	1	1	1	4	4	4	4
16	3	2	3	3	4	1	1	2
17	2	1	1	1	1	2	3	2
18	1	1	1	1	4	4	4	4
19	3	2	2	3	2	3	4	3
20	2	2	3	2	4	1	2	2
21	2	3	3	4	2	2	1	1
22	2	2	3	3	3	1	1	2
23	1	1	1	1	2	4	4	4
24	3	2	2	2	2	4	3	3
25	1	3	2	2	2	1	2	1
26	3	4	4	4	4	1	2	1
27	3	2	3	3	3	1	3	2
28	3	3	3	3	4	2	2	1
29	1	3	2	2	3	4	3	2
30	3	2	4	2	4	3	2	2
31	3	4	3	2	4	3	2	2
32	2	2	2	2	1	3	4	4
33	2	2	2	2	3	4	3	3
34	4	3	3	3	2	2	2	1
35	1	1	1	1	4	4	4	4
36	1	3	2	4	3	2	4	3
37	1	4	1	3	2	1	3	4
38	2	3	3	4	3	2	1	3
39	2	4	4	4	1	2	2	3
40	4	4	4	3	1	1	1	1
41	2	3	2	1	2	1	2	2
42	3	3	3	2	1	2	2	2
43	1	4	2	1	3	2	2	3
44	3	2	1	2	1	3	3	4
45	4	3	4	4	3	3	1	3
46	1	4	2	3	4	1	3	3
47	2	1	1	2	2	4	3	4
48	4	3	2	1	1	4	2	3
49	1	1	1	1	4	2	4	4
50	1	1	1	1	3	4	4	4
51	1	1	2	2	3	4	4	3
52	4	3	2	2	4	3	3	2
53	4	2	2	1	2	4	4	4
54	3	2	3	3	1	3	2	2
55	1	1	1	1	4	4	4	4

56	3	1	2	4	1	1	3	2
57	2	3	3	4	1	3	3	1
58	2	1	1	1	3	4	4	3
59	1	1	1	1	2	3	4	3
60	4	2	4	4	1	2	1	1
61	3	3	4	4	1	1	1	1
62	4	4	4	4	4	3	1	1
63	4	4	3	4	3	1	3	3
64	3	4	4	4	1	1	1	2
65	3	1	1	2	3	4	3	3
66	4	4	4	4	2	1	1	1
67	1	2	2	4	1	2	2	4
68	4	2	1	1	1	3	4	4
69	1	2	1	1	4	4	4	4
70	2	1	1	1	1	4	4	4
71	1	3	2	2	3	2	2	4
72	2	4	4	3	2	2	1	1
73	4	3	4	3	3	3	3	3
74	2	2	3	4	4	1	3	2
75	4	4	4	3	2	2	1	1
76	3	4	4	3	2	1	1	1
77	4	4	4	4	2	1	1	1

Visando obter uma rede com boa capacidade de generalização, os dados dos grupos produção e valor nutritivo, considerando a classificação de acordo com o índice de Taí (1977) em cada corte, foram ampliados de modo que os dados originais de cada classe do índice gerassem 50 novas observações, originando um total de 800 novos genótipos para produção e 800 para valor nutritivo, como apresentado na Tabela 2:

Tabela 2 – Esquema da ampliação dos dados dos genótipos de alfafa para escolha das redes para produção e valor nutritivo.

Grupo	Cortes	Classificação	Ampliação	Grupo	Cortes	Classificação	Ampliação
Produção	C1	1	50	Valor Nutritivo	C1	1	50
		2	50			2	50
		3	50			3	50
		4	50			4	50
	C2	1	50		C2	1	50
		2	50			2	50
		3	50			3	50
		4	50			4	50
	C3	1	50		C3	1	50

	2	50		2	50
	3	50		3	50
	4	50		4	50
	1	50		1	50
C4	2	50	C4	2	50
	3	50		3	50
	4	50		4	50

O processo de ampliação consistiu em gerar, a partir dos dados reais, uma variável aleatória com distribuição normal com uma matriz de zeros representando a média e variância equivalente a uma matriz sigma ($X \sim N(\phi, \Sigma)$) por meio da transformação linear $X = F'Y$, em que $Y \sim N(\phi, I)$ representava a variável aleatória correspondente aos dados originais e F é obtido pela decomposição espectral de Σ , tal que $\Sigma^{-1} = FF'$. Para ampliar os dados originais, novos valores de Y foram gerados considerando $Y \sim N(\phi, (F')^{-1}X)$. Para garantir a eficiência do processo, a pressuposição de que a distribuição dos dados deve ser conhecida, com média zero e variância igual a V deve ser atendida. Para isso, utilizou-se o teorema de Box-Muller, em que as variáveis z_1 e z_2 têm distribuição normal com média zero e variância V :

$$Z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2) \quad \text{e} \quad Z_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

em que U_1 e U_2 são números aleatórios gerados pela distribuição uniforme entre 0 e 1.

Os valores foram simulados conservando as médias, variâncias e covariâncias dos dados originais de cada um dos cortes separadamente. Assim, o conjunto de dados foi subdividido em 80% para treinamento da rede e 20% para validação. A ampliação foi realizada pelo programa GENES (Cruz, 2013).

2.2. Modelagem da Rede Neural Artificial

A modelagem via RNAs necessita de dois conjuntos de dados: um para treinamento e outro para validação. Neste trabalho, as redes foram treinadas e validadas com os dados ampliados e posteriormente foram utilizadas para classificar os dados reais dos quatro cortes. A arquitetura de rede utilizada para classificar os genótipos foi a Perceptron Multicamadas.

Foram estabelecidas duas conformações de RNA, uma para os caracteres do grupo produção e outra para os caracteres do grupo valor nutritivo. Utilizou-se quatro e cinco entradas, correspondentes aos caracteres dos grupos produção e valor nutritivo, respectivamente. Foram testadas as funções de ativação tangente hiperbólica (tansig) e logística (logsig). Para a camada de saída foi utilizada a função limiar (purelin), além disso, foi utilizado o algoritmo de treinamento *trainbr* – *Backpropagation* com um número de épocas igual a 7000.

A quantidade de neurônios nas camadas ocultas variou entre três e seis e a camada de saída foi composta por um único neurônio, representado por um vetor de elementos conhecidos, uma vez que o processo de aprendizado neste caso, é supervisionado. Após testes das várias possibilidades de topologias, foram escolhidas as redes que apresentaram a menores taxas de erro aparente. As RNAs foram processadas pelo software Matlab (Mathworks, 2012) a partir de scripts do aplicativo GENES (Cruz, 2013).

3. RESULTADOS E CONCLUSÕES

Topologia de rede para os grupos de caracteres

A melhor topologia de rede para cada grupo de caracteres foi escolhida considerando 5832 possibilidades, adotando como critério, a taxa de erro aparente. Dos 800 genótipos gerados via ampliação dos dados, 640 foram utilizados para treinamento da rede e 160 para validação. Dessa forma, para o grupo de caracteres produção a topologia utilizada incluiu seis neurônios na primeira camada oculta, três na segunda e seis na terceira e as funções de ativação foram tansig, tansig e logsig, respectivamente (Figura 1):

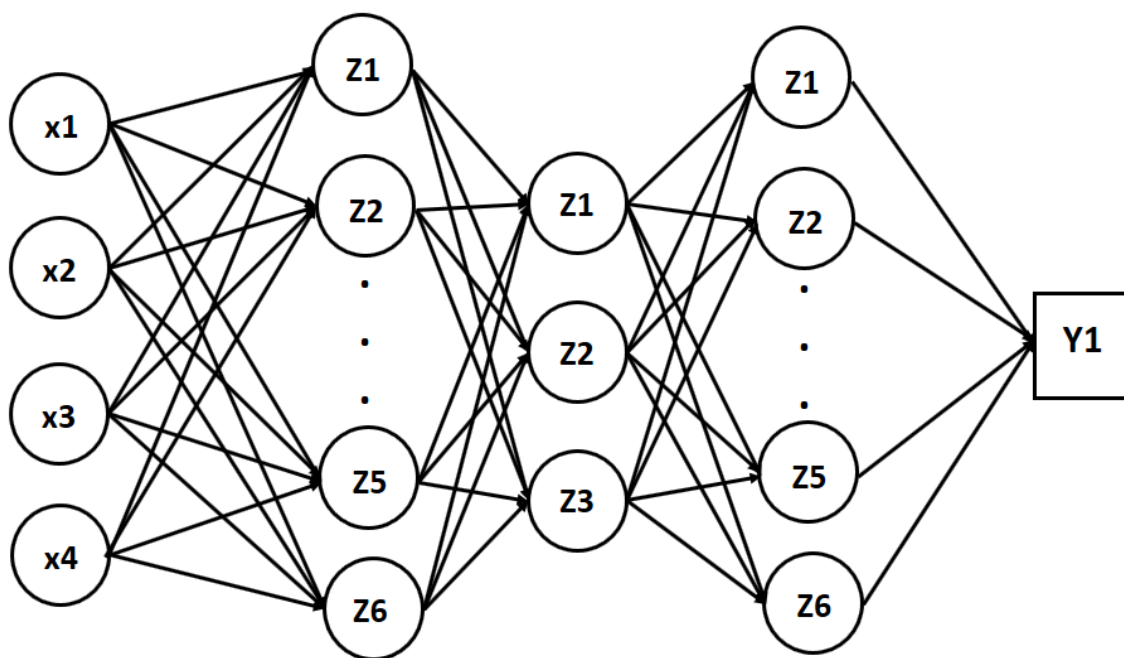


Figura 1 – Arquitetura da RNA para produção. Entradas de (x_1) a (x_4) representando cada caractere de produção. Três camadas ocultas compostas por seis, seis e três neurônios, com funções de ativação logística em todas as camadas. Na camada de saída, Y_1 representa a classificação que cada indivíduo obteve.

Para o grupo de caracteres valor nutritivo a topologia utilizada incluiu seis neurônios na primeira camada oculta, seis na segunda e três na terceira e a função de ativação para todas elas foi a logsig (Figura 2):

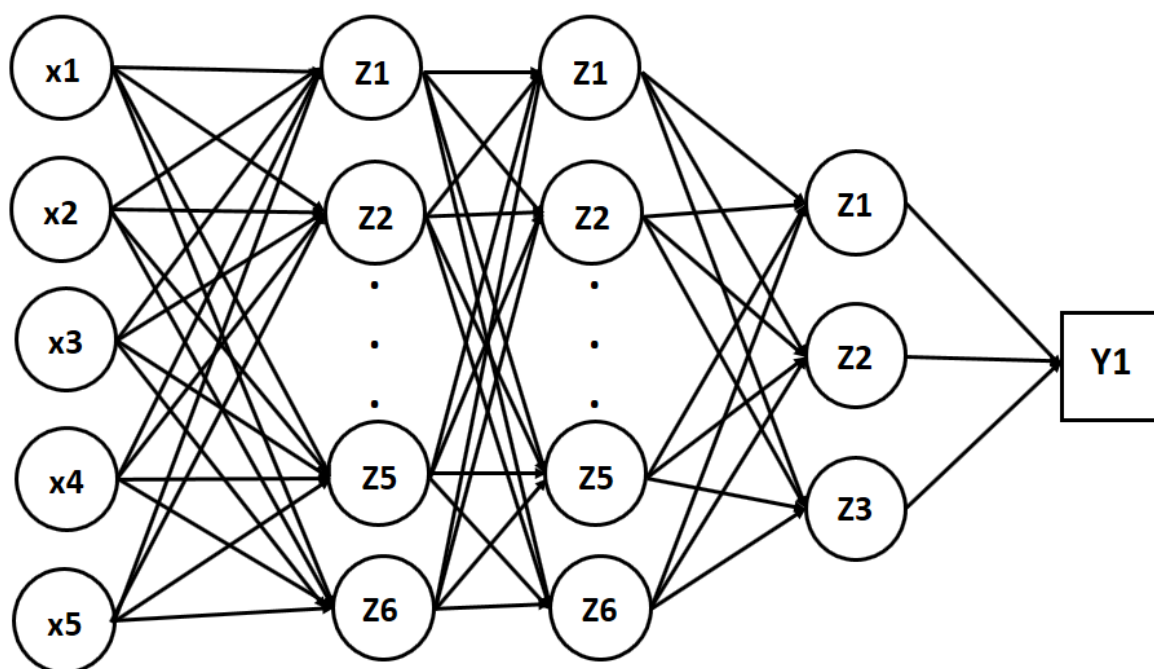


Figura 2 – Arquitetura da RNA para valor nutritivo. Entradas de (x_1) a (x_5) representando cada caractere de valor nutritivo. Três camadas ocultas compostas por seis, seis e três neurônios, com funções de ativação logística em todas as camadas. Na camada de saída, Y_1 representa a classificação que cada indivíduo obteve.

Na Tabela 3 são apresentados os resultados da RNA para cada grupo de caracteres. Em ambas as situações, as taxas de erros atingiram valores de no máximo 15%, além disso observou-se as maiores taxas de erro na validação do que no treinamento. Esses resultados são compatíveis com os resultados relatados por outros autores como por exemplo Carneiro (2015) e Sant’anna (2014) em estudos classificatórios, reafirmando o grande potencial de generalização das redes neurais como mencionado por Braga (2011). Além disso, a rede foi mais eficiente para classificar genótipos quanto ao valor nutritivo do que quanto à produção, mesmo apresentando topologias de semelhante complexidade, isso pode ser ocorrido devido às características do próprio conjunto de dados, mais distinguíveis quanto ao valor nutritivo do que quanto à produção.

Tabela 3 – Taxas de erro aparente, obtidas para os grupos de caracteres produção e valor nutritivo.

Procedimentos	Taxa de erro aparente – TEA (%)
---------------	---------------------------------

	Validação	Treinamento
RNA Produção	15,000	11,875
RNA Valor Nutritivo	11,875	6,563

Eficiência da utilização da rede neural artificial para classificação de genótipos de alfafa

As redes obtidas para os grupos de caracteres considerados neste estudo foram utilizadas para classificar os dados reais dos 77 genótipos avaliados em cada corte. Dessa forma, foram obtidas as tabelas de confusão para cada grupo de caracteres. Considerando a classificação dos genótipos quanto aos caracteres de produção (Tabela 4) houve altas taxas de concordância com o índice, como pode ser notado por exemplo, no corte 1, onde o menor percentual de acerto foi de 73,68%, neste mesmo corte a RNA classificou corretamente todos os indivíduos do grupo 1 (genótipos ótimos). No corte 4, onde os genótipos foram mantidos sob déficit hídrico, a rede acertou 94,74% dos genótipos ruins (grupo 4) e 78,95% dos genótipos medianos (grupo 3), observa-se que mesmo que as porcentagens de acertos nas classes ótima e boa tenham sido menores (55% e 57,89%, respectivamente), o alto índice de acertos nas classes ruim e mediana, garante a eficiência da seleção, pela eliminação dos genótipos inferiores. Nos demais cortes, as porcentagens de acerto da rede foram menores, no entanto dentre as classificações erradas, nota-se uma tendência de alocação dos indivíduos nas classes mais próximas, como ocorreu no corte 3, onde a rede acertou 50% das classificações do grupo 1 (desempenho ótimo) e 45% dos genótipos ótimos (que não haviam entrado no grupo 1) foram classificados como bons (grupo 2).

Tabela 4 – Porcentagem de acerto na classificação dos genótipos segundo a Rede Neural Artificial obtida para o grupo de caracteres de produção, baseada no índice de Taí (1977).

Corte	Classificação índice de Taí	Classificação Rede Neural Artificial			
		1	2	3	4
C1	1	20 (100%)*	0	0	0
	2	3 (15,79%)	14 (73,68%)	2 (10,53%)	0

	3	0	4 (21,05%)	15 (78,95%)	0
	4	0	0	2 (10,53%)	16 (84,21%)
C2	1	17 (85,00%)	0	2 (10,00%)	1 (5,00%)
	2	3 (15,79%)	13 (68,43%)	1 (5,26%)	2 (10,53%)
	3	0	6 (31,58%)	11 (57,89%)	2 (10,53%)
	4	1 (5,26%)	3 (15,79%)	2 (10,53%)	13 (68,43%)
	1	10 (50,00%)	9 (45,00%)	1 (5,00%)	0
C3	2	1 (5,26%)	10 (52,63%)	7 (36,84%)	1 (5,26%)
	3	0	3 (15,79%)	13 (68,42%)	3 (15,79%)
	4	0	0	1 (5,26%)	18 (94,74%)
	1	11 (55,00%)	8 (40,00%)	1 (5,26%)	0
C4	2	3 (15,79%)	11 (57,89%)	4 (21,05%)	1 (5,26%)
	3	0	2 (10,53%)	15 (78,95%)	2 (10,53%)
	4	0	0	1 (5,26%)	18 (94,74%)

* Porcentagem de acerto da RNA.

A análise do desempenho da RNA para classificação dos genótipos quanto aos caracteres de valor nutritivo é encontrada na Tabela 5. De modo geral, houve altas taxas de concordância com o índice assim como observado para a RNA de produção. Nos cortes 1, a RNA não conseguiu classificar dois genótipos e no corte 2, cinco. No entanto, esse fato não reduziu o poder de generalização da rede para os demais cenários. A taxa de acerto para o grupo de genótipos ótimos nos cortes 1 e 2 foram 75% e 90%, respectivamente. No corte 4, a rede acertou a classificação de 100% dos genótipos ótimos, 78,95% dos genótipos bons, 94,74% dos genótipos medianos e 89,47% dos genótipos considerados ruins. A eficiência da rede foi alta para discriminar genótipos ruins em todos os cortes, apresentando taxa mínima de acerto de 84,21%

no corte 2. O pior desempenho da rede aconteceu no corte 3, à semelhança das classificações da RNA de produção a alocação dos indivíduos que tiveram classificação errada aconteceu nas classes mais próximas.

Tabela 5 – Porcentagem de acerto na classificação dos genótipos segundo a Rede Neural Artificial obtida para o grupo de caracteres de valor nutritivo, baseada no índice de Taí (1977)

Corte	Classificação índice de Taí	Classificação Rede Neural Artificial			
		1	2	3	4
C1	1	15 (75,00%)*	1 (5,00%)	2 (10,00%)	0
	2	2 (10,53%)	12 (63,16%)	2 (10,53%)	3 (15,79%)
	3	0	1 (5,26%)	13 (68,43%)	5 (26,31%)
	4	0	0	1 (5,26%)	18 (94,74%)
C2	1	18 (90,00%)	0	0	0
	2	4 (21,05%)	10 (52,63%)	4 (21,05%)	0
	3	0	5 (26,31%)	11 (57,89%)	2 (10,52%)
	4	0	0	2 (10,53%)	16 (84,21%)
C3	1	10 (50%)	9 (45,00%)	1 (5,00%)	0
	2	1 (5,26%)	10 (52,63%)	7 (36,84%)	1 (5,26%)
	3	0	3 (15,79%)	13 (68,42%)	3 (15,79%)
	4	0	0	1 (5,26%)	18 (94,74%)
C4	1	20 (100%)	0	0	0
	2	3 (15,79%)	15 (78,95%)	1 (5,26%)	0
	3	0	0	18 (5,26%)	1 (5,26%)
	4	0	0	2 (10,53%)	17 (89,47%)

* Porcentagem de acerto da RNA.

Mesmo nos cortes onde a rede apresentou pior desempenho (quando comparado aos demais) foi possível observar a potencialidade da mesma para classificação de genótipos de alfafa. Resultados satisfatórios de performance das RNAs em análises classificatórias têm sido relatados na literatura por vários autores, como por exemplo Barbosa et al. (2011), que trabalhando com oito características de mamão papaia (*Carica papaya* L.), avaliaram o desempenho da rede em comparação com a análise discriminante de Anderson. Nesse estudo, as RNAs apresentaram taxa de acerto 94,44%, superior àquela obtida pela análise discriminante, os autores então concluíram que as RNAs são eficientes para estudos classificatórios. Conclusão semelhante já havia sido relatada por Braga et al. (2007) que reconhecem a superioridade das RNAs em relação aos modelos estocásticos. Para a cultura da alfafa, o sucesso dos estudos de natureza classificatória utilizando a abordagem das redes neurais foram relatados por Nascimento et al. (2013) e Barroso et al. (2013), ao comparar a eficiência de metodologias de adaptabilidade e estabilidade genotípica com as RNAs. Barroso et al. (2013) ainda destaca como uma desvantagem das redes em algumas situações, o tempo gasto com o processamento de dados e o desconhecimento da influência de cada característica sobre a variável resposta liberada pela RNA.

Tendo em vista que os programas de melhoramento da alfafa visam desenvolver cultivares de alta produtividade e com alto valor nutritivo, que o controle genético da maioria dos caracteres avaliados neste estudo é poligênico e que a acurácia do processo seletivo é altamente influenciada por esses fatores (Li e Brummer, 2012), a utilização de índices de seleção torna-se uma ferramenta muito útil para realizar seleção de genótipos da cultura da alfafa. Uma vez que o pesquisador estabeleça um bom índice que pondere bem os pesos dos caracteres, é possível utilizar a rede neural com sucesso para automatizar a seleção nos programas de melhoramento da cultura. A informação obtida *a priori* pelo índice pode ser utilizada para treinar uma rede que seja capaz de classificar novas observações com a eficiência exigida no processo seletivo. Para isso, é fundamental que haja um treinamento eficiente da rede, fornecendo exemplos que caracterizem diversas situações.

Para as redes obtidas neste trabalho, o treinamento ocorreu a partir do fornecimento de exemplos de cortes obtidos em diferentes épocas do ano e isso foi suficiente para que a rede conseguisse ótimo desempenho em todos os cortes. Assim,

estas poderão ser utilizadas em futuros cortes realizados pela Embrapa Pecuária Sudeste para classificação de genótipos do programa de melhoramento da cultura, quando à produção e valor nutritivo, auxiliando os pesquisadores na tomada de decisão. Devido ao dinamismo e plasticidade das RNAs, estas são capazes de aperfeiçoar o seu desempenho por meio da apresentação de novos exemplos para treinamento (Haykin, 2001). As redes propostas neste trabalho estão disponíveis em <ftp://ftp.ufv.br/dbg/RNA/>.

4. CONCLUSÕES

A abordagem das redes neurais artificiais constitui uma alternativa eficiente para estudos de natureza classificatória;

As redes apresentam alta eficiência para utilização em programas de melhoramento da cultura da alfafa.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Barbosa, C. D.; Viana, A. P.; Quintal, S. S. R.; Pereira, M. G. (2011) Artificial neural network analysis of genetic diversity in *Cariacica papaya* L. *Crop Breeding and Applied Biotechnology*. 11:224-231.
- Barroso, L. M. A.; Nascimento, M.; Nascimento, A. C. C.; Silva, F. F.; Ferreira, R. P. (2013) Uso do método de Eberhart e Russell como informação a priori para aplicação de redes neurais artificiais e análise discriminante visando a classificação de genótipos de alfafa quanto à adaptabilidade e estabilidade fenotípica. *Ver. Bras. Biom.*, 31:176-188.
- Basigalup, D.H.; Odorizzi, A. S. Melhoramento genético da alfafa. In: Ferreira, R. P.; Basigalup, D. H.; Gioco, J. O. (Ed.). *Melhoramento Genético da Alfafa*. São Carlos: Embrapa Pecuária Sudeste, 2011. p. 225-260.
- Braga, A. P.; Ferreira, A. C. P. L.; Ludermir, T. B. (2011) Redes Neurais Artificiais: teoria e aplicações (2nd ed.). Rio de Janeiro, LTC, 262p.
- Brasileiro, B. P. et al. (2015) Selection in sugarcane families with artificial neural networks. *Crop Breed. Appl. Biotechnol.*, 5:72-78.
- Carneiro, V. Q. (2015) Rede neural e lógica Fuzzy aplicadas no melhoramento do feijoeiro. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 108p, 2015.
- Cruz, C. D. (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum.*, 35:271-276.
- Galvão, C. O.; Valença, M. J. S.; Vieira, V. P. P. B.; Diniz, L. S. et al. (1999) Sistemas inteligentes: Aplicações a recursos hídricos e ciências ambientais. UFRGS: ABRH.

- Gianola, D.; Okut, H.; Kent A. W.; Rosa, J. M. R. (2011) Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genetics*, 45:34.
- Haykin, S. (2008) *Neural Networks and Learning Machines* (3rd ed.). Hamilton: Pearson – Prentice Hall, 906p.
- Ji, B.; Sun, Y.; Yang, S. and Wan, J. (2007) Artificial neural networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science*, 145:249-261.
- Li, X.; Brummer, E. C. (2012) Applied Genetics and Genomics in Alfalfa Breeding. *Agronomy*, 2:40-61.
- Mackay, D. J. C. (1994) Bayesian non-linear modelling for the prediction competition. In: ASHRAE Transactions, ASHRAE, Atlanta Georgia. 100:1053-1062.
- MathWorks (2012). MATLAB – The Language of Technical Computing. Disponível em: <http://www.mathworks.com/products/matlab>.
- McCulloch, W. S.; Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol.* 5:115-133.
- Nascimento, M.; Peternelli, L. A.; Crus, C. D.; Nascimento, A. C. C.; Ferreira, R. P.; Bhering, L. L. and Sangado, C. C. (2013) Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. *Crop Breeding and Applied Biotechnology*, 13:152-156.
- Norvig, P. and Russell, S. (2013) *Inteligência Artificial*. 3rd ed. CAMPOS, Rio de Janeiro, Elsevier, 958p.
- Peixoto, L. A. *Redes Neurais Artificiais na predição do valor genético*. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 97p, 2013.

Sant'anna, I. C. Redes Neurais Artificiais na discriminação de populações de retrocruzamento com diferentes graus de similaridade. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 115p, 2014.

Silva, G. N. Redes Neurais Artificiais: Novo paradigma para a predição de valores genéticos. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, 92p, 2014.

Tai, G. C. C. (1977) Index selection with desired gains. *Crop Sci.*, 17:182-183.

CAPÍTULO 3

EFICIÊNCIA DA UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS PARA CLASSIFICAÇÃO DE GENÓTIPOS QUANTO A CARACTERES DE PRODUÇÃO E VALOR NUTRITIVO

RESUMO

A utilização das Redes Neurais Artificiais (RNAs) pode ser útil para otimizar a seleção simultânea de caracteres, que tem sido a estratégia mais utilizada no melhoramento da alfafa. Entretanto, a eficiência de generalização das RNAs é dependente da qualidade do conjunto de dados para treinamento e validação. Assim, o objetivo desse trabalho foi avaliar eficiência das redes neurais treinadas a partir de dados ampliados de poucos cortes de alfafa para classificação de um novo conjunto de dados, bem como inferir sobre a importância do tamanho e diversidade do conjunto de dados de treinamento para generalização de informações. Os 77 genótipos avaliados foram previamente classificados segundo os escores do índice de Taí para os caracteres de produção e valor nutritivo. Assim, foram estabelecidas diferentes topologias de redes para cada grupo de caracteres, que foram treinadas com dados ampliados de três cortes e posteriormente, utilizadas para classificar dados reais de um novo corte. Considerando a classificação dos genótipos quanto aos caracteres de produção, observou-se altas taxas de erro aparente tanto no treinamento quanto na validação. Para o grupo de caracteres de valor nutritivo foram observadas pequenas taxas de erro tanto no treinamento quanto na validação. Entretanto as baixas taxas de erro não se mantiveram na utilização com um novo conjunto de dados, indicando a perda da habilidade de generalização, possivelmente pelo *overfitting* que pode ter ocorrido. O baixo desempenho das redes pode ser explicado uma vez que os limites mínimos das variáveis no conjunto de treinamento estão bem aquém daqueles encontrados no conjunto de teste e por isso, atribuir o insucesso de generalização por deficiência da modelagem pode ser um engano, pois não se deve esperar aprendizado em analogia ao cérebro humano, de situações tão díspares. Dessa forma, conclui-se que o uso de redes neurais será mais difundido à medida que haja mudança no paradigma da experimentação em que as decisões passarão a ser tomadas em bancos de dados, construídos ao longo do tempo de um programa de melhoramento, que constituirão de base para o processamento de dados.

1. INTRODUÇÃO

O melhoramento genético da alfafa visa o desenvolvimento de cultivares de alta produtividade e com alto valor nutritivo. Como a seleção nesta cultura é predominantemente fenotípica e a maioria dos caracteres de importância têm baixa herdabilidade, a acurácia do processo seletivo pode ser dificultada (Li e Brummer, 2012). Dessa forma, a utilização de técnicas de inteligência computacional, como por exemplo, as Redes Neurais artificiais (RNAs), pode ser útil para otimizar a seleção simultânea, que tem sido a estratégia mais utilizada no melhoramento da forrageira (Basigalup e Odorizzi, 2011).

As RNAs são técnicas de aprendizado baseadas em modelos de natureza biológica inspirados no funcionamento do cérebro humano (Norvig e Russel, 2008). O cérebro humano apresenta capacidade de processar informação complexas e não-lineares de forma muito superior à de qualquer máquina e, devido à plasticidade do sistema nervoso, a rede de neurônios biológicos pode se adaptar às novas situações a que é exposta (Haykin, 2001).

Por apresentarem estrutura não linear, as RNAs conseguem alta eficiência trabalhando com caracteres complexos (Galvão et al., 1999), sendo capazes de capturar relações entre variáveis que não são captadas pelos modelos estocásticos, podendo ainda generalizar as informações para novos casos (Mackay, 1994). Entretanto, a eficiência de generalização das RNAs é dependente da qualidade do conjunto de dados para treinamento e validação, além da correta escolha da topologia a ser utilizada. Estes são os principais problemas relatados quanto à utilização das redes neurais (De Veaux et al., 1997).

Segundo Kominakis et al. (2002) quanto maior o conjunto de dados de treinamento de uma rede, mais o desempenho desta pode ser aumentado, já que a diversificação das informações apresentadas garante o reconhecimento de padrões futuros. Dessa forma, é possível dizer que a capacidade de generalização de uma RNA está diretamente ligada ao tamanho do conjunto de dados de treinamento.

De modo geral, os trabalhos disponíveis na literatura dispõem de séries históricas de dados para treinamento das RNAs, como mostrado no trabalho de Dantas et al., 2016, onde a utilização das redes neurais visava a previsão da precipitação de períodos chuvosos. Esses autores utilizaram dados coletados durante

38 anos e encontraram resultados satisfatórios para descrever as variações na precipitação ao longo dos anos. No melhoramento genético, a obtenção de grandes séries históricas de dados é um pouco mais incomum, se comparado às áreas médica e da engenharia, por exemplo. Entretanto, a obtenção de um bom conjunto de dados aliada aos avanços obtidos na área da informática tem possibilitado a realização de trabalhos de RNAs cada vez mais abrangentes pela geração, *in silico*, de grandes volumes de dados, através da simulação (Sant'anna, 2014).

Durante o processo de seleção em alfafa são realizados muitos cortes para estimar a produção de matéria seca e o valor nutritivo da forragem nas diferentes estações de cultivo. Para cada ciclo da cultura, a seleção pode ser determinada por índices visando obter ganhos em todo o conjunto de caracteres. Nesse contexto, as informações coletadas nos cortes da forrageira ao longo do tempo podem ser reunidas em um banco de dados e este pode ser utilizado para treinar uma RNA. A eficiência de generalização da RNA será tanto maior quanto mais diversificadas forem as informações disponíveis para treinamento.

Assim, o objetivo desse trabalho foi avaliar eficiência das redes neurais treinadas a partir de dados ampliados de poucos cortes de alfafa para classificação de um novo conjunto de dados, bem como inferir sobre a importância do tamanho e diversidade do conjunto de dados de treinamento para generalização de informações.

2. MATERIAL E MÉTODOS

2.1. Material vegetal e conjunto de dados

Para este estudo, foram utilizados dados de 77 genótipos de alfafa provenientes da Embrapa Pecuária Sudeste, avaliados em quatro cortes realizados nos meses de novembro de 2015, fevereiro, maio e agosto de 2016. Todos os cortes foram irrigados, exceto o corte 4, colhido sob déficit hídrico. O delineamento utilizado em cada experimento foi o de blocos casualizados, com três repetições, de acordo com a Equação 1:

$$Y_{ij} = \mu + G_i + B_j + e_{ij} \quad (\text{Equação 1})$$

Em que Y_{ij} é a observação referente ao tratamento i avaliado no bloco j ; μ é a média geral; G_i é o efeito do i -ésimo tratamento, com $i = 1, 2, \dots, I$; B_j é o efeito do j -ésimo bloco, com $j = 1, 2, \dots, J$; e e_{ij} é o erro experimental, tal que $e_{ij} \sim \text{NID}(0; \sigma^2)$.

Os caracteres avaliados foram: Altura de plantas dada pela média de cinco plantas em centímetros na área útil da parcela quando as plantas atingissem 10% de florescimento; Produção de matéria seca (kg/ha) medida pelos cortes feitos a uma altura de 8 a 10 cm do nível do solo, sempre que cada cultivar atingia 10% da floração ou quando as brotações basais atingissem altura média de 3 a 5 cm; Relação caule/folha: feita na matéria verde, durante os cortes de produção.

Além disso foram feitas análises de proteína bruta (PB); digestibilidade “*in vitro*” da matéria seca (DIVMS); fibra em detergente neutro (FDN); fibra em detergente ácido (FDA); lignina e suscetibilidade a doenças – estimada visualmente sendo a severidade determinada de acordo com a porcentagem de área foliar atacada em cada parcela, classificando as cultivares da seguinte maneira: a) nota 0 - susceptível (>50% de severidade da doença); b) nota 1 - baixa resistência (49 a 30% de severidade da doença); c) nota 2 - moderadamente resistentes (29 a 11% de severidade da doença); d) nota 3 - resistentes (< 11% de severidade da doença).

Os caracteres avaliados foram divididos em dois grupos: Produção e Valor Nutritivo. As variáveis do grupo “Produção” foram: altura de plantas, produção de matéria seca, relação caule/folha e suscetibilidade a doenças, já o grupo “Valor nutritivo” foi composto pelos caracteres: proteína bruta, lignina, digestibilidade “*in vitro*” de matéria seca, fibra em detergente neutro e fibra em detergente ácido. Para fins de seleção simultânea, os genótipos foram classificados conforme desempenho em produção e valor nutritivo segundo o índice de seleção de Taí (1977).

Considerando que existam n caracteres avaliados e destes, m ($m \leq n$) são principais, o agregado genotípico (H) e o índice (I) são dados por:

$$H = a_1g_1 + a_2g_2 + \dots + a_mg_m$$
$$I = b_1x_1 + b_2x_2 + \dots + b_mx_m + b_{m+1}x_{m+1} + \dots + b_{m+1}x_{m+1}$$

em que:

a = pesos econômicos

g = valores genotípicos

x = valores fenotípicos

b = coeficiente dos caracteres no índice

O vetor de ganhos desejados pode ser estabelecido com base em um desvio padrão genético para cada caractere (Cruz et al., 2014).

O vetor b pode ser estimado a partir das equações 1 e 2:

$$P\hat{b} = Ga \quad (1)$$

$$G\hat{b} = \Delta g_d \quad (2)$$

em que:

G = matriz de covariâncias genotípicas entre as variáveis

P = matriz de covariâncias fenotípicas entre as variáveis

Δg_d = vetor de ganhos desejados

Substituindo 2 em 1:

$$a = GP^{-1}G \Delta g_d$$

Considerando que:

$$a = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad e \quad \Delta g_d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

em que:

c_1 e d_1 são vetores de dimensão m x 1

c_2 e d_2 são vetores de dimensão (n-m) x 1

Considerando ainda:

$$c_2 = \phi$$

$$\omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$$

em que:

ω_{11} , ω_{12} , ω_{21} e ω_{22} são matrizes de dimensão $m \times m$, $m \times (n-m)$, $(n-m) \times m$ e $(n-m) \times (n-m)$, respectivamente

Assim:

$$a = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} \omega_{11}d_1 + \omega_{12}d_2 \\ \omega_{21}d_1 + \omega_{22}d_2 \end{bmatrix}$$

e

$$\Delta g_d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ -\omega_{22}^{-1} \omega_{21} d_1 \end{bmatrix}$$

Assim, o vetor b pode ser estimado tanto por 1 quanto por 2.

Após a obtenção dos escores, os genótipos foram ordenados em ordem decrescente segundo os valores obtidos nos índices de produção e valor nutritivo. A partir disso, foram determinados quatro subgrupos para cada índice e os genótipos foram classificados como ótimos, bons, medianos e ruins em cada um deles. O subgrupo 1 denominado “ótimo”, foi composto pelos 20 melhores genótipos, ou seja, aqueles que obtiveram maiores escores nos índices de cada grupo. Os demais grupos (2, 3 e 4) foram denominados bom, mediano e ruim, respectivamente e eram compostos por 19 genótipos cada.

Para treinamento de cada rede, os dados dos grupos produção e valor nutritivo, considerando a classificação de acordo com o índice de Taí (1977) nos cortes 2, 3 e 4, foram ampliados de modo que os dados originais de cada classe do índice gerassem 50 novas observações, originando um total de 600 novos genótipos para produção e 600 para valor nutritivo, como apresentado na Tabela 1.

Tabela 1 – Esquema da ampliação dos dados dos genótipos de alfafa para escolha das redes para produção e valor nutritivo.

Grupo	Cortes	Classificação	Ampliação	Grupo	Cortes	Classificação	Ampliação
Produção	C2	1	50	Valor nutritivo	C2	1	50
		2	50			2	50
		3	50			3	50

	4	50		4	50
	1	50		1	50
C3	2	50	C3	2	50
	3	50		3	50
	4	50		4	50
	1	50		1	50
C4	2	50	C4	2	50
	3	50		3	50
	4	50		4	50

O processo de ampliação consistiu em gerar, a partir dos dados reais, uma variável aleatória com distribuição normal com uma matriz de zeros representando a média e variância equivalente a uma matriz sigma ($X \sim N(\phi, \Sigma)$) por meio da transformação linear $X = F'Y$, em que $Y \sim N(\phi, I)$ representava a variável aleatória correspondente aos dados originais e F é obtido pela decomposição espectral de Σ , tal que $\Sigma^{-1} = FF'$. Para ampliar os dados originais, novos valores de Y foram gerados considerando $Y \sim N(\phi, (F')^{-1}X)$. Para garantir a eficiência do processo, a pressuposição de que a distribuição dos dados deve ser conhecida, com média zero e variância igual a V deve ser atendida. Para isso, utilizou-se o teorema de Box-Muller, em que as variáveis z_1 e z_2 têm distribuição normal com média zero e variância V :

$$z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2) \quad \text{e} \quad z_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

em que U_1 e U_2 são números aleatórios gerados pela distribuição uniforme entre 0 e 1.

Os valores foram simulados conservando as médias, variâncias e covariâncias dos dados originais de cada um dos cortes separadamente. O conjunto de dados de treinamento foi subdividido em 80% para treinamento da rede (480 genótipos) e 20% para validação (120 genótipos). A ampliação foi realizada pelo programa GENES (Cruz, 2013). Uma vez que a rede foi treinada, avaliou-se sua eficiência na classificação de um conjunto de dados, representativos do corte 1, cujas informações não foram consideradas no processo de ampliação.

2.2. Modelagem da Rede Neural Artificial

Neste trabalho, as redes foram treinadas e validadas com os dados ampliados, provenientes dos cortes 2, 3 e 4, mantendo o corte 1 (excluído do processo de ampliação) como informação nova para futura classificação. A arquitetura de rede utilizada para classificar os genótipos foi a Perceptron Multicamadas. Foram estabelecidas duas conformações de RNA, uma para os caracteres do grupo produção e outra para os caracteres do grupo valor nutritivo. Utilizou-se quatro e cinco entradas, correspondentes aos caracteres dos grupos produção e valor nutritivo, respectivamente. Foram testadas as funções de ativação tangente hiperbólica (tansig) e logística (logsig). Para a camada de saída foi utilizada a função limiar (purelin), além disso, foi utilizado o algoritmo de treinamento *trainbr* – *Backpropagation* com um número de épocas igual a 7000.

Foram testados 10 diferentes cenários representados por topologias de rede distintas, que diferiam entre si pelo número de camadas ocultas e pelo número de neurônios em cada uma delas. As RNAs foram processadas pelo software Matlab (Mathworks, 2012) a partir de scripts do aplicativo GENES (Cruz, 2013). As topologias testadas e utilizadas podem ser vistas na Tabela 2:

Tabela 2 – Esquema de treinamento e utilização dos dados dos quatro cortes de alfafa quanto a caracteres de produção e valor nutritivo.

Topologias	Camadas ocultas	Número de neurônios por camada	Utilização
1	1	3	Corte 1
2	1	6	Corte 1
3	1	9	Corte 1
4	2	3 – 3	Corte 1
5	2	3 – 6	Corte 1
6	2	6 – 6	Corte 1
7	3	3 – 3 – 3	Corte 1
8	3	3 – 6 – 3	Corte 1
9	3	6 – 3 – 6	Corte 1
10	3	6 – 6 – 6	Corte 1

3. RESULTADOS E DISCUSSÃO

As topologias treinadas e validadas com os dados ampliados para os grupos de caracteres considerados neste estudo foram utilizadas para classificar os novos dados desconhecidos (Corte 1). Dessa forma, foram obtidas as tabelas de confusão para cada cenário em relação aos dois grupos de caracteres avaliados. Considerando a classificação dos genótipos quanto aos caracteres de produção (Tabela 3), observou-se que as taxas de erro aparente foram, no geral, altas para as etapas de treinamento e validação, atingindo valores desde 3,96% no treinamento do cenário 10, até 30,00% na validação do cenário 3. Além disso, as maiores taxas de erro ocorreram sempre na validação, como encontrado em outros trabalhos (Sant'anna, 2015; Carneiro, 2014). Um dos problemas encontrados na etapa de treinamento de uma rede neural é, sem dúvidas, a qualidade do conjunto de dados apresentados à rede e os parâmetros escolhidos (Pandolfi et al., 2011). No presente trabalho, os dados de treinamento parecem ter sido insuficientes para que as redes conseguissem generalizar as informações e duas razões principais podem explicar os resultados encontrados. A primeira delas diz respeito ao tamanho reduzido do conjunto de dados de treinamento (480 genótipos) que prejudicou o ajuste e classificação das redes, como observado nos estudos de Kominakis et al., 2002 e Jaiswal et al., 2005. A segunda trata da existência de lacunas nos dados, que leva à perda de eficiência do sistema, nessa situação, a adição de novas entradas de rede poderia aumentar a acurácia da classificação (Betker et al., 2003).

Tabela 3 – Taxas de erro aparente nas etapas de treinamento, validação e utilização das redes neurais artificiais em genótipos de alfafa quanto a caracteres de produção.

Cenário	Taxa de erro aparente (%)		
	Treinamento	Validação	Teste
1	25,62	28,33	83,11
2	22,08	28,33	74,02
3	18,95	30,00	92,21
4	21,87	26,67	74,02
5	18,54	26,67	74,02
6	9,58	18,33	72,73
7	18,75	28,33	74,02
8	10,83	17,50	74,02
9	7,08	21,67	74,02
10	3,96	20,83	72,73

Quanto à complexidade das topologias adotadas, as redes com maior número de camadas e neurônios, no geral, mantiveram altas as taxas de erro, resultado já relatado por Zanchettin e Ludermir (2005) que encontraram erros de classificação crescentes e proporcionais ao aumento do número de neurônios adicionados à rede em seu estudo. As taxas de erro aparente quando as redes dos 10 cenários foram testadas no novo conjunto de dados atingiram valores elevados, variando desde 72,73% nos cenários 6 e 10, até 92,21% no cenário 3 (Tabela 3), mostrando a incapacidade de generalização das redes, fato que é justificado pelo pequeno conjunto de treinamento utilizado, que não foi representativo do conjunto de dados apresentado (Tabela 4).

Tabela 4 – Valores comparativos das características de produção mensuradas em alfafa, no conjunto de dados utilizados para treinamento e para uso da rede neural.

Característica	Treinamento (N=600)				Teste (N=77)				F	t
	Média	Máximo	Mínimo	DP	Média	Máximo	Mínimo	DP		
Altura	39,12	51,65	7,43	9,11	51,5	56,67	41,33	3,05	++	**
Prod MS	1429,13	2350	306,67	394,36	2287,25	3177,67	1371,33	354,72	ns	**
C/F	0,57	0,85	0,27	0,11	0,83	1,06	0,71	0,06	++	**
Doenças	2,83	3	2	0,19	2,4	2,67	2,33	0,14	++	**

Prod MS = produção de matéria seca, C/F = relação caule folha. ++ : variância heterogênea pelo teste F a 1% de probabilidade; ** : médias diferentes estatisticamente pelo teste t a 5 e 1% de probabilidade.

Este trabalho preliminar demonstra a necessidade de se utilizar um conjunto de dados mais apropriado para que a rede possa aprender de forma efetiva com os exemplos que lhes são apresentados. Percebe-se, na Tabela 4, que os limites mínimos das variáveis no conjunto de treinamento estão bem aquém daqueles encontrados no conjunto de teste. Além disto, todas as demais estatísticas apontam discrepância de informação de um conjunto em relação ao outro nesta análise elementar univariada que nem mesmo contempla informações sobre interação genótipos x ambientes. Atribuir o insucesso de generalização por deficiência da modelagem da rede pode ser um engano, pois não devemos esperar aprendizado, em analogia ao cérebro humano, de situações tão díspares. Diversos trabalhos disponíveis na literatura demonstram a grande potencialidade das redes, inclusive de sua capacidade de extrapolação para dados de validação, em conjunto de dados

relacionados por informações básicas como médias, variâncias e covariâncias. Entretanto, os resultados obtidos neste trabalho nos levam a reconhecer que o uso de pequenos conjuntos de dados, como aqueles gerados rotineiramente na pesquisa agropecuária, é sujeito a críticas pois haverá grande lacuna de informação entre aprendizado e generalização. Acredita-se que as redes terão grande valor na medida que haja mudança no paradigma da experimentação em que as decisões passarão a ser tomadas em bancos de dados, construídos ao longo do tempo de um programa de melhoramento, que constituirão de base para o processamento de dados.

A classificação dos genótipos quanto aos caracteres de valor nutritivo (Tabela 5), levaram a pequenas taxas de erro tanto no treinamento quanto na validação. Na realidade, essas baixas taxas de erro não podem ser interpretadas como fator totalmente positivo, já que em nenhum cenário estudado, a utilização da rede no novo conjunto de dados apresentou taxa de erro baixa. Isso significa que a rede perdeu a habilidade de generalização, possivelmente pelo sobreajuste (*overfitting*) que pode ter ocorrido. O *overfitting* se deve principalmente ao fato de a rede se especializar no conjunto de treinamento, nesses casos, há memorização dos dados e suas particularidades, levando a perdas na capacidade de generalização (Santos, et al., 2005). Como o conjunto de dados inicial ampliado, foi subdividido em treinamento e validação, nesta última etapa a rede também conseguiu obter taxas de erro baixas, exceto para o cenário 10 (17,50%).

Tabela 5 – Taxas de erro aparente nas etapas de treinamento, validação e utilização das redes neurais artificiais em genótipos de alfafa quanto a caracteres de valor nutritivo.

Cenário	Taxa de erro aparente (%)		
	Treinamento	Validação	Teste
1	11,67	11,67	55,84
2	8,12	10,00	70,13
3	7,92	5,83	61,04
4	8,54	5,00	54,54
5	6,67	5,00	48,05
6	2,08	5,83	68,83
7	5,21	5,00	50,65
8	4,58	5,00	50,65
9	1,67	9,17	63,64
10	0,00	17,50	55,84

O aumento da complexidade das redes levou a uma tendência de redução das taxas de erro de treinamento, que chegou a 0% no cenário 10. Entretanto, como já enfatizado anteriormente, esses resultados não demonstram maior eficiência, fato que pode ser notado pelas altas taxas de erro aparente, que variaram de 48,05% no cenário 5 até 70,13% no cenário 2, ao se utilizar as redes no novo conjunto de dados. Como relatado por Pandorfi et al. (2011), o maior número de camadas de uma rede neural além de elevar a complexidade e tempo de processamento, leva a um menor poder de generalização, por isso é fundamental que se ajuste uma rede neural considerando todas as peculiaridades do conjunto de dados para evitar problemas de ajuste.

Pode-se inferir que para o conjunto de caracteres de valor nutritivo, o conjunto de dados de treinamento, assim como observado para os caracteres de produção, também foi insuficiente para que as redes conseguissem generalizar as informações para novas observações. Verifica-se, na Tabela 6, discrepância menores entre as estatísticas básicas do conjunto de dados de treinamento e o de teste, o que contribuiu para uma ligeira melhoria do desempenho da rede, mas em níveis não tão animadores. Novamente fica clara a importância do tamanho e qualidade das informações do conjunto de dados para ajuste de uma rede neural, já que a eficiência desta aumenta proporcionalmente com o aumento do conjunto de dados e informações deste (Jaiswal et al., 2005). Considerando o alto potencial de geração de informações do programa de melhoramento da alfafa, a reunião de um banco de dados históricos da cultura seria possível e eficiente para que as redes neurais destinadas à classificação de genótipos também o fossem.

Tabela 6 – Valores comparativos das características de qualidade, mensuradas em alfafa, no conjunto de dados utilizados para treinamento e para uso da rede neural.

Característica	Treinamento (N=600)				Teste (N=77)				F	t
	Média	Máximo	Mínimo	DP	Média	Máximo	Mínimo	DP		
Lignina	7.41	8.23	6.12	0,46	8.51	11.32	6.07	1.21	ns	ns
PB	26.57	30.02	24.69	1,04	25.7	28.43	23.28	1.11	++	**
DIVMS	70.54	73.05	67.84	1,16	69.27	73.37	65.1	1.69	ns	**
FDN	42.05	50.39	35.97	3,83	44.37	52.96	35.65	4.03	++	**
FDA	30.27	32.71	27.25	0,92	31	41.62	26.2	2.38	ns	**

PB = proteína bruta, DIVMS = digestibilidade “*in vitro*” da matéria seca, FDN = fibra em detergente neutro, FDA = fibra em detergente ácido. ++ : variância heterogênea pelo teste F

a 1% de probabilidade; ** : médias diferentes estatisticamente pelo teste t a 5 e 1% de probabilidade.

A eficiência das RNAs tem sido amplamente demonstrada na literatura, nas diferentes áreas da ciência, como pode ser confirmado por Barbosa et al. (2011), Barroso et al. (2013), Brasileiro et al. (2015), Carneiro (2015), Gianola et al. (2011), Ji et al. (2007), Kominakis et al (2002), Nascimento et al. (2013), Pandorfi et al. (2011), Peixoto (2013), Sant'anna (2014), Santos, et al. (2005), Silva (2014), Zanchettin e Ludermir (2005), entre outros. Em todos os casos, é preciso que o pesquisador apresente à rede, informações de qualidade. A partir de um conjunto de dados bastante representativo e que caracterize diversas situações, é possível estabelecer uma boa rede neural e utilizá-la com sucesso nos programas de melhoramento.

4. CONCLUSÕES

As redes neurais artificiais são eficientes para classificar indivíduos se estas conhecerem bem os padrões de classificação desejados. Como os limites mínimos das variáveis no conjunto de treinamento foram muito diferentes daqueles encontrados no conjunto de teste, o resultado encontrado neste trabalho confirma a necessidade de atenção aos dados fornecidos à rede.

A abordagem das redes neurais artificiais é ineficaz quando o conjunto de dados para treinamento é limitado a poucas informações.

As redes neurais serão mais difundidas à medida que as decisões dos programas de melhoramento deixarem de ser tomadas com base em pequenos conjuntos de dados e passarem a ser tomadas com base em séries históricas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Barbosa, C. D.; Viana, A. P.; Quintal, S. S. R.; Pereira, M. G. (2011) Artificial neural network analysis of genetic diversity in *Cariacica papaya* L. *Crop Breeding and Applied Biotechnology*. 11:224-231.
- Barroso, L. M. A.; Nascimento, M.; Nascimento, A. C. C.; Silva, F. F.; Ferreira, R. P. (2013) Uso do método de Eberhart e Russell como informação a priori para aplicação de redes neurais artificiais e análise discriminante visando a classificação de genótipos de alfafa quanto à adaptabilidade e estabilidade fenotípica. *Ver. Bras. Biom.*, 31:176-188.
- Basigalup, D.H.; Odorizzi, A. S. Melhoria genética da alfafa. In: Ferreira, R. P.; Basigalup, D. H.; Gioco, J. O. (Ed.). *Melhoramento Genético da Alfafa*. São Carlos: Embrapa Pecuária Sudeste, 2011. p. 225-260.
- Brasileiro, B. P. et al. (2015) Selection in sugarcane families with artificial neural networks. *Crop Breed. Appl. Biotechnol.*, 5:72-78.
- Betker, A.L.; Szturm, T.; Moussavi, Z. Application of feedforward backpropagation neural network to center of mass estimation for use in a clinical environment. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE EMBS, 25., 2003, Cancun. Proceedings... Cancun: IEEE EMBS, 2003. p.17-21.
- Carneiro, V. Q. (2015) Rede neural e lógica Fuzzy aplicadas no melhoramento do feijoeiro. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 108p, 2015.
- Cruz, C. D. (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum.*, 35:271-276.
- Dantas, D.; Luz, T. M. O.; Souza, M. J. H.; Barbosa, G. P.; Cunha, E. G. S. (2016) Uso de Redes Neurais Artificiais na previsão da precipitação de períodos chuvosos. *Revista Espinhaço*, 5:11-18.

- De Veaux, R., Ungar, L. (1997). A brief introduction to neural networks, Technical Report, Williams College, Williamstown, MA.
- Galvão, C. O.; Valença, M. J. S.; Vieira, V. P. P. B.; Diniz, L. S. et al. (1999) Sistemas inteligentes: Aplicações a recursos hídricos e ciências ambientais. UFRGS: ABRH.
- Gianola, D.; Okut, H.; Kent A. W.; Rosa, J. M. R. (2011) Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genetics*, 45:34.
- Haykin, S. (2008) Neural Networks and Learning Machines (3rd ed.). Hamilton: Pearson – Prentice Hall, 906p.
- Jaiswal, S.; Benson, E.R.; Bernard, J.C. et al. (2005) Neural network modelling and sensitivity analysis of a mechanical poultry catching system. *Biosystems Engineering*, 92:59-68.
- Ji, B.; Sun, Y.; Yang, S. and Wan, J. (2007) Artificial neural networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science*, 145:249-261.
- Kominakis, A.P.; Abas, Z.; Maltaris, I. et al. (2002) A preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. *Computers and Eletronics in Agriculture*, 35:35-48.
- Li, X.; Brummer, E. C. (2012) Applied Genetics and Genomics in Alfalfa Breeding. *Agronomy*, 2:40-61.
- Mackay, D. J. C. (1994) Bayesian non-linear modelling for the prediction competition. In: ASHRAE Transactions, ASHRAE, Atlanta Georgia. 100:1053-1062.

MathWorks (2012). MATLAB – The Language of Technical Computing. Disponível em:
<http://www.mathworks.com/products/matlab>.

Nascimento, M.; Peternelli, L. A.; Crus, C. D.; Nascimento, A. C. C.; Ferreira, R. P.; Bhering, L. L. and Sangado, C. C. (2013) Artificial neural networks for adaptability and stability evaluation in alfafa genotypes. *Crop Breeding and Applied Biotechnology*, 13:152-156.

Norvig, P. and Russell, S. (2013) Inteligência Artificial. 3rd ed. CAMPOS, Rio de Janeiro, Elsevier, 958p.

Pandorfi, H.; Silva, I. J. O.; Sarnighausen, V. C. R.; Vieira, F. M. C.; Nascimento, S. T.; Guiselini, C. (2011) Uso de redes neurais artificiais para predição de índices zootécnicos nas fases de gestação e maternidade na suinocultura. *R. Bras. Zootec.*, 40:676-681.

Peixoto, L. A. Redes Neurais Artificiais na predição do valor genético. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 97p, 2013.

Sant'anna, I. C. Redes Neurais Artificiais na discriminação de populações de retrocruzamento com diferentes graus de similaridade. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 115p, 2014.

Santos, A. M.; Seixas, J. M.; Pereira, B. B.; Medronho, R. A. (2005). Usando redes neurais artificiais e regressão logística na predição da Hepatite A. *Rev. bras. epidemiol.*, 8:117-126

Silva, G. N. Redes Neurais Artificiais: Novo paradigma para a predição de valores genéticos. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, 92p, 2014.

Tai, G. C. C. (1977) Index selection with desired gains. *Crop Sci.*, 17:182-183.

Zanchettin, C, & Ludermir, T. B. (2005). Sistemas neurais híbridos para reconhecimento de padrões em narizes artificiais. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, 16:159-172.

6. CONCLUSÕES GERAIS

O conjunto de genótipos de alfafa avaliados neste trabalho apresenta valores para os caracteres de valor nutritivo considerados ótimos, e por isso, a seleção para esses caracteres pode ser mais flexível. Entretanto, para os caracteres de produção a seleção deve ser efetiva visto que os objetivos nos programas de melhoramento quase sempre envolvem aumento da produtividade. A existência de variabilidade aliada às altas médias caracteriza a potencialidade do conjunto de genótipos para utilização no programa de melhoramento de alfafa da Embrapa Pecuária Sudeste.

Métodos de seleção univariadas se mostraram inferiores se comparados à seleção simultânea, que proporcionou uma distribuição de ganhos para o conjunto de caracteres mais equilibrada em todos os cortes avaliados.

A reunião das informações coletadas nos genótipos de alfafa em um banco de dados temporal se mostrou útil para otimizar a seleção simultânea baseada na abordagem das redes neurais artificiais. A partir do estabelecimento do índice de seleção de Taí, as redes foram treinadas e demonstraram sua eficiência para classificação de genótipos, podendo ser utilizadas com alta eficiência no programa de melhoramento da cultura da alfafa.

Entretanto, quando se limitou o conjunto de dados de treinamento a informações de cortes realizados em três estações do ano e utilizou-se a rede treinada para dados coletados em uma particular estação até então desconsiderada, estas foram ineficazes. Este resultado não invalida a utilização dessa abordagem no melhoramento de plantas, pelo contrário, demonstra a importância do conjunto de dados apresentados na etapa de treinamento da rede.

A abordagem das redes neurais artificiais é uma alternativa eficiente para estudos de natureza classificatória desde que seja bem treinada, conservando seu poder de generalização. Sua utilização será mais difundida à medida que as decisões dos programas de melhoramento deixarem de ser tomadas com base em pequeno conjunto de dados experimentais e passarem a ser tomadas com base em séries históricas.

APÊNDICE

Capítulo 1

Relação de genótipos selecionados (S) e não selecionados (NS) para os grupos de caracteres produção e valor nutritivo considerando os cortes 1, 2, 3 e 4 (C1, C2, C3 e C4, respectivamente)

Genótipo	Produção				Valor Nutritivo			
	C1	C2	C3	C4	C1	C2	C3	C4
1	S	NS	NS	NS	S	S	S	S
2	NS	NS	NS	NS	S	NS	NS	NS
3	NS	NS	NS	NS	NS	S	NS	S
4	S	NS	S	S	NS	S	NS	NS
5	NS	NS	NS	NS	NS	S	NS	NS
6	NS	NS	NS	NS	NS	NS	S	NS
7	NS	NS	NS	NS	NS	NS	S	S
8	NS	NS	NS	NS	NS	NS	NS	NS
9	NS	NS	S	S	NS	NS	NS	NS
10	NS	NS	S	NS	NS	NS	NS	NS
11	NS	NS	NS	NS	NS	S	NS	NS
12	S	S	S	NS	NS	NS	NS	NS
13	NS	NS	NS	S	NS	NS	NS	NS
14	S	S	S	S	NS	NS	NS	NS
15	NS	NS	NS	NS	S	S	S	S
16	NS	S	S	NS	NS	NS	NS	NS
17	S	NS	NS	NS	NS	S	S	S
18	NS	NS	NS	NS	S	S	S	S
19	NS	NS	NS	NS	NS	NS	NS	NS
20	NS	S	NS	NS	NS	NS	NS	NS
21	NS	NS	S	S	NS	NS	NS	NS
22	NS	S	S	S	NS	NS	NS	NS
23	NS	NS	NS	NS	S	S	S	S
24	NS	NS	NS	NS	NS	NS	NS	NS
25	NS	S	NS	S	S	NS	NS	NS
26	NS	S	NS	NS	NS	NS	NS	NS
27	NS	S	NS	NS	NS	NS	NS	NS
28	NS	NS	NS	NS	NS	NS	NS	NS
29	NS	NS	NS	NS	S	NS	NS	NS
30	NS	NS	NS	NS	NS	NS	NS	NS
31	NS	NS	NS	NS	NS	NS	NS	NS
32	S	NS	NS	NS	NS	NS	NS	NS
33	NS	NS	NS	NS	NS	NS	NS	NS
34	NS	NS	NS	NS	NS	NS	NS	NS

35	NS	NS	NS	NS	S	S	S	S
36	NS	NS	NS	NS	S	NS	NS	NS
37	NS	S	NS	NS	S	NS	NS	S
38	NS	NS	S	NS	NS	NS	NS	NS
39	S	NS	NS	NS	NS	NS	NS	NS
40	S	S	S	S	NS	NS	NS	NS
41	NS	S	NS	S	NS	NS	S	NS
42	S	NS	NS	S	NS	NS	NS	NS
43	NS	NS	NS	NS	S	NS	S	NS
44	S	NS	NS	NS	NS	NS	NS	S
45	NS	NS	S	S	NS	NS	NS	NS
46	NS	S	NS	NS	S	NS	NS	NS
47	NS	NS	NS	NS	NS	S	NS	S
48	S	NS	NS	NS	NS	NS	S	NS
49	NS	NS	NS	NS	S	S	S	S
50	NS	NS	NS	NS	S	S	S	S
51	NS	NS	NS	NS	S	S	NS	NS
52	NS	NS	NS	NS	NS	NS	NS	NS
53	NS	NS	NS	NS	NS	NS	S	NS
54	S	NS	NS	NS	NS	NS	NS	NS
55	NS	NS	NS	NS	S	S	S	S
56	S	S	NS	NS	NS	S	NS	NS
57	S	NS	NS	S	NS	NS	NS	NS
58	NS	NS	NS	NS	NS	S	S	S
59	NS	NS	NS	NS	S	S	S	S
60	S	NS	S	S	NS	NS	NS	NS
61	S	S	S	S	NS	NS	NS	NS
62	NS	NS	S	NS	NS	NS	NS	NS
63	NS	S	NS	NS	NS	NS	NS	NS
64	S	S	S	S	NS	NS	NS	NS
65	NS	NS	NS	NS	NS	S	NS	S
66	NS	S	S	S	NS	NS	NS	NS
67	S	NS	NS	NS	S	NS	NS	NS
68	S	NS	NS	NS	NS	NS	S	S
69	NS	NS	NS	NS	S	NS	S	S
70	S	NS	NS	NS	NS	S	S	S
71	NS	NS	NS	NS	S	NS	NS	NS
72	NS	NS	S	S	NS	NS	NS	NS
73	NS	NS	NS	NS	NS	NS	NS	NS
74	NS	S	NS	NS	NS	NS	NS	NS
75	NS	NS	S	S	NS	NS	NS	NS
76	NS	S	S	S	NS	NS	NS	NS
77	NS	S	S	S	NS	NS	NS	NS
