

PREDIÇÃO DA TAXA DE PROGRESSO DA FERRUGEM DO CAFEIEIRO POR MEIO DE *ENSEMBLES*¹

Thiago T. Thamada²; Luiz H. A. Rodrigues³; Carlos A. A. Meira⁴

¹Trabalho financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

²Mestrando, Faculdade de Engenharia Agrícola da Unicamp, Campinas-SP, ic.tiba@gmail.com

³Professor, PhD, Faculdade de Engenharia Agrícola da Unicamp, Campinas-SP, lique@feagri.unicamp.br

⁴Pesquisador, PhD, Embrapa Informática Agropecuária, Campinas-SP, carlos.meira@embrapa.br

RESUMO: Este trabalho apresenta um processo de desenvolvimento de modelos preditivos por meio de *ensembles*, que irão prever aumentos na taxa de progresso da ferrugem do cafeeiro, a partir de dados meteorológicos e de incidência mensal da doença em lavouras de alta ou baixa carga pendente de fruto. Desenvolver *ensembles* visa identificar e entender os fatores climáticos que favorecem a ocorrência da ferrugem tardia. Posteriormente, os *ensembles* serão incorporados ao SafCafe, um sistema de alerta, para auxiliar técnicos da Fundação Procafé na elaboração de boletins mensais de aviso fitossanitários.

PALAVRAS-CHAVE: *Hemileia vastatrix*, modelos de alerta, modelagem, mineração de dados, ferrugem tardia

PREDICTING INFECTION RATE OF COFFEE RUST BY ENSEMBLES

ABSTRACT: The aim of this study is to present a process of predictive models development by ensembles to forecast the growth of the monthly progress rate for coffee rust from meteorological and monthly incidence data on high or low fruit load crops. Ensembles development aim to identify and understand climate factors which favor the late rust occurrence. Later, the ensembles will be embedded into SafCafe, a warning system, to assist technicians of Fundação Procafé to prepare monthly warning phytosanitary newsletters.

KEYWORDS: *Hemileia vastatrix*, warning models, modeling, data mining, late rust

INTRODUÇÃO

Causada pelo fungo *Hemileia vastatrix* Berk. et Br., a ferrugem é a principal doença do cafeeiro. Em condições climáticas favoráveis a ferrugem provoca perdas de, em média, 35% na produção podendo ultrapassar 50%. O cafeeiro possui ciclo bienal de produção, ou seja, alterna anos de alta e baixa carga pendente de frutos. A ferrugem é mais agressiva em anos de alta carga por razões ainda desconhecidas. Porém, aceita-se que a drenagem de fitossintetizados das folhas para os frutos aliado ao alto índice de enfolhamento das plantas no começo do período chuvoso e condições climáticas favoráveis à epidemia da ferrugem sejam as razões por sua maior agressividade em plantas com alta carga pendente (Zambolim et al., 2005).

A curva de progresso típica da doença tem início entre dezembro e janeiro, atinge seu máximo em torno de junho e, a partir de então, decresce devido às baixas temperaturas e à desfolha das plantas. O controle da ferrugem é realizado, principalmente, por aplicações de fungicidas a partir de dezembro, seguindo calendário fixo (Chalfoun et al., 2001; Zambolim et al., 2005).

Variações no clima como o aumento de temperatura e alterações no regime de chuva podem alterar a curva de progresso da ferrugem, atrasando seu início ou mantendo altos níveis da doença ao fim de seu ciclo, em agosto (Chalfoun et al., 2001). Em casos de desenvolvimento tardio da ferrugem é preciso rever e readaptar suas medidas de controle, já que as aplicações de fungicidas por calendário fixo se tornam ineficientes (Chalfoun e Carvalho, 1999).

Posicionar adequadamente as aplicações de fungicidas mantém baixa a intensidade da doença durante seu ciclo. Um sistema de alerta de doença de planta é uma ferramenta capaz de auxiliar nessa questão. Esse tipo de sistema pode incorporar modelos preditivos, desenvolvidos por meio de técnicas de mineração de dados e algoritmos computacionais, para prever quando uma doença atingirá níveis críticos na lavoura (Hardwick, 2006).

Sistema de alerta e modelos preditivos para ferrugem do cafeeiro já foram desenvolvidos (Meira et al., 2009; Thamada et al., 2013; Girolamo Neto et al., 2014). Estes modelos foram criados a partir de dados (meteorológicos e de incidência mensal da doença) referentes a todo ano agrícola e com base em apenas uma técnica de mineração de dados dentre rede neural, SVM (máquina de vetor suporte) e árvore de decisão.

A fim de desenvolver modelos com desempenho preditivo superior aos atuais pode-se utilizar o *ensemble*, técnica de mineração de dados que reúne diversos modelos em um comitê de votação onde a predição de cada modelo é considerada para inferir a predição final. *Ensemble* pode ser utilizado para melhorar o desempenho preditivo dos

modelos que o compõe (Opitz e Maclin, 1999; Hastie et al., 2009). Outro ponto relevante para estudo do progresso da ferrugem é aprender sobre os fatores climáticos que auxiliam ou promovem o desenvolvimento tardio da doença, podendo ser interessante limitar o período de estudo.

O objetivo deste trabalho é apresentar um processo de desenvolvimento de *ensembles* para ferrugem do cafeeiro capazes de prever o aumento da taxa de progresso da doença na lavoura, criados a partir de dados coletados durante o período relevante para desenvolvimento da epidemia, dezembro a junho. Assim, os *ensembles* poderiam auxiliar no posicionamento adequado das medidas de controle, principalmente em anos de desenvolvimento tardio da ferrugem.

MATERIAL E MÉTODOS

O desenvolvimento dos *ensembles* para ferrugem do cafeeiro será realizado usando a metodologia *CRISP-DM* (Chapman et al., 2000). O ciclo de vida da metodologia envolve seis fases: compreensão do domínio, entendimento dos dados, preparação dos dados, modelagem, avaliação e distribuição. A metodologia é iterativa e não há rigidez em sua sequência lógica, sendo possível avançar ou voltar entre as fases ao longo da execução de um projeto, caso seja necessário. Este trabalho realizará um novo ciclo no processo de desenvolvimento de modelos preditivos iniciado por Meira et al. (2009) e continuado por Girolamo Neto et al. (2014).

Com base nas fases do *CRISP-DM* a Fig. 1. mostra uma visão macro dos procedimentos a serem desenvolvidos para criação dos *ensembles* e obtenção de sua predição final.

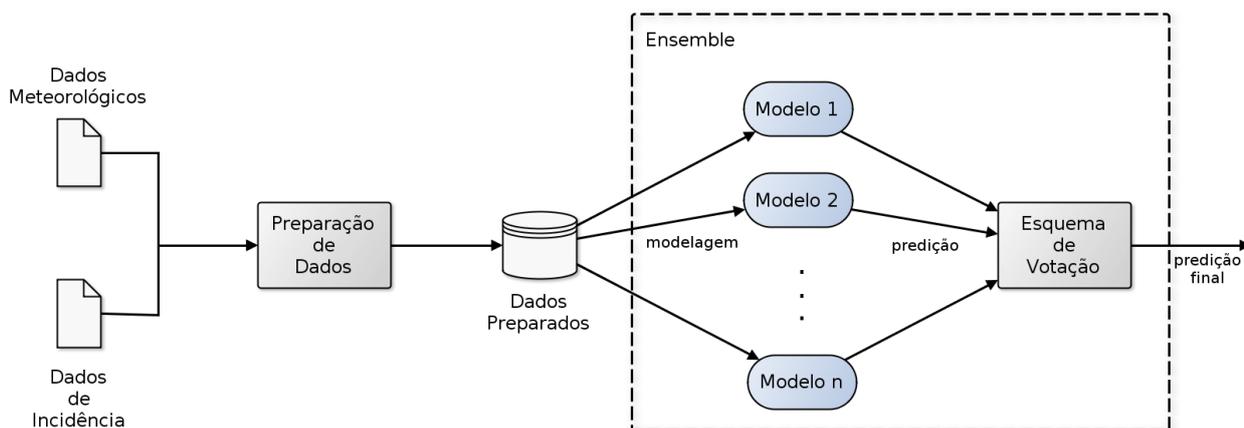


Fig. 1. Visão macro do processo de desenvolvimento de *ensembles*.

Os dados que serão utilizados no projeto foram coletados em fazendas experimentais da Fundação Procafé localizadas nas cidades mineiras de Boa Esperança (03/2010 a 06/2014), Carmo de Minas (03/2006 a 06/2014) e Varginha (09/1998 a 06/2014). A coleta dos dados meteorológicos ocorreu em intervalos de trinta minutos ao longo de um mês e corresponde às variáveis climáticas: temperaturas média, máxima e mínima, precipitação e umidade relativa do ar (UR). Valores sobre a incidência mensal da ferrugem em lavouras de cafeeiros com alta e baixa carga pendente de frutos também foram coletados para as três fazendas experimentais.

Dados meteorológicos serão submetidos a testes a fim de verificar sua consistência (por exemplo, valor de temperatura máxima deve ser maior que o valor de temperatura mínima) e detectar registros defeituosos ou faltantes. Caso os registros de um determinado mês estejam muito comprometidos, todos os dados referentes a esse mês serão descartados.

As granularidades entre os dados meteorológicos (nível horário) e de incidência (nível mensal) são diferentes. Assim, o conjunto de dados meteorológicos será transformado, com base em Meira (2008), do nível horário para diário e, por fim, para mensal (Fig. 2.). Dessa forma será possível relacionar os dados meteorológicos com os de incidência da doença. Também serão derivadas, baseado em Meira (2008) e Girolamo Neto et al. (2014), novas variáveis climáticas considerando os períodos de incubação (Moraes et al., 1976), de molhamento foliar e de infecção (PINF).

Uma vez que o intervalo de estudo para ferrugem tardia é de dezembro a junho, somente os registros coletados durante esse período serão selecionados para formar um conjunto de dados preparado, pronto para ser utilizado no desenvolvimento de modelos preditivos simples e *ensembles*.

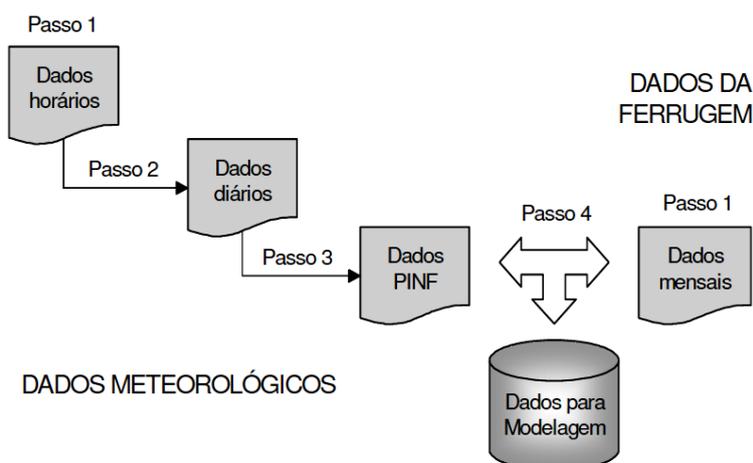


Fig. 2. Esquema da transformação dos dados meteorológicos (Meira, 2008).

O atributo meta, variável a ser predita, será a taxa de progresso mensal (TP) da ferrugem na lavoura. Desse modo, o *ensemble* irá prever se no próximo mês a TP será maior ou igual a um determinado limiar a partir de dados meteorológicos do mês atual. Serão usados dois limiares distintos: 5 e 10 pontos percentuais (p.p.). Estes valores foram escolhidos com base em Zambolim et al. (1997), que indicam o início do controle da ferrugem caso a porcentagem de folhas doentes na lavoura seja de, aproximadamente, 5%, e em Kushalappa et al. (1984), que propuseram 10% como valor limite de folhas infectadas para aplicação de fungicidas. O atributo meta será binário, ou seja, caso o *ensemble* indique aumento na TP superior ao limiar, a predição final do *ensemble* terá valor 1; caso contrário, será 0.

No processo de modelagem, técnicas de mineração de dados serão aplicadas aos dados preparados a fim de construir os modelos preditivos simples. Serão utilizadas as técnicas: *SVM*, rede neural e árvore de decisão. A escolha dessas técnicas se justifica por suas influências e popularidade entre a comunidade científica (Wu et al., 2008) e pelo trabalho de Girolamo Neto et al. (2014), que conseguiram modelos com bons desempenhos preditivos usando *SVM*. Construídos os modelos, estes irão compor um *ensemble* que, através de um sistema de votação entre as predições dos modelos simples, irá emitir o valor da predição final.

Serão desenvolvidos *ensembles* com base em dados de incidência mensal coletados em lavouras com alta e baixa carga pendente e nos dois valores de limiar (5 e 10 p.p.), exceto para o cenário baixa carga e 10 p.p. (Tabela 1). A recusa pelo desenvolvimento de modelos para baixa carga e 10 p.p. se justifica pelo baixo desempenho preditivo que estes apresentaram em trabalho anterior (Meira e Rodrigues, 2009).

Haverá dois conjuntos de dados, um contendo registros das três fazendas experimentais (Tudo) e outro com dados de Varginha. A escolha pelo segundo conjunto é justificado pelo intervalo de tempo em que os dados foram coletados, superior a doze anos, suficiente para que se identifique quais fatores influenciam no progresso de uma doença de planta (Coakley, 1988). Cenários que serão usados estão indicados na Tabela 1.

Tabela 1. Cenários a serem usados para desenvolver os *ensembles*.

Conjunto de dados	Carga pendente	Limiar (p. p.)
Tudo	Alta	5
Tudo	Alta	10
Tudo	Baixa	5
Varginha	Alta	5
Varginha	Alta	10
Varginha	Baixa	5

O desenvolvimento dos *ensembles* será feito por meio das técnicas *bagging*, *boosting*, *stacking* e floresta aleatória. Cada uma dessas técnicas possui características que as diferenciam como, por exemplo, o manuseio do conjunto de dados durante a criação dos modelos simples e modo de seleção das variáveis climáticas que irão compor cada modelo simples (Opitz e Maclin, 1999; Hastie et al., 2009).

O esquema de votação a ser implementado será sem peso, ou seja, a predição inferida por cada modelo simples terá o mesmo peso no resultado final do *ensemble*. Todo o desenvolvimento de modelos preditivos simples e *ensembles* será realizado por meio de *scripts* escritos na linguagem de programação R, que contém bibliotecas específicas para tarefas de mineração de dados, análise exploratória de registros e plotagem de gráficos, requisitos necessários nessa pesquisa. O ambiente de desenvolvimento a ser usado é o RStudio, ferramenta gratuita e com suporte para instalação e atualização das bibliotecas do R.

A avaliação do desempenho preditivo dos *ensembles* será realizada com dados do ano agrícola 2014/15, ou seja, de dezembro de 2014 a junho de 2015. Dessa forma poderá ser conferida se a performance dos *ensembles* se mantém para dados que não foram usados em suas criações.

RESULTADOS E DISCUSSÃO

Atualmente encontra-se em operação o SafCafe (Sistema de Alerta para Ferrugem do Cafeeiro), sistema *web* para emissões de alerta sobre a ferrugem do cafeeiro, desenvolvido pela Embrapa com apoio do Consórcio Pesquisa Café (Thamada et al., 2013). O SafCafe incorpora modelos desenvolvidos por Girolamo Neto et al. (2014), que prevêem aumentos na TP no próximo mês para lavouras situadas em Boa Esperança, Carmo de Minas e Varginha.

O sistema realiza a predição com base em dados meteorológicos e de incidência mensal da ferrugem na lavoura e os cenários avaliados pelo sistema são: alta e 5 p.p.; alta e 10 p.p.; e baixa e 5 p.p.. Caso os modelos indiquem aumento na TP superior ao limiar, o alerta é positivo; caso contrário, é negativo (Thamada et al., 2013).

Recentemente, modelos preditivos para Varginha, incorporados ao SafCafe, foram avaliados com dados meteorológicos referentes a três anos agrícolas (Meira et al., 2014), que não foram usados por Girolamo Neto et al. (2014). A avaliação ocorreu para 2011/12, 2012/13 e 2013/14. Modelos para baixa carga e 5 p.p. apresentaram performance preditiva fraca, 71,2% de acurácia para todo o período. Na avaliação de 2013/14 foi constatado que os modelos de alta carga e 5 p.p. não identificaram o progresso tardio da ferrugem, que ocorreu a partir de abril de 2014.

O uso do sistema é exclusivo para os técnicos da Fundação Procafé, que utilizam as predições como informações auxiliares na elaboração de boletins mensais de avisos fitossanitários para a região sul de Minas Gerais. Os boletins são distribuídos aos cafeicultores e interessados na cultura do café.

Este projeto se insere dentro do projeto que envolve o SafCafe e pode ser entendido como um novo ciclo de desenvolvimento de modelos preditivos para ferrugem do cafeeiro, que anteriormente contou com colaborações de Meira et al. (2009) e Girolamo Neto et al. (2014).

Espera-se, ao fim desta pesquisa, que a técnica *ensemble*, juntamente com uma nova proposta de preparação dos dados, seja capaz de prever com bom desempenho preditivo o aumento da TP, especialmente em anos de ocorrência da ferrugem tardia. *Ensembles* desenvolvidos com base em dados de lavouras de baixa carga deverão manter seu desempenho preditivo quando avaliados com dados mais recentes. A avaliação dos *ensembles* por meio de dados externos (ano agrícola 2014/15), que não foram usados na modelagem, deve ser um bom indicador para mensurar seus desempenhos preditivos.

O processo de preparação dos dados originais até o conjunto de dados prontos para modelagem visará remover variáveis reportadas como irrelevantes (Meira et al., 2009; Girolamo Neto et al., 2014) e criar novas variáveis climáticas baseadas em temperatura, precipitação e UR, fundamentais para o desenvolvimento da ferrugem (Zambolim et al., 2005). A criação de novas variáveis climáticas também tem o objetivo de identificar os fatores climáticos que favorecem a ocorrência da ferrugem tardia e permitam um maior entendimento no estudo da doença.

Futuramente, com a confirmação do bom desempenho preditivo dos *ensembles* por meio de avaliações feitas nos próximos anos agrícolas, pretende-se incorporar os *ensembles* gerados neste projeto no SafCafe como forma de atualizar os modelos preditivos incorporados ao sistema. Assim, o SafCafe apresentará bom desempenho preditivo e irá utilizar modelos desenvolvidos focados na detecção do aumento da TP somente nos períodos relevantes para a propagação de uma epidemia da ferrugem e que, possivelmente, consigam detectar anos em que a ferrugem tardia irá ocorrer.

CONCLUSÕES

Espera-se que o processo de desenvolvimento de *ensembles* para a ferrugem do cafeeiro apresentado neste trabalho possa gerar modelos de bom desempenho preditivo para todos os cenários estudados e que sejam capazes de identificar a ocorrência da ferrugem tardia. Nesse caso os *ensembles* serão incorporados ao SafCafe, sendo uma ferramenta auxiliar na elaboração de boletins mensais de avisos fitossanitários emitidos pela Fundação Procafé.

AGRADECIMENTOS

À Fundação Procafé, pelo fornecimento dos dados, e ao CNPq e Consórcio Pesquisa Café, pelo apoio financeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

- CHALFOUN, S. M.; CARVALHO, V. L. de. Controle químico da ferrugem (*Hemileia vastatrix* Berk & Br.) do cafeeiro através de diferentes esquemas de aplicação. Pesquisa Agropecuária Brasileira, v. 34, n. 3, p. 363-367, 1999.
- CHALFOUN, S. M.; CARVALHO, V. L. de; PEREIRA, M. C. Efeito de alterações climáticas sobre o progresso da ferrugem (*Hemileia vastatrix* Berk. & Br.) do cafeeiro (*Coffea arabica* L.). Ciência e Agrotecnologia, Lavras, v. 25, n. 5, p. 1248-1252, 2001.
- CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. CRISP-DM 1.0: step-by-step data mining guide. Illinois: SPSS, 2000.
- COAKLEY, S. M. Variation in climate and prediction of disease in plants. Annual Review of Phytopathology, v. 26, p. 163-181, 1988.
- GIROLAMO NETO, C. di; RODRIGUES, L. H. A.; MEIRA, C. A. A. Modelos de predição da ferrugem do cafeeiro (*Hemileia vastatrix*) por técnicas de mineração de dados. Coffee Science, v. 9, n. 3, p. 408-418, 2014.
- HARDWICK, N. V. Disease forecasting. In: COOKE, B. M.; JONES, D. G.; KAYE, B. (Eds.). The epidemiology of plant diseases. 2nd edition. Wageningen: Springer, 2006. p. 239-267.
- HASTIE, T.; TIBSHINARI, R.; FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edition. New York:Springer, 2009. 739 p.
- KUSHALAPPA, A. C.; AKUTSU, M.; OSEGUERA, S. H.; CHAVES, G. M.; MELLES, C. Equations for predicting the rate of coffee rust development based on net survival ratio for monocyclic process of *Hemileia vastatrix*. Fitopatologia Brasileira, v. 9, p. 255-271, 1984.
- MEIRA, C. A. A. Processo de descoberta de conhecimento em bases de dados para a análise e o alerta de doenças de culturas agrícolas e sua aplicação na ferrugem do cafeeiro. 198p. Tese (Doutorado em Engenharia Agrícola) - Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas. 2008.
- MEIRA, C. A. A.; RODRIGUES, L. H. A. Modelos em árvore de decisão para alerta da ferrugem do cafeeiro. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 6., 2009, Vitória. Anais... Vitória: Consórcio Pesquisa Café, 2009.
- MEIRA, C. A. A.; RODRIGUES, L. H. A.; MORAES, S. A. de. Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. Pesquisa Agropecuária Brasileira, v. 44, n. 3, p. 233-242, 2009.
- MEIRA, C. A. A.; THAMADA, T. T.; HOLZHAUSEN, P. P. P. Avaliação do SAFCAFE – Sistema de Alerta da Ferrugem do Cafeeiro em três anos agrícolas. In: Congresso Brasileiro de Pesquisas Cafeeiras, 40., 2014, Serra Negra. Anais... Varginha: Fundação Procafé, p. 220-221, 2014.
- MORAES, S. A.; SUGIMORI, M. H.; RIBEIRO, I. J. A.; ORTOLANI, A. A.; PEDRO JUNIOR, M. J. Período de incubação de *Hemileia vastatrix* Berk. e Br. em três regiões do Estado de São Paulo. Summa Phytopathologica. Piracicaba, v. 2, n. 1, p. 32-38, 1976.
- OPITZ, D.; MACLIN, R. Popular Ensemble Methods: An Empirical Study. Journal of Artificial Intelligence Research, v. 11, p. 169-198, 1999.
- THAMADA, T. T.; GIROLAMO NETO, C. di; MEIRA, C. A. A. Sistema de alerta da ferrugem do cafeeiro: resultado de um processo de mineração de dados. In: Congresso Brasileiro de Agroinformática, 9., 2013, Cuiabá. Anais... Cuiabá: Editora UFMT, 2013.
- WU, X.; KUMAR, V.; QUINLAN, J. R.; GHOSH, J.; YANG, Q.; MOTODA, H.; MCLACHLAN, G. J.; NG, A.; LIU, B.; YU, P. S.; ZHOU, Z.; STEINBACH, M.; HAND, D. J.; STEINBERG, D. Top 10 algorithms in data mining. Knowledge and Information Systems, v. 14, p. 1-37, 2008.
- ZAMBOLIM, L.; VALE, F. X. R.; PEREIRA, A. A.; CHAVES, G. M. Café (*Coffea arabica* L.): controle de doenças – doenças causadas por fungos, bactérias e vírus. In: VALE, F. X. R. do; ZAMBOLIM, L. (Eds.) Controle de doenças de plantas: grandes culturas. Viçosa: UFV, 1997. p. 83-139.
- ZAMBOLIM, L.; VALE, F. X. R.; ZAMBOLIM, E. M. Doenças do cafeeiro. In: KIMATI, H.; AMORIM, L.; REZENDE, J. A.M.; BERGAMIN FILHO, A.; CAMARGO, L. E. A. (Eds.) Manual de Fitopatologia: Doenças das Plantas Cultivadas. 4th ed. volume 2. São Paulo: Agronômica Ceres. 2005. 663 p.