

Plano de Trabalho de Bolsista de Treinamento Técnico

Projeto: Um Protocolo para Ajuste e Seleção de Modelos de Abundância de Espécies

Responsável: Paulo Inácio de Knecht López de Prado, IB-USP]

Resumo

A bolsa de Treinamento Técnico solicitada destina-se a um profissional da área de Ciências Exatas com conhecimento e experiência em programação nas linguagens R e C. O objetivo deste plano é o desenvolvimento de um pacote em ambiente R que implemente os protocolos analíticos propostos no plano de pesquisa. Embora a equipe seja capaz de programar funções em R para uso próprio, sua depuração, otimização e integração em um pacote R, que tem um rígido controle de qualidade, é uma tarefa que demanda especialidade em programação, em dedicação exclusiva.

1 Objetivo

Criação de um pacote do ambiente R para ajuste e seleção de modelos de distribuição de abundância de espécies, e sua publicação no repositório público de pacotes da linguagem R (<http://www.r-project.org/cran>).

2 Metodologia

No ambiente R, um pacote consiste de um conjunto de objetos, classes e métodos, integrados para realizar um determinado tipo de análise estatística. O pacote aqui proposto irá implementar a estrutura lógica dos protocolos analíticos que serão desenvolvidos no projeto principal, que incluem ajustes, comparação e diagnóstico de modelos de distribuição de abundância, bem como simulações de comunidades sob modelos teóricos.

Por ser uma linguagem orientada a objetos, cada função pode ser programada, depurada e testada como uma unidade independente. Outra conveniência é a existência de interface com outras linguagens, como C e FORTRAN, o que permite otimização de rotinas, por meio do uso da linguagem mais adequada. Assim, um pacote pode ser construído em módulos, que são os objetos (Venables and Ripley, 2000), e novas funções podem ser adicionadas facilmente, na medida da necessidade dos usuários. Por outro lado, a unidade de um pacote depende de uma estrutura definida *a priori*, que reflete a lógica das análises implementadas. Em termos de programação, isto se dá pela definição das classes de objetos e seus respectivos métodos, que estabelecem as relações entre eles.

Por fim, os pacotes em R seguem uma padronização muito rigorosa, que define a estrutura de subdiretórios, documentação, parâmetros de configuração (R Development Core Team, 2008), que são verificados automaticamente pelo servidor do repositório de pacotes. A inclusão de um novo pacote neste repositório só é possível com esta verificação, garantindo um padrão mínimo de qualidade.

Combinando estes aspectos da linguagem R, o desenvolvimento do pacote aqui proposto seguirá as seguintes etapas:

Definição da estrutura: a estrutura geral do pacote, com suas funções, classes e métodos será definida, a partir da lógica do protocolo analítico que será desenvolvido no projeto principal. Para isto, o bolsista realizará reuniões com a equipe e o pesquisador principal, além de estudo individual do código e documentação de pacotes correlatos. Definida uma estrutura, o bolsista a apresentará a toda a equipe, para discussão e aperfeiçoamento. As diretrizes gerais para planejamento de pacotes seguirão as duas referências básicas da linguagem (R Development Core Team, 2008; Venables and Ripley, 2000). O produto esperado desta fase é o plano do pacote.

Programação: as funções e demais objetos serão desenvolvidos seguindo a cadeia de dependências, e as necessidades de aplicação pela equipe. O produto desta fase são todas os códigos de todas as funções e demais objetos que irão compor o pacote.

Depuração e testes: cada função passará pelas rotinas básicas de depuração em R (Venables and Ripley, 2000), bem como por testes de uso pela equipe (ver rotinas de testes com simulação, no projeto principal). O produto desta fase são os códigos revistos e depurados.

Documentação: conforme os padrões de pacotes em R, será preparada uma página de documentação em inglês para cada objeto desenvolvido. Estas páginas de ajuda são parte obrigatória dos pacotes, e seguem um formato padronizado (R Development Core Team, 2008). O produto desta fase é o conjunto de páginas de documentação.

Finalização: documentação e funções serão organizadas na estrutura de diretórios exigida para compor um pacote, que inclui uma série de arquivos de configuração, controle e documentação (R Development Core Team, 2008). O produto desta fase será o pacote em sua primeira versão.

3 Cronograma de Resultados

PRODUTO	TRIMESTRE							
	1	2	3	4	5	6	7	8
Estrutura do pacote	X							
Códigos dos Objetos					X			
Códigos depurados						X		
Documentação							X	
Pacote								X

4 Justificativa

Toda a equipe do projeto é capaz de criar funções no ambiente R para uso próprio. No entanto, a integração e documentação destas funções em um pacote para distribuição para outros usuários exige a dedicação em tempo integral de um programador. Uma pessoa com este perfil também é necessária para um trabalho sistemático de depuração e otimização das funções desenvolvidas, o que muitas vezes exige conhecimento de outras linguagens computacionais. Assim, a bolsa de Treinamento Técnico nível IV solicitada permite que a equipe de pesquisadores do projeto se ocupe com as questões de lógica e fundamentação ecológica dos protocolos, deixando aspectos técnicos de implementação para um profissional com competência específica.

O bolsista se beneficiará de um treinamento no desenvolvimento de pacotes em ambiente R, cuja aceitação no mercado e na academia tem crescido rapidamente. A interação com um grupo de usuários com demandas analíticas complexas e conhecimento básico de programação aumenta as possibilidades

de diálogo, e melhor compreensão dos perfis de potenciais clientes no meio acadêmico.

Referências

R Development Core Team, 2008. Writing R Extensions. R Foundation for Statistical Computing, Vienna, Austria.

Venables, W. and B. Ripley, 2000. S programming. Springer Verlag.