

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

RODRIGO BÜHLER

BIOMON: MODELO DE AMBIENTE PARA AUTOMAÇÃO DE ANÁLISE DE DADOS
EM BIOMONITORAMENTO

SÃO LEOPOLDO
2015

Rodrigo Bühler

BIOMON: MODELO DE AMBIENTE PARA AUTOMAÇÃO DE ANÁLISE DE DADOS
EM BIOMONITORAMENTO

Dissertação apresentada como requisito parcial
para a obtenção do título de Mestre pelo
Programa de Pós-Graduação em Computação
Aplicada da Universidade do Vale do Rio dos
Sinos — UNISINOS

Orientador:
Prof. Dr. Jorge L.V. Barbosa

São Leopoldo
2015

RESUMO

O biomonitoramento tem importante participação na compreensão das dinâmicas da natureza. Compreender a maneira como a vida reage às interferências externas ajuda na predição e tratamento dos efeitos fruto dessas alterações. Além disso, possibilita o entendimento sobre os processos de recuperação de áreas contaminadas, ajuda a prever efeitos das mudanças climáticas e até mesmo entender como é possível compartilhar o ambiente com diferentes espécies de forma a minimizar os danos. Com o objetivo de agregar valor na etapa de análise de dados de campo, o presente trabalho visa apresentar uma proposta de software para apoio ao pesquisador. O modelo apresentado contempla a interação entre o pesquisador responsável pela coleta e organização dos dados e o estatístico encarregado de elaborar e por vezes implementar agrupamentos de técnicas estatísticas que descrevem uma análise estatística. Levando em conta o fato de que a área de estudos do Biomonitoramento abrange diversas abordagens de interação com a fauna e flora, a presente proposta está modelada de forma que sua adaptação para diferentes métodos aconteça com um mínimo de esforço. Para tanto, são propostos módulos ou pacotes de funções para a interface, controle de ações e requisições do usuário, bem como execução da análise estatística. Como um meio para apresentar uma avaliação prática do modelo proposto, um protótipo foi implementado no ambiente de programação estatística R. Este protótipo foi utilizado em um teste de aceitação de tecnologia baseado no TAM - *Technology Acceptance Model*, onde pesquisadores convidados a utilizar o protótipo e responder a um questionário sobre a experiência. Desta forma foram geradas evidências que confirmam que os objetivos foram alcançados.

Palavras-chave: Biomonitoramento. Estatística. Método do elástico. Modelo de Equações Estruturais. O projeto R. Ecologia.

ABSTRACT

The biomonitoring has an important role in the understanding of the dynamics of nature. Understanding the way life reacts to external interferences helps us to predict and treat the effects result of these changes. In addition to that, it provides a better understanding of the steps of self-recovery processes in areas damaged by toxic accidents, regarding the weather changes or even knowledge on how it is possible to share the same space with different species, minimizing the damages. Aiming at adding value to the step of data analysis, the current dissertation presents a software proposal for supporting the researcher. The presented model covers the interaction between the researcher, who is responsible for collecting and organizing the data filed, and the statistician, who is in charge of defining and sometimes implementing sets of statistical techniques describing a statistical analysis. Taking into account that the Biomonitoring area of study covers different approaches in terms of fauna and flora interactions, the current proposal is modelled in a way that its adaptation to handle such a diverse set of options would happen with a minimum effort. For that, modules or packages with functions are designed to handle the interface, the user actions and requests, as well as the statistical analysis. As a way of presenting a practical evaluation of the proposed model, a prototype was implemented in the statistical development environment R. This prototype was applied in a technology acceptance test based on TAM - Technology Acceptance Model, where researchers were invited to execute the prototype and answer a survey about the experience. This way evidences were created confirming that the objectives were achieved.

Keywords: Biomonitoring. Statistics. Rubber band method. Structural Equation Modelling. The R project. Ecology.

LISTA DE FIGURAS

1	Etapas do Biomonitoramento e suas Áreas de Conhecimento	16
2	Visão geral do Biomon	34
3	Camada de Visões	35
4	Camada de Controladores	35
5	Camada do Modelo	36
6	Visão Lógica	37
7	Visão de Processos	38
8	Visão do Desenvolvimento	39
9	Visão Física	39
10	Sequencia de Passos	40
11	Editor de Textos	47
12	Planilha Eletrônica	47
13	Modelo de Matriz de Dados	51
14	Interface do Biomon app	52
15	Botões de <i>Upload</i>	53
16	Drop-down menu grupo <i>Analysis</i>	54
17	Drop-down menu grupo <i>Data Matrices</i>	54
18	Botão <i>Execute</i>	54
19	Diagrama da Classe Analyz	55
20	Diagrama de Sequência	56
21	Facilidade de Uso Avaliação	60
22	Facilidade de Uso Questões	60
23	Avaliação da Utilidade Percebida	61
24	Utilidade Percebida Questões	61

LISTA DE TABELAS

1	Comparativo entre os Trabalhos Relacionados	31
2	Leiaute do <i>Script</i> CSV	45
3	Tipo de Dado para Parâmetros	45
4	Variáveis Globais	46
5	Sample_Matrix_T.csv	48
6	Matrizes de Dados de Campo	51
7	Pesquisadores Entrevistados	57
8	Afirmações de itens Likert	58
9	Contribuição do modelo Biomon	65

LISTA DE ABREVIATURAS

ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
CCA	Canonical Correspondence Analysis
CSV	Comma Separated Values
GLM	General Linear Model
GPL	General Public License
GUI	Graphical User Interface
HTML	Hyper Text Markup Language
MANOVA	Multivariate Analysis of Variance
MVC	Model-Views-Controllers
PCA	Principal Component Analysis
SEM	Structured Equation Model
TAM	Technology Acceptance Model
UNISINOS	Universidade do Vale dos Sinos

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Motivação	15
1.2 Definição do Problema e Questão de Pesquisa	17
1.2.1 Definição do Problema	17
1.2.2 Questão de Pesquisa	18
1.3 Objetivos	18
1.4 Metodologia	19
1.5 Organização do Texto	19
2 FUNDAMENTAÇÃO TEÓRICA	21
2.1 Biomonitoramento	21
2.2 Coleta e Organização de Dados	22
2.2.1 Método do Elástico	22
2.3 Técnica Estatística	23
2.3.1 Modelo de Equações Estruturais (SEM)	23
2.4 Suporte Tecnológico	24
2.4.1 Linguagem e Ambiente de Computação Estatística R	25
3 TRABALHOS RELACIONADOS	27
3.1 Multiv & SYNCSA	27
3.2 Past	28
3.3 PC-ORD	28
3.4 CANOCO	29
3.5 IBM SPSS	29
3.6 Comparativo	30
4 BIOMON	33
4.1 Visão geral	33
4.2 Camadas do Biomon	34
4.2.1 Camada de Visões	34
4.2.2 Camada dos Controladores	34
4.2.3 Camada do Modelo	35
4.3 Arquitetura do Biomon	36
4.3.1 Visão Lógica	37
4.3.2 Visão de Processos	37
4.3.3 Visão do Desenvolvimento	37
4.3.4 Visão Física	38
4.3.5 Visão +1	39
4.4 Considerações sobre o capítulo	40
5 PROTÓTIPO	43
5.1 Ambiente de programação	43
5.1.1 The R Project	43
5.2 Arquitetura do Protótipo	43
5.3 O Estatístico	44
5.3.1 Arquivo de Análise	44
5.3.2 Caso de Uso	47

5.4 O Pesquisador	50
5.4.1 Matrizes de Dados	51
5.4.2 Interface Gráfica	52
5.4.3 Caso de Uso	52
5.5 Discussão	54
5.6 Código	55
6 AVALIAÇÃO	57
6.1 Metodologia	57
6.2 Caso de Teste	57
6.3 Questionário	58
6.4 Considerações sobre a Avaliação	59
6.5 Facilidade de Uso	59
6.6 Utilidade Percebida	59
7 CONSIDERAÇÕES FINAIS	63
7.1 Conclusão	63
7.2 Contribuições	64
7.3 Trabalhos Futuros	65
REFERÊNCIAS	67
APÊNDICE A CASO DE TESTE	71
APÊNDICE B AVALIAÇÃO	75

1 INTRODUÇÃO

Atualmente muito se discute a respeito do impacto do homem no meio ambiente. Diferentes correntes de pesquisadores procuram justificar seu ponto de vista com afirmativas baseadas nas mais variadas fontes e teorias. Como mencionado em (MARKERT; BREURE; ZECHMEISTER, 2003) é notório que o mundo contemporâneo passa por mudanças cada vez mais rápidas. Este mesmo autor destaca que o aumento de poluentes conhecidos e desconhecidos aliados com mudanças climáticas e a diminuição da biodiversidade se mostram uma ameaça para todos os ecossistemas. Isso revela que estes sistemas complexos apresentam inter-relações que demandam esforços na coleta e integração de diferentes fontes de informação, a fim de proporcionar ferramentas para seu acompanhamento.

A este acompanhamento se dá o nome de biomonitoramento onde a qualidade ambiental é avaliada a partir do acompanhamento de organismos presentes na área monitorada. Estes organismos são conhecidos como bioindicadores e biomonitores, (MARKERT; BREURE; ZECHMEISTER, 2003) e têm se mostrado excelentes ferramentas para o biomonitoramento.

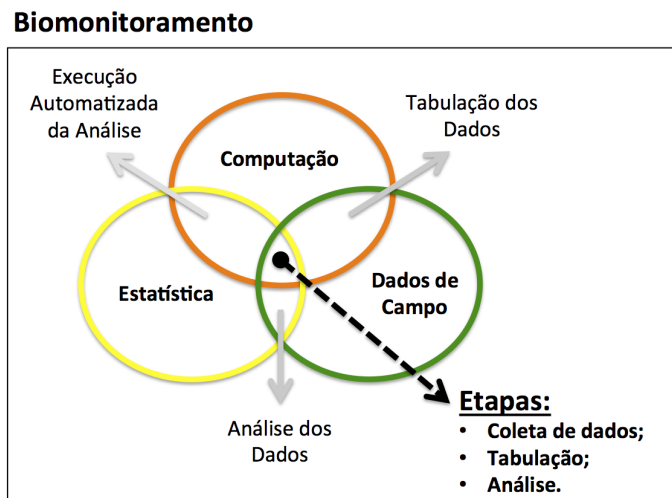
Idealmente são organismos encontrados em qualquer lugar facilitando a comparação entre diferentes regiões (CONTI; CECCHETTI, 2001). No caso dos organismos bioindicadores, são observados, suas famílias (PLATA; LÜCKING; LUMBSCH, 2008), função e *status* entre outras características. Desta forma qualquer alteração em relação ao que se conhece sobre estes organismos pode fornecer informações qualitativas do impacto no ambiente em questão. Já os organismos biomonitores têm como característica acumular poluentes do ambiente em sua composição, o que pode fornecer ao pesquisador um resultado quantitativo do impacto da poluição no ambiente ao realizar estudos na sua estrutura e composição.

1.1 Motivação

O biomonitoramento trata da avaliação da qualidade do ambiente por meio da coleta de dados de campo, contemplando informações de agentes conhecidos causadores de alterações ou estresse ambiental e dados dos organismos bioindicadores que respondem a estas alterações (ELLENBERG; WOODS-SCHANK; ENTWICKLUNGSTECHNOLOGIEN, 1991). Diferentes metodologias são empregadas nesta tarefa, o que implica nas mais variadas formas de registrar estas informações, como índices que representam uma determinada área em um ponto do tempo. Este estudo é efetuado em três etapas (Figura 1): coleta de dados em campo, tabulação dos dados e teste ou análise dos dados para obtenção de informações sobre o ambiente.

Informações como espécie de árvores, qualidade do ar, luminosidade, posição geográfica, análise química de componentes presentes em organismos como plantas e líquens são exemplos de variáveis observadas e catalogadas em uma pesquisa de campo para o biomonitoramento (LAVOREL et al., 2013). Após este levantamento, os dados são tabulados em matrizes que correspondem às variáveis ambientais, como por exemplo as variáveis de espaço e variáveis de

Figura 1: Etapas do Biomonitoramento e suas Áreas de Conhecimento



Fonte: Elaborada pelo autor

comunidades biológicas (KOCH, 2012), a partir daí a análise de dados é executada, independentemente do método utilizado na sua coleta.

Para a avaliação das variáveis os dados são submetidos a técnicas estatísticas que procuram demonstrar o relacionamento entre as variáveis que foram observadas em campo, a fim de encontrar padrões e respostas nestas relações, avaliar a representatividade da amostra coletada em relação à comunidade que esta representa, entre outros aspectos.

São alguns exemplos de técnicas estatísticas utilizadas no processamento dos dados de campo, a análise de variância univariada e multivariada, análise de regressão, análise de caminho (*path analysis*), Bayes e Modelo de Equações Estruturais (SEM), tendo cada uma um objetivo específico sobre uma mesma base de dados, ajudando o pesquisador a obter respostas para explicar o que é observado na natureza.

Para dar suporte à análise de grandes bases de dados ou mesmo como forma de automatizar esta tarefa, ferramentas estatísticas foram desenvolvidas em diferentes plataformas e linguagens de programação. Fornecendo as mais variadas formas de entrada de dados, configurações e representações gráficas destas análises, estas ferramentas de maneira geral apresentam um suporte generalista de forma a atender as mais variadas necessidades. Quando trata-se especificamente de biomonitoramento, ferramentas como Smart PLS (SMART PLS, 2014), SPSS (SPSS, 2014), Multiv (MULTIV, 2014) e o ambiente de computação estatística R (THE R PROJECT, 2014), entre outros, são frequentemente relacionados como ferramentas de apoio na análise dos dados.

O pesquisador de posse da descrição dos passos a serem seguidos em uma análise estatística, elaborados por um estatístico, faz uso do aparato tecnológico de software disponível a fim de chegar ao resultado desejado. De maneira geral os referidos softwares são desenvolvidos e distribuídos de forma a cobrir um assunto específico como por exemplo negócios (SPSS, 2014) ou ainda uma abordagem estatística específica como análise de variância em matrizes de dados ou

teste de aleatorização (MULTIV, 2014). Esta característica faz com que haja um esforço extra por parte do pesquisador que precisa fazer uso de diferentes ferramentas de software com seus próprios procedimentos de utilização como interface e entrada de dados para uma determinada sequência de passos descritos por uma análise.

Pesquisadores em biomonitoramento têm buscado suporte nestes diferentes softwares estatísticos a fim de obter respostas às suas questões de pesquisa. O fato de se tratarem de ferramentas especializadas em um determinado aspecto estatístico, faz com o que pesquisador precise elaborar uma combinação de diferentes programas ou ambientes estatísticos em sua pesquisa. Este fato acarreta em um esforço extra de preparação dos dados para diferentes formas de entrada de dados bem como a manipulação dos mais variados paradigmas de interface com usuário. A ausência de uma ferramenta dinâmica ambientada nas características do biomonitoramento faz com que o pesquisador precise dominar habilidades diferentes das inerentes a sua área de pesquisa.

1.2 Definição do Problema e Questão de Pesquisa

1.2.1 Definição do Problema

Considerando a inexistência de uma solução onde em um mesmo ambiente o pesquisador tenha condições de executar todos os passos de uma análise, surge a oportunidade da proposição de um software para suporte às etapas do biomonitoramento como entrada de dados, organização das técnicas estatísticas bem como sua execução. Este software deve requerer do pesquisador habilidades inerentes a pesquisa e um mínimo de complexidade tecnológica, transferindo para o estatístico a elaboração das análises estatísticas e para a arquitetura proposta meios de promover esta cooperação.

Para tal, algumas características se mostram estratégicas na tarefa de disponibilizar o software proposto:

- elaborar um protocolo onde é descrita a forma como o estatístico deve informar as técnicas estatísticas e seus parâmetros a fim de criar arquivo de análise estatística para ser lido e interpretado pelo software;
- entregar ao usuário uma interface adaptada aos termos do biomonitoramento onde este deve informar os dados de campo (LEGENDRE; LEGENDRE, 2012), bem como o arquivo elaborado pelo estatístico com a análise a ser executada (KOCH, 2012);
- disponibilizar um conjunto de técnicas estatísticas para serem acessadas pelos passos descritos pelo estatístico, como por exemplo análise de variância multivariada (MANOVA) e teste de aleatorização Monte-Carlo, Bayes e SEM e aplicadas sobre os dados informados pelo pesquisador como em (PILLAR; ORLÓCI, 1996) e (MCCUNE; GRACE; URBAN, 2002);

- como forma de promover a adaptabilidade do software em novas versões para atendimento a diferentes segmentos da área de pesquisa é fundamental a modularização de sua arquitetura como forma de reduzir o impacto do desenvolvimento de tais novas opções.

1.2.2 Questão de Pesquisa

Frente a inexistência de uma solução que atenda as características da análise de dados de campo em biomonitoramento, onde haja suporte à entrada de dados e flexibilidade na elaboração de análises estatísticas em um ambiente único, surge a seguinte questão de pesquisa: "Poderia tal ambiente efetivamente agregar valor ao trabalho do pesquisador na execução do biomonitoramento, possibilitando a este a combinação em uma mesma interface da entrada de dados de campo e a escolha de diferentes análises estatísticas específicas para execução sobre estes dados?"

1.3 Objetivos

O objetivo geral desta trabalho é a modelagem de uma proposta de software para suporte a análise de dados em biomonitoramento, onde é possível a elaboração e execução de uma sequência de passos estatísticos em um mesmo ambiente. Além deste do suporte ao pesquisador o modelo deve apresentar uma modelagem que fomente sua extensão ou adaptação, com baixo impacto no software como um todo, para a cobertura de diferentes áreas do biomonitoramento. Para que este objetivo seja alcançado foram definidos os seguintes objetivos específicos:

1. Descrever a fundamentação teórica destacando a etapa de coleta de campo e organização dos dados bem como o procedimento de elaboração dos conjuntos de técnicas estatísticas;
2. Apresentar trabalhos relacionados na forma de ferramentas computacionais utilizadas na organização dos dados e na execução dos métodos estatísticos definidos para atender a determinada abordagem adotada pelo pesquisador;
3. Detalhar o modelo de software proposto de forma a ilustrar a maneira como seu desenho e implementação podem responder à questão de pesquisa e resultar em um software realmente útil para a análise de dados de campo com características que o habilitam para ser facilmente estendido cobrindo, desta forma, diferentes áreas da pesquisa;
4. Descrever a implementação de um protótipo funcional a partir do modelo proposto em um ambiente de computação estatística R (THE R PROJECT, 2014), protótipo este desenvolvido com o propósito de servir como prova de conceito;
5. A partir da utilização do protótipo por pesquisadores da área do biomonitoramento executar uma avaliação de aceite de software com estes a fim de avaliar sua experiência na utilização do Biomon;

6. Apresentar as considerações finais em relação ao cumprimento de cada etapa e fazer um balanço geral dos resultados alcançados e trabalhos futuros.

1.4 Metodologia

A fim de viabilizar o modelo proposto, foi realizada uma pesquisa com o objetivo de identificar os passos executados pelo pesquisador quando da avaliação da qualidade ambiental de uma área por meio de biomonitoramento. A pesquisa procurou cobrir as etapas e conceitos envolvidos, bem como publicações relacionadas com biomonitoramento, onde foram aplicadas de forma prática os métodos e técnicas discutidos no presente trabalho.

A partir deste estudo foi proposto o Biomon, um modelo de software que permite a cobertura das mais variadas estruturas de dados, e técnicas estatísticas para análise destes dados, em um ambiente único ambientado nos termos e necessidades da área do biomonitoramento.

Como forma de responder a questão de pesquisa proposta foi implementado um protótipo no ambiente de computação estatística R (THE R PROJECT, 2014), que foi avaliado por pesquisadores convidados da área de biomonitoramento. Seu parecer quanto a utilização do protótipo em uma situação real serviu de método de validação do modelo proposto.

1.5 Organização do Texto

Esta dissertação está dividida em seis capítulos. O capítulo um cobre a presente introdução já o capítulo dois contém a fundamentação teórica das etapas seguidas no biomonitoramento. No capítulo três são apresentados trabalhos relacionados, na forma de softwares de estatística, que são aplicados em situações reais de análise de dados de campo por pesquisadores. O quarto capítulo se refere ao detalhamento do modelo proposto em sua arquitetura e desenho. No capítulo cinco será apresentada a implementação de um protótipo funcional implementado com a função de servir como prova de conceito e utilizado em uma avaliação de aceite de software detalhada no capítulo seis. Já no capítulo sete serão demonstrados os resultados obtidos em um balanço final e listados possíveis trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

De maneira a contextualizar a presente proposta as seções a seguir irão cobrir conceitos de biomonitoramento como coleta e organizações de dados coletados em campo bem como as análises estatísticas e suporte tecnológico disponível para tratamento destas informações. O contexto abordado se refere ao ambiente no qual Biomon está inserido e propõe suporte.

Com o objetivo de apresentar um caso de uso real será apresentado um método de coleta e análise de dados em biomonitoramento onde é feito o estudo dos níveis de degradação e recuperação de áreas de floresta com a utilização de líquens como biomonitores. Segundo literatura especializada os líquens são os organismos mais adequados para esta tarefa (CONTI; CECCHETTI, 2001). Já para demonstrar a forma como técnicas estatísticas são empregadas no biomonitoramento será abordada uma técnica de teste de hipóteses com a utilização do conceito de análise de caminho (*path analysis*).

Quanto ao suporte tecnológico serão numeradas algumas opções de software disponíveis e comumente referenciadas em material científico como em (KOSICKI; CHYLARECKI, 2012) e (GENTLEMAN et al., 2004). Importante observar que em um mesmo estudo, por vezes, é possível notar a utilização de diferentes softwares complementares a fim de executar as técnicas estatísticas descritas por uma análise.

2.1 Biomonitoramento

O biomonitoramento trata da identificação das reações dos organismos frente as interferências em seu hábitat (ELLENBERG; WOODS-SCHANK; ENTWICKLUNGSTECHNOLOGIEN, 1991). Diferentes bioindicadores, ambientes e agentes poluentes são observados, como em (CONTI; CECCHETTI, 2001) onde líquens são utilizados como biomonitores para avaliar níveis de poluição ou como em (PLATA; LÜCKING; LUMBSCH, 2008) que também fez uso de líquens, mas desta vez para determinar se dada floresta está livre de distúrbios ou interferências de agentes externos.

Independentemente do objetivo, estudos em biomonitoramento seguem uma estrutura semelhante, onde o pesquisador propõe uma hipótese para um efeito a ser observado. A partir de então procura com a ajuda de organismos bioindicadores, relacionar alterações nestes organismos com as interferências monitoradas. As características a serem analisadas no ambiente alvo definem tanto o organismo a ser coletado ou observado quanto os atributos a serem estudados. Além disso são coletados dados das chamadas variáveis ambientais como em (HEPP et al., 2010) onde além do bioindicador, um microinvertebrado do fundo de rios, são coletadas informações como velocidade da corrente do rio, temperatura da água, nível de oxigênio, pH, entre outros.

As coletas são concentradas em diferentes áreas ou parcelas (*plots*), de forma a produzir evidências que caracterizem cada ponto e proporcionem subsídio para comparações e inferên-

cias a partir das diferenças. Além disso, também é comum a definição de uma área controle como em (HEPP et al., 2010) e (SILVA et al., 2000) onde foram definidas regiões referência para comparação com as demais.

Embora compartilhem uma mesma estrutura, cada método descreve à sua maneira a forma como os dados serão coletados em campo. Um maior detalhamento a respeito da coleta de dados e análise destes serão descritos nas seções seguintes.

2.2 Coleta e Organização de Dados

Diferentes métodos de coleta de dados em campo podem ser encontrados em publicações da área da ecologia com ênfase em biomonitoramento. Variações em áreas como florestas, lagos ou desertos, em unidades amostrais tais como árvores, pedras, água doce ou salgada e, de diferentes bioindicadores como líquens, peixes, organismos invertebrados ou bactérias, determinam a forma como os dados serão coletados. Por vezes, pesquisas semelhantes como (JEON et al., 2008) e (HEPP et al., 2010) apresentam metodologias diferentes.

Em (JEON et al., 2008) a proposta se refere a um sistema de alarme quanto ao nível de poluição da água. O método proposto é a análise do comportamento de um agente bioindicador por meio da análise de imagens registradas por filmadoras ou *frame grabbers*. Já (HEPP et al., 2010) procura avaliar o impacto da agricultura e urbanização em rios por meio da coleta e análise de um determinado organismo bioindicador. Estes dois exemplos ilustram a necessidade de suporte a diferentes entradas de dados para uma mesma finalidade de análise tendo em vista que as técnicas estatísticas empregadas para análise destes dados são, senão idênticas, semelhantes.

2.2.1 Método do Elástico

De maneira a ilustrar com mais detalhes uma metodologia de coleta de dados, foi escolhido o método conhecido como Método do Elástico proposto em (MARCELLI, 1992), que trata especificamente de estudos com líquens. A partir de uma hipótese inicial este método propõe a avaliação da cobertura de líquens (bioindicador) nas árvores (unidades amostrais) como subsídio para ao teste de hipótese.

Aos dados da coleta, sugere o autor, devem ser agregadas informações adicionais do ambiente onde se encontram as unidades amostrais, tais como luminosidade, humidade e velocidade do vento, entre outros. Também são coletados dados a respeito da própria unidade amostral, como pH e cor da casca, todos estes dados são tabulados e relacionados seguindo a metodologia proposta.

Deste relacionamento são elaboradas tabelas de variáveis ambientais, de onde se parte para a transformação dos dados em informações com a aplicação das mais variadas técnicas estatísticas. Em (KÄFFER; GANADE; MARCELLI, 2009) são relacionadas as técnicas: Análise de Co-Variância (ANCOVA), Modelo de Regressão Linear (GLM) e Coeficiente de Sørensen.

A próxima seção trata das técnicas estatísticas, utilizadas no estudo dos dados de campo. Mais especificamente será tratada a técnica conhecida como Modelo de Equações Estruturais (SEM) utilizada no estudo de relações de causa-efeito, ideal no entendimento da contribuição de cada uma das variáveis na degradação ou recuperação de dada área.

2.3 Técnica Estatística

A estatística tem sua contribuição no apoio ao desenvolvimento de modelos e suporte à predições na ciência, e na sua confrontação com dados reais observados em campo (GOTELLI et al., 1995). Estudos científicos têm como objetivo observar causas para fenômenos observáveis e a coleta de dados é um meio para este fim. A partir desta coleta é formulada uma hipótese que por sua vez gera predições, estas últimas são testadas com novas rodadas de observações. Em situações onde as predições não corroboram o que foi proposto, a hipótese precisa ser revista e o processo se reinicia.

Na tarefa de entender ou explicar os fenômenos observados, as análises de causa-efeito têm papel fundamental, e podem ser executadas por meio de diferentes metodologias. Entre as principais técnicas é possível mencionar *Path Analysis* (SHIPLEY, 2009), Bayes (ARHONDITSIS et al., 2007) e SEM (WALKER et al., 2013).

2.3.1 Modelo de Equações Estruturais (SEM)

A técnica de modelagem de relações de causa-efeito SEM, sigla em inglês para Modelo de Equações Estruturais, fornece ao pesquisador uma ferramenta para quantificar e testar modelos teóricos (PUGESEK; TOMER; VON EYE, 2003). A partir da definição dos parâmetros que demonstram a magnitude do efeito de uma relação direta ou indireta entre variáveis de um modelo, o pesquisador pode avaliar uma proposta de hipótese. A este processo se dá o nome de Aspecto Confirmatório de SEM, já a repetição deste procedimento com novos dados proporciona a possibilidade de análise do impacto destes no modelo, este é o chamado Aspecto Exploratório de SEM.

A principal diferença entre SEM e outras técnicas semelhantes, é o suporte a variáveis latentes, este tipo de variável não pode ser observada diretamente na coleta de dados mas é caracterizada de forma indireta sendo composta por dados de outras variáveis que associadas.

Em (MAZANCOURT et al., 2013), os autores propõem que um ecossistema se mantém estabilizado em razão da biodiversidade por meio de três mecanismos: a resposta ambiental, a demografia e um baixo valor de erro nas observação. Diferentes técnicas estatísticas foram aplicadas sobre bases de dados preexistentes de diferentes regiões, como forma de caracterizá-las. A partir de então foram elaborados modelos SEM onde os mecanismos propostos foram testados quanto a sua participação no fenômeno observado.

Já (LAVOREL et al., 2013) faz uso de SEM como forma de identificar os fatores da gestão

de pastagens que possuem maior impacto na polinização de plantas. Os autores propõem uma metodologia para identificação desta relação, a partir de então elaboram um modelo SEM de efeito-resposta e testam os dados coletados em campo.

Os trabalhos citados são alguns exemplos da utilização de SEM em biomonitoramento, com o objetivo de explicar a participação de diferentes fatores em um dado fenômeno.

2.4 Suporte Tecnológico

Para o suporte a análise de dados, diferentes opções de software estatístico podem ser encontradas. De maneira geral estes softwares fornecem suporte a uma abordagem ou conceito estatístico oferecendo ao usuário diferentes paradigmas de operação. Desta forma o pesquisador que necessita aplicar diferentes conceitos estatísticos em uma análise precisa também interagir com diferentes contextos de software.

Em (PILLAR; ORLÓCI, 1996) os autores, no papel de pesquisador e estatístico, implementaram suas ferramentas, o Multiv (MULTIV, 2014) e o SYNCSA (SYNCSA, 2014), e as utilizam como forma de sustentar sua proposta. Este é um exemplo de implementação de software dedicado ao teste de hipótese, abordando o conceito sustentado pelo autor. A medida que diferentes índices, como *t*Student ou Chi^2 , ou técnicas, como Bayes ou SEM, sejam agregadas à análise dos dados de campo, outras ferramentas precisam ser agregadas ao trabalho do pesquisador, como em (TER BRAAK, 1992) que fez uso do software CANOCO (CANOCO, 2014) em suas pesquisas.

Biomon propõe um modelo de software onde o estatístico e o pesquisador interagem colaborativamente para elaborar um arquivo descrevendo a integração de técnicas estatísticas e seus parâmetros que deverá ser interpretado pelo software proposto e executado, juntamente com os dados de campo informados pelo pesquisador. A necessidade desta integração pode ser observada em (PLATA; LÜCKING; LUMBSCH, 2008), onde são aplicadas técnicas como *One-way ANOVA* e *Scheffé post-hoc* bem como Análise de Componentes Principais (PCA) com o auxílio de dois diferentes softwares: PC-ORD (PC-ORD, 2014) e Statistica (STATISTICA, 2014), cada um com sua própria interface de funcionamento e forma de entrada de dados.

Procurando oferecer uma alternativa para a implementação do Biomon, a próxima seção apresenta o ambiente R (THE R PROJECT, 2014) de computação estatística, utilizado na produção acadêmica de diversas áreas, desde análises de integração das funções do cérebro (ROELSTRAETE; ROSSEEL, 2011), passando por tratamento de imagens digitais (POLZEHL; TABELOW, 2006) e biologia (BONNET; PEER, 2002). O ambiente R (THE R PROJECT, 2014) fornece um conjunto de bibliotecas para tratamento estatístico de dados nativo, bem como pacotes desenvolvidos por outros desenvolvedores, com isso não há a necessidade de re-escrever as técnicas já suportadas. Desta forma a presente proposta pode limitar-se a descrever como a integração destas técnicas estatísticas em uma análise é suportada e integrada em um ambiente onde o pesquisador informa seus dados e escolha a abordagem estatística que melhor se adequa

a sua necessidade.

2.4.1 Linguagem e Ambiente de Computação Estatística R

R (THE R PROJECT, 2014) é uma linguagem e ambiente para computação estatística e produção de gráficos, considerada uma implementação de código aberto de outra linguagem, chamada S desenvolvida inicialmente na Bell Laboratories, atualmente Lucent Technologies. O R é atualmente distribuído como software livre sob os termos da licença GPL da *Free Software Foundation*.

A interface com o usuário é essencialmente em linha de comando que, por ser um software de arquitetura aberta, pode ser incrementado com a adição de novos pacotes que estendem sua gama de possibilidades. A implementação destes pacotes exige do desenvolvedor conhecimentos técnicos de desenvolvimento, podendo inclusive agregar ao seu pacote a integração com códigos de outras linguagens como C, C++ e Fortran, por exemplo.

Quanto a operação, o R (THE R PROJECT, 2014) permite ao usuário a execução individual de cada cálculo, ou a criação de um arquivo *script* onde diferentes comandos são encadeados e executados sequencialmente. Com isso é possível, com poucos conhecimentos técnicos específicos, automatizar uma rotina de análise de dados abrangendo diversas técnicas estatísticas.

Algumas iniciativas como em (RSTUDIO, 2014), procuram tornar a utilização do R (THE R PROJECT, 2014) uma tarefa mais intuitiva disponibilizando uma interface gráfica tanto para a sua operação, quanto desenvolvimento de novos pacotes.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta propostas de softwares que são utilizados em trabalhos de biomonitoramento como ferramenta principal, ou em conjunto com outras na análise dos dados de campo. Entre os softwares estatísticos disponíveis, foram escolhidas opções aplicadas na análise de dados em biomonitoramento com suporte a matrizes de dados univariados e multivariados ou mesmo apenas suporte estatístico à algum passo necessário para a análise de dados de campo.

Algumas das ferramentas apresentadas a seguir foram desenvolvidas por pesquisadores como apoio às suas publicações, e a partir de então utilizadas em outros trabalhos, sendo aplicadas nos mais diferentes assuntos. A intenção é promover uma comparação entre os trabalhos relacionados usando critérios que permitam avaliar a contribuição proposta pelo Biomon.

3.1 Multiv & SYNCOSA

No artigo de (PILLAR; ORLÓCI, 1996) os autores propõem a implementação de um algoritmo para a execução de aleatorização de testes para confirmação de hipótese. Desta proposta surgiram os softwares Multiv (MULTIV, 2014) para análise exploratória multivariada e o SYNCOSA (SYNCOSA, 2014) para análise de correlação de matrizes. Fazendo uso de uma ou mais técnicas estatísticas estes softwares podem ser utilizados isoladamente ou em conjunto com softwares complementares como em (CRUZ et al., 2010), (CARBONI; ACOSTA; RICOTTA, 2013) e (HEPP et al., 2010).

Ambos Multiv (MULTIV, 2014) e SYNCOSA (SYNCOSA, 2014) são distribuídos livremente nas plataformas Apple Mac OS TM e MS Windows TM (com limitações quanto a geração de gráficos na versão MS Windows TM). Ambos softwares têm sua operação por linha de comando onde são fornecidas opções ao usuário para configuração, importação dos dados e execução das análises disponíveis.

A entrada de dados é feita por um arquivo texto contendo uma matriz de variáveis observadas pelas unidades amostrais. A partir da entrada dos dados algumas operações podem ser efetuadas nestes a fim de editar os títulos das linhas ou colunas e a partir daí as análises estatísticas são executadas.

Diferente do Multiv (MULTIV, 2014), o SYNCOSA (SYNCOSA PACKAGE FOR R, 2014) teve implementada uma versão para o ambiente R (DEBASTIANI; PILLAR, 2012) suportando as funcionalidades desenvolvidas anteriormente neste novo ambiente, desta forma estando disponível em uma gama maior de plataformas.

3.2 Past

O artigo (HAMMER; HARPER; RYAN, 2001) trata do software estatístico para paleontologia Past (PAST, 2014), onde os autores procuram reunir métodos para auxílio em pesquisa levando em conta o aspecto quantitativo. Apesar do contexto dirigido à paleontologia, os métodos cobertos são conhecidos e aplicados nas análises em biomonitoramento, mas ainda assim não são facilmente encontrados em distribuições de software estatístico padrão. Entre suas funcionalidades está o suporte a dados univariados e multivariados bem como análises temporais de dados, análise filogenética e representação gráfica.

Um exemplo da utilização do Past (PAST, 2014), em conjunto com outro software, pode ser observado em (KNEITEL; LESSIN, 2010).

Past (PAST, 2014) é distribuído livremente para plataforma MS WindowsTM apenas. Sua interface gráfica apresenta ao usuário uma tabela onde os dados podem ser informados manualmente ou importados a partir de um arquivos texto, com diferentes opções de quebras como forma de identificar as colunas. Depois de informada a base de dados, as operações ocorrem por meio do acesso ao menu *pull-down*, onde são aplicadas as técnicas ou funções escolhidas sobre os dados informados. Algumas funções disponibilizam ao usuário, ajustes nos seus parâmetros a fim de controlar sua execução ou apresentação.

3.3 PC-ORD

O software PC-ORD (PC-ORD, 2014) é distribuído comercialmente para a plataforma MS WindowsTM, apesar de haver uma versão chamada *demo* esta é limitada a apresentar a interface e algumas funcionalidades da ferramenta, nenhum cálculo é executado. Seu manual (MCCUNE; MEFFORD, 2011), atualmente na versão 6, descreve de forma geral seu funcionamento, em relação a sua operação o software foi construído com seu sistema de ajuda sensível ao contexto provendo suporte adequado às dificuldades do usuário.

Com foco na análise de dados ecológicos, disponibiliza uma variedade de métodos estatísticos para operação com dados multivariados. A terminologia utilizada é adaptada à área da ecologia, o que permite ao pesquisador reduzir a curva de aprendizado fornecendo uma interface familiar ao contexto da pesquisa.

A partir do menu principal é possível importar as matrizes com os dados em vários formatos, como arquivos texto e planilhas MS ExcelTM.

Assim como os outros softwares já mencionados, PC-ORD (PC-ORD, 2014) cobre um contexto limitado de técnicas estatísticas, com isso pode haver a necessidade de ser utilizado em combinação com outra ferramenta como em (PLATA; LÜCKING; LUMBSCH, 2008).

3.4 CANOCO

Em (BLASIUS; GREENACRE, 2014), o autor apresenta um relato de sua proposta de utilização da Análise de Correspondência Canônica (CCA) no biomonitoramento. Neste mesmo artigo é apresentado o software CANOCO (CANOCO, 2014) desde sua origem, descrevendo o suporte a esta técnica de ordenação e suas variantes.

CANOCO (CANOCO, 2014) é comercializado em versões para a plataforma MS WindowsTM, atualmente se encontra na versão 5, frequentemente mencionado em artigos que fazem uso das técnicas de ordenação, como se pode observar em (ALCARAZ; CAIOLA; IBÁÑEZ, 2011), (AYUKE et al., 2011) e (WILSON et al., 2009).

A entrada de dados é feita a partir de uma planilha MS ExcelTM, onde diferentes matrizes de dados podem ser importados simultaneamente a partir de diferentes abas de uma mesma planilha. No geral sua operação acontece da mesma forma que os demais softwares similares, a partir de menus *pull-down* e janelas *popup* onde as análises podem ser parametrizadas.

3.5 IBM SPSS

Como parte do pacote de soluções analíticas da IBMTM para o mercado corporativo, o SPSS (SPSS, 2014) é comumente mencionado em artigos de ecologia no apoio a análise de dados. Apesar de ter foco em análise de dados de negócio, possui suporte a muitos métodos estatísticos divididos em diferentes módulos que podem ser abstraídos para a ecologia. Como exemplo pode-se mencionar o pacote AMOS que trata de Modelo de Equações Estruturais, onde é possível estabelecer um modelo de relação de causa e efeito, e sua parametrização, e a partir disso avaliar o modelo proposto com dados reais, tanto de uma pesquisa de campo como de um negócio ou produto.

Apesar da interface promover ao usuário uma experiência intuitiva, o conjunto de soluções analíticas SPSS (SPSS, 2014), por seu foco na área do negócios, exige do pesquisador conhecimento técnico tanto da ferramenta quanto de suas funcionalidades. Desta forma, podendo adaptar sua necessidade à execução da análise correta no pacote específico.

O SPSS (SPSS, 2014) é um software comercial licenciado, e portanto, precisam ser adquiridas licenças de uso para sua operação. Cada pacote da solução possui seus próprios requerimentos em termos de sistema operacional, o que pode, em alguns casos limitar sua instalação a um determinado ambiente. Assim como alguns dos demais softwares já mencionados, SPSS (SPSS, 2014) também importa as matrizes a partir de uma planilha eletrônica, bem como de arquivos texto ou outras fontes.

3.6 Comparativo

Como forma de apresentar um comparativo entre os trabalhos relacionados, aqui representados por softwares de análise de dados, a tabela 1 apresenta como cada software atende aos critérios adotados e descritos a seguir:

- **Interface com o usuário:** No que se refere a experiência do usuário com a utilização de um software esta se apresenta mais intuitiva quando há acesso a uma interface gráfica com campos de entrada de dados devidamente descritos por seus títulos, botões para acesso a funções e as mais variadas formas de *menus* para seleção de opções. Para tanto o Biomon dedica uma camada (biblioteca ou pacote) especificamente a este conceito com forma de apresentar ao usuário uma experiência com um mínimo de complexidade tecnológica, de forma que o pesquisador não precise necessariamente decorar listas de comandos para executar alguma ação ou lidar com mensagens de erro ou alerta de difícil interpretação enviados pelo software por imperícia do usuário.

Para a avaliação da interface com o usuário serão considerados os seguintes aspectos:

- **Linha de comando:** para os casos onde o acesso do usuário as funções do software devem ser feitas por linha de comando em um terminal;
- **Interface gráfica:** para os casos onde o acesso é feito por meio de interface gráfica com janelas, botões e menus.
- **Suporte:** A arquitetura do Biomon é dirigida para suporte a um contexto de análise composto por dados e um conjunto de técnicas estatísticas. Com o objetivo de auxiliar o pesquisador na análise de dados de campo sem que este tenha que lidar com diferentes soluções de software, cada um com seu próprio paradigma de operação. O Biomon depende apenas de elaboração dos arquivos de análise, criados pelo estatístico ou compartilhado entre os pesquisadores, para cobertura de diferentes abordagens estatísticas em um mesmo ambiente.

Para avaliação do tipo de suporte fornecido pelos trabalhos relacionados, softwares, serão considerados os seguintes aspectos:

- **Contexto de análise:** para os trabalhos relacionados, que suportam um contexto de análise ou assunto disponibilizando métodos estatísticos para suporte a este, proporcionando ao pesquisador um ambiente único para análise de seus dados;
- **Método estatístico:** para os trabalhos relacionados, que atendem a um método ou abordagem estatística desta forma podendo ser aplicado em análises estatísticas de diferentes contextos em combinação com outros métodos.
- **Extensibilidade:** A presente proposta procura fornecer ao usuário uma experiência próxima a sua área de atuação, seja com o desenvolvimento de uma instância de software

direcionada a um contexto ou propondo uma arquitetura que fomente o desenvolvimento de novas versões da referida instância com o reaproveitamento de camadas (pacotes ou bibliotecas) comuns de software. O que geralmente se observa é o desenvolvimento de soluções de software dedicadas a uma área foco ou abordagem estatística fim, onde a arquitetura geralmente é proprietária não permitindo desta forma sua extensão ou adaptação para outros fins. Apesar disso, alguns exemplos implementados sob licenças de domínio público tem sua estrutura acessível à comunidade de desenvolvedores, mas ainda assim não se apresentam devidamente preparados para serem facilmente estendidos ou adaptados para áreas diferentes das que foi originalmente designado.

Para a avaliação da extensibilidade do software serão considerados os seguintes aspectos:

- **Código aberto:** para softwares onde o acesso ao código é de domínio público podendo ser gerada uma nova instância do mesmo software e aplicadas adaptações ou ajustes por terceiros;
- **Código proprietário:** para softwares onde o acesso ao seu código seja controlado pela equipe de responsável ou empresa proprietária dos direitos autorais, desta forma não sendo possível o acesso direto pela comunidade de desenvolvedores para cópia ou adaptação.

Tabela 1: Comparativo entre os Trabalhos Relacionados

Software	Interface	Suporte	Extensibilidade
Multiv & SYNCSA	Linha de comando	Método estatístico	Código aberto
Pct.SYNCSA para R	Linha de comando	Método estatístico	Código aberto
Past	Interface gráfica	Contexto de análise	Código proprietário
PC-ORD	Interface gráfica	Método estatístico	Código proprietário
CANOCO	Interface gráfica	Método estatístico	Código proprietário
IBM SPSS	Interface gráfica	Método estatístico	Código proprietário

Fonte: Elaborada pelo autor

A interface e seus componentes por si só não podem caracterizar um software como adequado ou não para utilização pelo pesquisador, tanto isto é válido que mesmo ferramentas com operação por linha de comando ainda são acessadas e desenvolvidas para análise de dados de campo, como por exemplo Multiv (MULTIV, 2014) e SYNCSA (SYNCSA, 2014). Apesar de requerer uma curva de aprendizado maior que soluções gráficas, seu suporte a abordagens estatísticas ambientadas na área do biomonitoramento ainda se apresenta como solução para suporte a assuntos não cobertos por soluções com uma interface mais intuitiva.

Em relação ao suporte fornecido ao pesquisador, o presente modelo procura contextualizar sua operação em uma área da pesquisa proporcionando meios para que todo suporte estatístico necessário seja adicionado de maneira dinâmica. Como resultado o pesquisador não precisa

recorrer a diferentes soluções de software com diferentes interfaces, diferentes abordagens de entrada de dados e metodologias de operação.

Além dos aspectos práticos observados como interface com o usuário e tipo de suporte estatístico oferecido, Biomon prima por uma solução que tecnicamente possa ser estendida para diferentes abordagens de análise de dados de campo em biomonitoramento. Esta extensão compreende a aplicação de adaptações no software original para suporte, por exemplo, a uma nova distribuição de dados de entrada resultando em uma nova instância do Biomon. Apesar de algumas soluções de código livre terem sido listadas nenhuma delas apresenta como característica uma arquitetura dedicada a fomentar sua extensão com baixo impacto no software original. Detalhes do modelo proposto serão descritos nos próximos capítulos.

4 BIOMON

Como forma de organizar a apresentação do modelo as próximas seções serão representadas pelos conceitos de *blackbox* e *whitebox*, estabelecidos por (ROSENBLUETH; WIENER; BIGELOW, 1943). O primeiro se refere a execução de uma unidade sem que se saiba seu funcionamento interno, desta forma a seção Visão Geral apresenta o modelo em dividido em camadas sob uma visão global sem entrar em detalhes a respeito de sua composição interna.

O conceito de *white box* se refere a apresentação da estrutura interna do modelo que será descrito em camadas especializadas na seção Camadas do Biomon. Ainda considerando a estrutura interna do modelo a seção Arquitetura do Biomon apresenta um maior nível de detalhamento e descreve a interação entre as camadas.

4.1 Visão geral

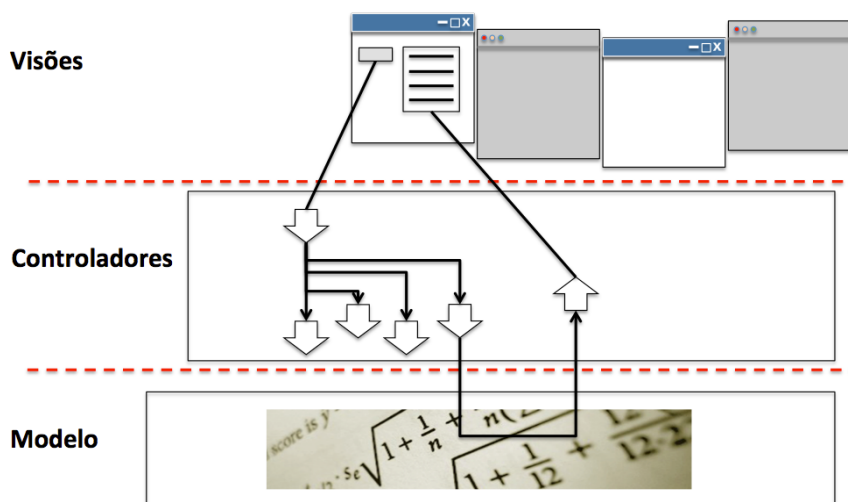
Biomon é uma proposta de software que procura contemplar sua extensibilidade em nível de desenvolvimento, como forma de proporcionar a cobertura de diferentes contextos com um mínimo de esforço. Outro aspecto considerado pelo software é sua usabilidade, onde sua operação e configuração são compartilhadas entre o especialista em estatística, aqui chamado de estatístico, e o usuário final ou pesquisador. Ambos papéis podem, sem prejuízo para a usabilidade, serem desempenhados por um usuário mais experiente ou capacitado.

Em relação a extensibilidade esta é alcançada com um arquitetura que contempla o padrão de modelagem Modelo-Visões-Controladores figura 2, padrão este que teve suas primeiras referências em (REENSKAUG, 1979a) e (REENSKAUG, 1979b). Levando em conta o fato de que a camada do Modelo se encarregaria de disponibilizar os métodos para interpretação do arquivo elaborado pelo estatístico, e executar as técnicas estatísticas disponíveis este não necessitaria qualquer ajuste específico para suporte a diferentes contextos.

A camada de Controladores deve responder as alterações aplicadas na camada de Visões a fim de suportar eventuais novos objetos de interface ou comportamentos implementados para contemplar uma nova instância do software. A interface gráfica (GUI - *Graphical User Interface*) pode fornecer ao usuário uma experiência aclimatada nos termos do contexto que o software se propõe. Desta forma diferentes áreas de pesquisa poderiam demandar desenvolvimentos na camada de Visões a fim de atender a um contexto específico, resultando em um novo software com relativamente pouco esforço.

A próxima seção descreve em detalhes o escopo de cada uma das camadas no modelo proposto.

Figura 2: Visão geral do Biomon



Fonte: Elaborada pelo autor

4.2 Camadas do Biomon

4.2.1 Camada de Visões

A camada de Visões ou interface tem papel fundamental em uma característica que Biomon procura promover que é fornecer ao usuário final, pesquisador, uma experiência ambientada nos termos da área de pesquisa que a instância do software se propõe. Outro aspecto importante presente nesta camada de desenvolvimento é o fato de que os ajustes aplicados em seus componentes tem pouco ou nenhum impacto no restante do software. Desta forma versões do mesmo software para linhas de pesquisa similares podem ser distribuídas com diferentes interfaces preservando seu funcionamento interno, o que reduz a curva de desenvolvimento destes softwares sensivelmente.

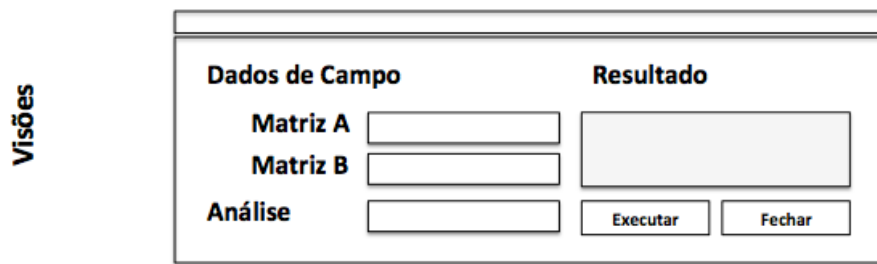
A figura 3 apresenta um exemplo do que se espera desta camada. No contexto do modelo proposto o usuário deve ter condições de informar seus dados de campo, a análise estatística que pretende executar e ao final ter acesso ao resultado desta análise.

4.2.2 Camada dos Controladores

A camada de Controladores (Figura 4) reúne as rotinas que vão responder à interação do usuário com o software pela interface onde são acessados elementos como botões e menus. Os controladores tem a função de dar suporte as requisições destes objetos executando chamadas as rotinas descritas e implementadas na camada do modelo.

Uma ação pode ser representada por uma chamada ou um conjunto de chamadas encadea-

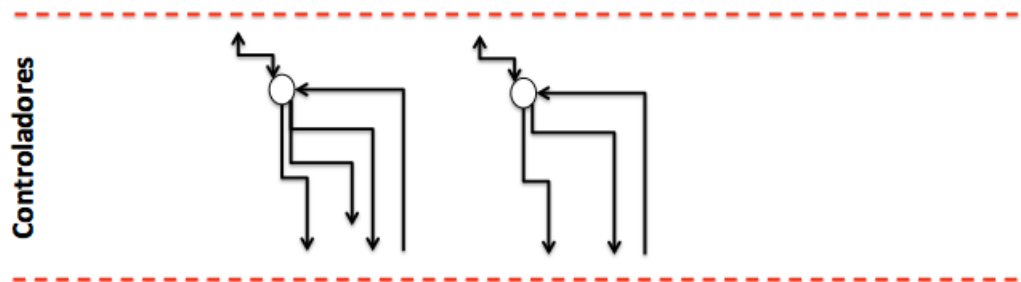
Figura 3: Camada de Visões



Fonte: Elaborada pelo autor

das. Desta forma alterações no comportamento destas ações podem ser aplicadas na camada de Controladores apenas, sem necessariamente causar impacto no restante do código. Da mesma forma que atualizações nas camadas de Visões e modelo podem acarretar efeitos colaterais nos controladores apenas não necessariamente no software como um todo.

Figura 4: Camada de Controladores



Fonte: Elaborada pelo autor

4.2.3 Camada do Modelo

A camada do Modelo (figura 5), como sugere o nome, reúne toda a lógica responsável pela mecânica de funcionamento do Biomon como proposta de software para automação da análise de dados. Nesta camada é implementado o suporte estatístico, seja de forma direta com a criação de classes de cálculo ou por chamadas à bibliotecas externas. Também é atribuição do modelo prover rotinas para leitura e interpretação do arquivo de análise criado pelo estatístico e informado pelo pesquisador juntamente com as referências para seus dados de campo.

A principal característica do Biomon é que o diferencia dos demais trabalhos relacionados é o fato de que as análises estatísticas, caracterizadas por passos que descrevem técnicas estatísticas, são informadas em tempo de execução na forma de um arquivo com formatação específica. Desta forma o usuário tem a possibilidade de organizar conjuntos de arquivos de análise que

mais se adequam a sua necessidade em um mesmo ambiente de maneira simples e intuitiva. A formatação deste arquivo de análise vai depender muito do ambiente onde o software estiver sendo implementado mas de maneira geral o que se espera é um baixo nível de complexidade para o estatístico no momento de formatar tal arquivo, abaixo um exemplo:

- **Arquivo no formato CSV:** se trata de um formato comumente aplicado em troca eletrônica de informações e suportado pela maioria (senão a totalidade) dos softwares de planilhas eletrônicas, maiores detalhes podem ser obtidos a partir da sua RFC (*Request For Comment*) (RFC 4180, 2015);
- **Informações:** para a descrição dos passos a serem executados em uma análise de dados, o arquivo CSV deve conter no mínimo uma lista de funções ou métodos a serem executados bem como as referências aos seus parâmetros, para serem interpretados e executados pelo Biomon;
- **Referência dinâmica:** o arquivo de análise deve permitir ao estatístico fazer referência aos dados informados pelos usuários de forma dinâmica permitindo, desta forma, que um mesmo conjunto de passos descritos em um arquivo de análise possam ser usados com diferentes conjuntos de dados.

Figura 5: Camada do Modelo



Fonte: Elaborada pelo autor

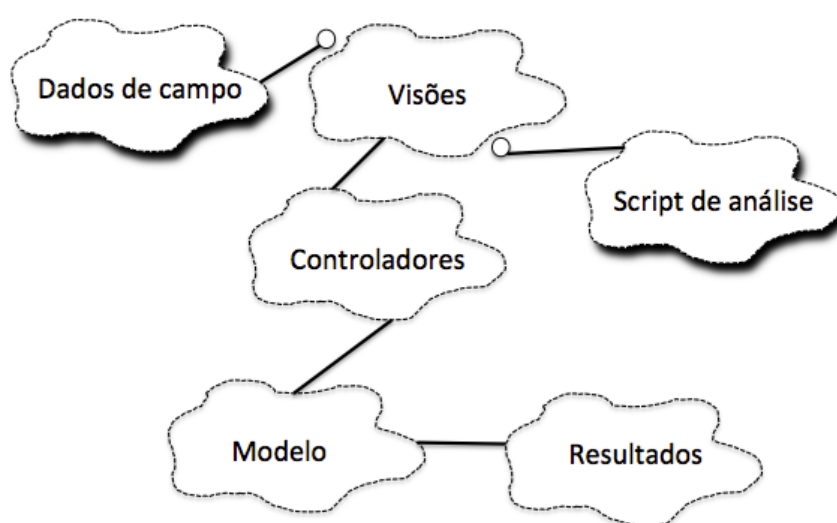
4.3 Arquitetura do Biomon

Para a arquitetura do modelo foi considerada a metodologia de descrição de modelo de software *4+1 View Model of Software Architecture* (KRUCHTEN, 1995) com algumas adaptações a fim de adequar o modelo do software proposto. A seguir serão apresentadas as visões propostas pela metodologia com seus componentes e a forma como estes se integram descrevendo o software modelado:

4.3.1 Visão Lógica

A figura 6 apresenta a visão lógica do modelo, onde os requerimentos funcionais são apresentados. As representações “Dados de campo” e “Script de análise” são objetos que o usuário deve interagir diretamente por meio da camada de interface do software. A camada de Visões também disponibiliza meios para que o usuário execute ações que serão roteadas pela camada de Controladores para os métodos definidos na camada do Modelo que lêem e interpretam o arquivo de análise e retornam o resultado obtido.

Figura 6: Visão Lógica



Fonte: Elaborada pelo autor

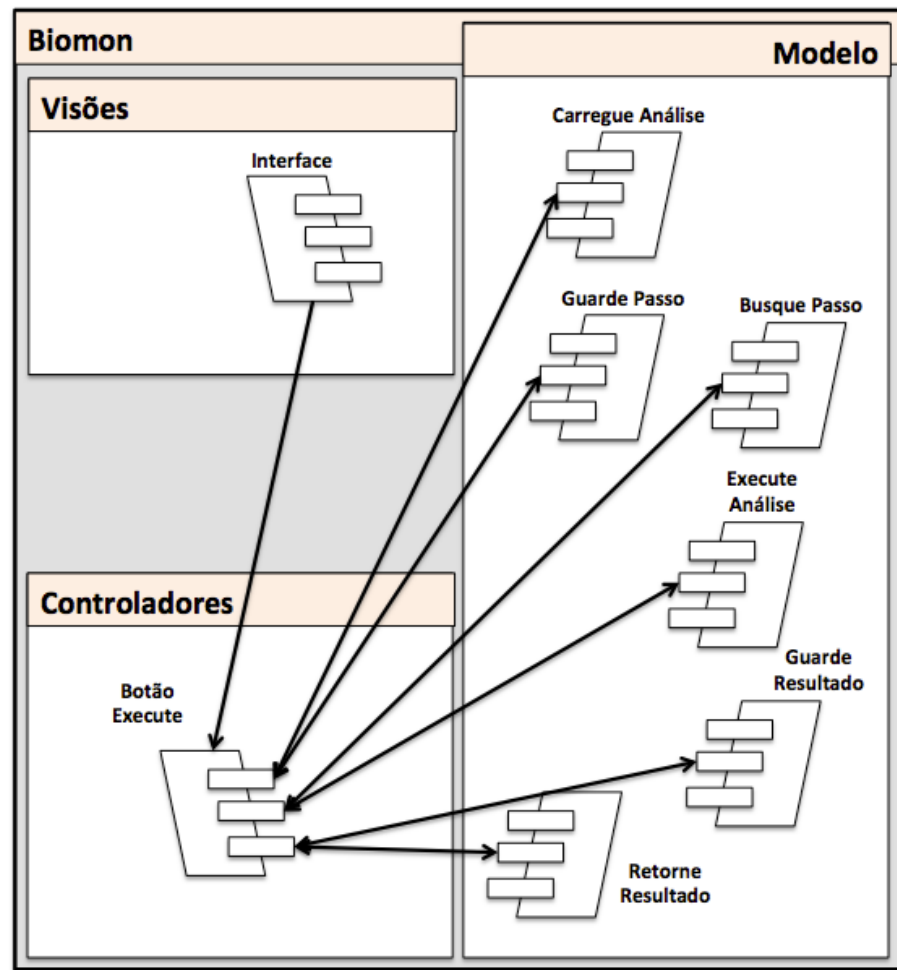
4.3.2 Visão de Processos

A figura 7 proporciona uma visão mais detalhada da interação dos componentes internos do modelo. A visão de processos considera um aspecto mais detalhado da arquitetura levando em conta requerimentos não funcionais, detalhando a forma como os métodos são distribuídos objetivando a especialização e manutenibilidade.

4.3.3 Visão do Desenvolvimento

A figura 8 apresenta a modularização implementada no desenvolvimento do modelo proposto e o que se espera de cada módulo ou camada. A representação *bottom-up* descreve a interação do usuário e a forma como esta interação é tratada pelo software em cada uma das camadas. É possível observar que a camada **1.Visão** possui uma participação mais interativa, permitindo ao usuário executar ações que serão tratadas pela camada **2.Controladores** que

Figura 7: Visão de Processos



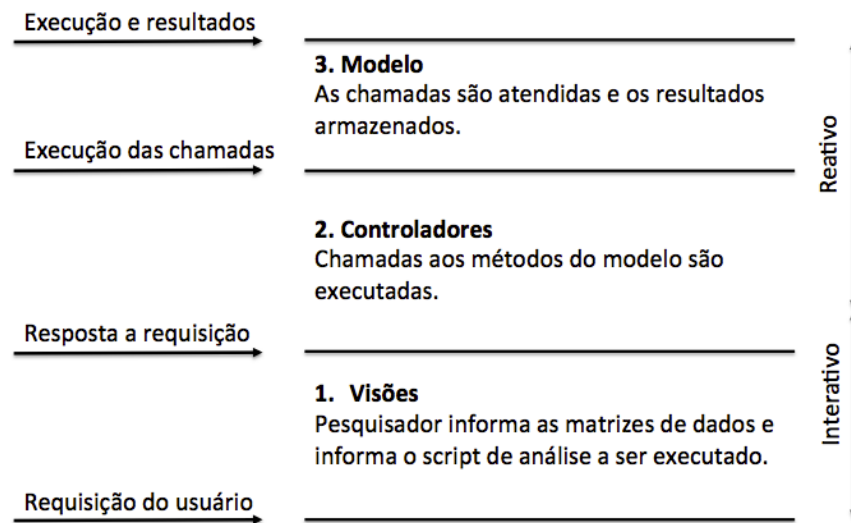
Fonte: Elaborada pelo autor

por sua vez participa de maneira interativa e reativa transformando as requisições dos usuários em uma sequência de chamadas aos métodos definidos na camada **3.Modelo**, que por sua vez responde a estas chamadas.

4.3.4 Visão Física

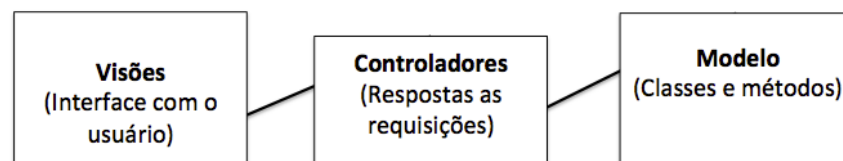
A visão física procura apresentar a distribuição física da solução desenhada, cobrindo sistemas distribuídos e arquitetura de *hardware*. O modelo proposto é organizado em camadas ou pacotes, por isso a figura 9 se limita a apresentar esta distribuição. A camada de Visões agrega as funções responsáveis por fornecer ao usuário final a interface com seus artefatos gráficos como botões, *menus* e campos de edição. Já os controladores são rotinas internas que fazem o roteamento das requisições executadas na interface para os métodos implementados no Modelo, que por sua vez disponibiliza as funcionalidades de leitura e interpretação do arquivo e análise bem como a execução dos passos descritos neste.

Figura 8: Visão do Desenvolvimento



Fonte: Elaborada pelo autor

Figura 9: Visão Física



Fonte: Elaborada pelo autor

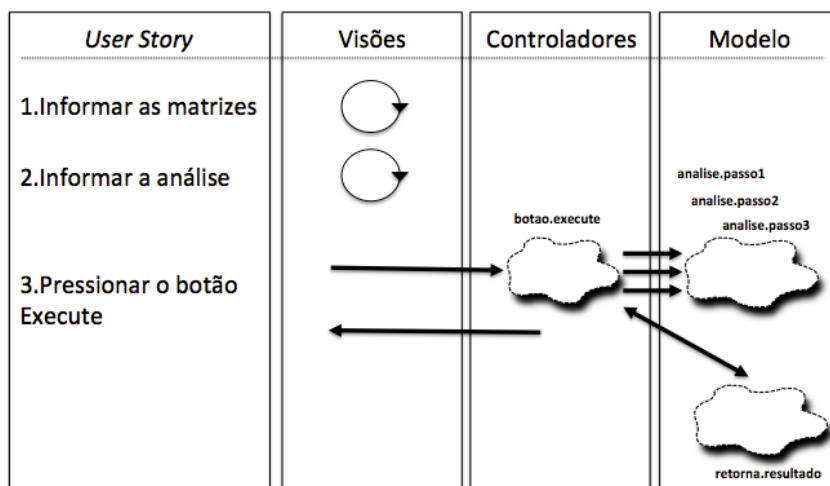
4.3.5 Visão +1

A visão +1 procura apresentar a integração das visões apresentadas até aqui em casos práticos também conhecidos como casos de uso. Para um melhor entendimento na forma como o pesquisador deve interagir com o software e executar suas análises, a seguir será descrita a sequência de passos (figura 10) a serem executados pelo pesquisador para executar uma operação básica.

- O pesquisador usa menus *drop-down* do grupo **Matrizes de dados** para informar cada uma das matrizes que pretende utilizar na análise;
- O pesquisador usa o menu *drop-down* do grupo **Análise** para informar o *script* com a análise que pretende executar sobre a(s) matriz(es) informada(s);
- O pesquisador seleciona o botão **Execute**;

- O pesquisador tem o resultado da execução do último passo de sua análise apresentado no quadro logo abaixo do botão **Execute**.

Figura 10: Sequencia de Passos



Fonte: Elaborada pelo autor

4.4 Considerações sobre o capítulo

O presente capítulo procurou apresentar os conceitos e objetivos do modelo proposto, o Biomon, de forma a guiar o desenvolvimento de um software para automação de análise de dados e apoio ao biomonitoramento. Foram destacadas as características que procuram diferenciar o modelo proposto das opções já disponíveis. Características estas que não se limitam a sua operação pelo usuário mas propõe um modelo de desenvolvimento que encoraja o desenvolvimento de novas versões ou instâncias do software com o objetivo de atender a diferentes áreas da pesquisa fornecendo ao usuário uma interface adaptada à sua linha de pesquisa.

Entre as características propostas como diferenciais em relação aos softwares estatísticos disponíveis pode-se destacar o compartilhamento de responsabilidades entre o especialista em estatística, o estatístico, e o usuário final do software ou pesquisador. O que se busca com esta divisão é liberar o pesquisador do trabalho de formatar conjuntos de softwares para análise estatística a fim de executar um estudo mas se focar no que é mais importante, a análise dos resultados. Por outro lado o estatístico tem a oportunidade de descrever uma análise em um arquivo pré-formatado e distribuí-lo como qualquer outro arquivo eletrônico, permitindo que sua proposta de análise esteja acessível facilmente pela comunidade científica.

Com sua arquitetura organizada em camadas, o que se propõe é um modelo de software cuja manutenção de seus componentes tenha efeito restrito a sua camada, e eventuais efeitos colaterais, sejam fruto de mudanças de requisito planejados. Da mesma forma procura-se encorajar o

desenvolvimento de novas instâncias do software para cobrir diferentes áreas de pesquisa. Sua estrutura modularizada permite a criação de um novo software a partir de uma outra instância, aplicando ajustes em sua interface sem necessariamente gerar efeitos colaterais nos mecanismos de interpretação do arquivo de análise ou execução das técnicas estatísticas.

No próximo capítulo serão apresentadas todas as etapas do desenvolvimento de um protótipo em linguagem de programação R (THE R PROJECT, 2014) aplicando todos os conceitos descritos até então. Serão detalhados todos os aspectos de desenvolvimento, do desenho a implementação, bem como a execução de um arquivo de análise de exemplo. Como forma de validar tanto o protótipo quando o modelo proposto será executada uma série de testes de aceite com pesquisadores em ecologia onde dados reais de pesquisa serão analisados pelo protótipo. Ao final, a execução será avaliada pelos usuários usando um formulário próprio bem como coletada a opinião deste de forma descritiva.

5 PROTÓTIPO

O presente capítulo tem como propósito apresentar a implementação de um protótipo, chamado Biomon app, a fim de demonstrar a aplicação prática dos conceitos apresentados até então. Este mesmo protótipo será usado em uma atividade de avaliação de aceitação de tecnologia com a aplicação de um caso de teste e avaliação com pesquisadores da área de biomonitoramento.

5.1 Ambiente de programação

A linguagem de programação escolhida para o desenvolvimento foi o R (THE R PROJECT, 2014). A escolha ocorreu principalmente pela sua arquitetura fortemente dedicada à computação estatística disponibilizando uma extensa gama de bibliotecas para execução de rotinas de análise estatística e por possuir uma comunidade ativa de desenvolvedores que implementam e suportam novos pacotes que estendem as funcionalidades do ambiente.

5.1.1 The R Project

O ambiente de computação estatístico R (THE R PROJECT, 2014) é uma derivação de código aberto de outra linguagem chamada S (THE S SYSTEM, 2014). Apesar das diferenças boa parte dos códigos desenvolvidos em S são interpretados pelo R sem maiores problemas. A principal característica do R em relação ao seu predecessor é o cuidado com os gráficos gerados que resultam em um material de melhor qualidade para publicações, além do fato de sua licença estar sob os termos da *GNU General Public License* (GENERAL PUBLIC LICENSE, 2014) da *Free Software Foundation* (FREE SOFTWARE FOUNDATION, 2014).

O R (THE R PROJECT, 2014) tem seus interesses mantidos pela *The R Foundation*, uma organização fundada pelo time responsável pelo *core* do ambiente, que entre outros objetivos procura fomentar o contínuo desenvolvimento do ambiente e sua divulgação. Como ponto inicial para maiores informações a página *The R Project* (THE R PROJECT FOR STATISTICAL COMPUTING, 2014) fornece uma boa quantidade de material a respeito do ambiente, sua origem e de que forma é possível contribuir com o desenvolvimento de novos pacotes.

5.2 Arquitetura do Protótipo

Conforme já foi estabelecido pelo modelo o desenho do protótipo considera o padrão de desenvolvimento em camadas Modelo-Visões-Controladores. O protótipo foi desenvolvido em três pacotes, um para cada camada, que juntos representam uma instância do Biomon app para automação de análise de dados de campo. As responsabilidades e o escopo de cada camada são descritos a seguir:

- **Camada do Modelo - Pacote *analyz*:** A camada do modelo contempla a mecânica prin-

cial do software. É nesta camada que são definidos os métodos de leitura e interpretação do *script* de análise estatística desenvolvido pelo estatístico. O modelo contempla também o suporte às variáveis globais (tabela 4) que fazem com que os passos descritos na análise tenham acesso as matrizes de dados informadas pelo pesquisador.

- **Camada de Controladores - Pacote biomonCore:** Os controladores são rotinas que farão o roteamento das requisições do usuário que por meio da interface acessa botões e menus e dispara ações para serem tratadas pelos métodos da camada do modelo.
- **Camada de Visões - Pacote biomon:** A camada de Visões contempla a interface do usuário com o software. Nesta camada são implementados os componentes visuais que permitirão ao usuário informar os dados de campo, escolher a análise que melhor se adequar a sua abordagem de análise e executá-la. Por se tratar da interação com o usuário final foram consideradas tanto uma interface simplificada, com poucos objetos, como termos adaptados ao contexto, como por exemplo o nome de cada uma das matrizes contempladas.

5.3 O Estatístico

O estatístico tem participação central no ciclo de vida do Biomon app, ele ou ela é responsável por conectar a engenharia implementada com as necessidades dos pesquisadores. Para atingir tal objetivo, algumas regras são estabelecidas com o objetivo e conectar o conhecimento estatístico e a análise de dados. Conhecimentos básicos em programas de planilha eletrônica são necessários para a transposição das funções do ambiente de desenvolvimento R em linhas e colunas de um arquivo de análise CSV. Experiência em R se apresenta útil para o momento de decidir a sequência de passos a serem seguidos em uma análise.

5.3.1 Arquivo de Análise

O leiaute do arquivo de análise representa a especificação definida como forma de comunicar ao Biomon app os passos a serem seguidos para determinada abordagem estatística. O arquivo pode ser criado manualmente em um editor de texto, mas esta tarefa pode se tornar trabalhosa para a definição de análises mais elaboradas ou para funções onde são necessários um grande número de parâmetros. Por isso é recomendável a utilização de um software de planilha eletrônica para a tarefa de elaboração e geração do arquivo de análise. A maioria destes softwares (senão a sua totalidade) possuem a funcionalidade de leitura e exportação de arquivos CSV. Os itens a seguir descrevem o que é esperado de um *script* de análise:

- **Composição:** Cada linha deve conter no mínimo quatro colunas como apresentado na tabela 2. A primeira coluna deverá conter uma identificação, a segunda o nome de uma

função do R, a terceira o tipo do parâmetro informado e na quarta o valor deste parâmetro. Não há a necessidade de informar os parâmetros do tipo caracter/*string* entre aspas, todos parâmetros são convertidos para seus tipos informados em tempo de execução.

Tabela 2: Leitura do *Script CSV*

Coluna	Conteúdo
Identificação	A identificação é um texto qualquer que identifica o passo descrito
Função R	Função R que será executada pelo interpretador
Tipo do Parâmetro	Tipo de dado que o valor informado para o parâmetro deve ser convertido
Valor do Parâmetro	Valor informado para o parâmetro

Fonte: Elaborada pelo autor

- **Tipos de parâmetros suportados:** Como descrito anteriormente, cada valor informado para os parâmetros necessariamente precisa ter seu tipo de dado informado. Isso porque o interpretador do Biomon app vai converter o valor de cada parâmetro para seu tipo de dado correspondente antes de executar a função R. A tabela 3 apresenta os tipos de dados suportados e suas funções correspondentes no ambiente R.

Tabela 3: Tipo de Dado para Parâmetros

Tipo de dado	Função equivalente no R
null	as.null(variable)
numeric	as.numeric(variable)
double	as.double(variable)
character	as.character(variable)
logical	as.logical(variable)
vector	c(variable)
list	list(variable)

Fonte: Elaborada pelo autor

- **Demais parâmetros:** Algumas funções do R requerem mais do que apenas um parâmetro. Para estes casos as colunas três e quatro devem ser repetidas com seus devidos valores, ou em outras palavras, os próximos parâmetros devem ser representados por um novo par de colunas com o tipo de dado e o valor do próximo parâmetro. Este procedimento deve ser repetido até que todos os parâmetros sejam informados.
- **Valores já calculados:** Valores já calculados em passos anteriores podem ser informados como parâmetros nos passos subsequentes em novos cálculos com a utilização do *token @*. Cada execução, também chamada de passo, tem seu resultado armazenado em uma tabela de resultados em uma posição correspondente a sua linha de execução no arquivo CSV. Desta forma quando um valor já calculado precisa ser informado como parâmetro em um passo seguinte o pesquisador precisa apenas informar o seguinte:

- Na coluna do tipo de dado deverá ser informado o *token* @;
- Na coluna do valor do parâmetro deverá ser informado o número da linha que gerou o resultado esperado.

Desta forma o interpretador irá informar como parâmetro para a função R um valor calculado que já possui seu tipo de dado definido, por esta razão não há a necessidade de descrever esta informação novamente.

- **Tipos de dados complexos:** Algumas funções do R requerem não apenas um dado simples, mas um tipo complexo, como uma matriz, vetor ou *data frame*. Para tanto o estatístico deve combinar as técnicas descritas acima, por exemplo, em um passo criar um vetor de valores e no passo seguinte informar como parâmetro o resultado do passo anterior como parâmetro. Desta forma o parâmetro informado será não apenas um valor mas um vetor de valores.
- **Variáveis globais:** Quando o estatístico estiver elaborando o *script* de análise este não terá acesso aos dados do pesquisador, ou estes podem variar dependendo do conjunto de dados a ser analisado pelo mesmo arquivo de análise. Para resolver esta questão o Biomon app disponibiliza ao estatístico um conjunto de variáveis globais que são preenchidas com os dados informados pelo pesquisador em tempo de execução do software. De forma similar como foi descrito para os valores já calculados, as variáveis globais podem ser informadas como parâmetros em um passo, e isso é feito com a utilização do *token* \$ como descrito a seguir:
 - Na coluna do tipo de dado deverá ser informado o *token* \$;
 - Na coluna do valor do parâmetro deverá ser informado o nome da variável global (veja tabela 4).

Tabela 4: Variáveis Globais

Variável Global	Descrição
gbl.analysis	Contém o caminho onde os <i>scripts</i> de análise serão armazenados
gbl.matrices	Contém o caminho onde as matrizes de dados serão armazenadas
gbl.mAnalysis	Contém o nome do arquivo da matriz com os passos a serem executados
gbl.mAttributes	Contém o nome do arquivo da matriz com os dados de atributos
gbl.mSpecies	Contém o nome do arquivo da matriz com os dados das espécies
gbl.mSpace	Contém o nome do arquivo da matriz com os dados de espaço
gbl.mFactor	Contém o nome do arquivo da matriz Fator ou T
gbl.mEnvironment	Contém o nome do arquivo da matriz com os dados de ambiente

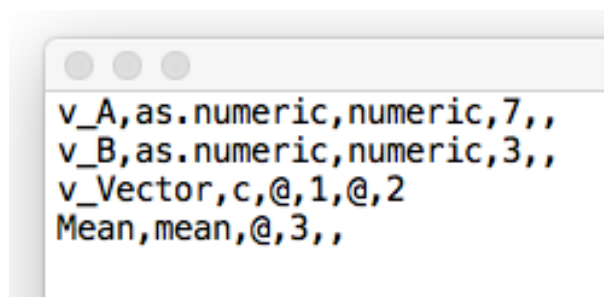
Fonte: Elaborada pelo autor

- **Arquivo CSV:** Maiores informações a respeito deste formato podem ser obtidas em (RFC 4180, 2015).

O resultado final de um arquivo de análise deve parecer com as figuras abaixo. O arquivo representa o cálculo da média de dois valores pré definidos. O objetivo deste exemplo é apresentar um arquivo de análise simples como exemplo.

- A figura 11 apresenta a aparência do *script* de análise em um arquivo de texto puro.

Figura 11: Editor de Textos



```
v_A,as.numeric,numeric,7,,
v_B,as.numeric,numeric,3,,
v_Vector,c,@,1,@,2
Mean,mean,@,3,,
```

Fonte: Elaborada pelo autor

- A figura 12 apresenta a aparência do *script* de análise em um software de planilha eletrônica.

Figura 12: Planilha Eletrônica

	A	B	C	D	E	F
1	v_A	as.numeric	numeric	7		
2	v_B	as.numeric	numeric	3		
3	v_Vector	c	@	1	@	2
4	Mean	mean	@	3		
5						
6						

Fonte: Elaborada pelo autor

É recomendado ao estatístico algum conhecimento em biomonitoramento e R levando-se em conta que o Biomon app está aclimatado nestes assuntos. Não há a necessidade de acesso a dados reais de pesquisa para a elaboração do arquivo de análise, apenas alguns exemplos para teste são suficientes. Esta independência é chave para o sucesso das análises elaboradas e compatibilidade com diferentes conjuntos de dados coletados em campo.

5.3.2 Caso de Uso

Para apresentar um caso de uso real da elaboração de uma análise a seguir são detalhados os passos para criação de um arquivo para o cálculo da matriz T.

- **Arquivo de Análise:** Para este caso de uso será criado um arquivo de análise com o nome `Sample_Matrix_T.csv` que vai calcular a matriz T e gerar um arquivo CSV no repositório de matrizes do Biomon, deixando este disponível para outras análises. O cálculo da matriz T será descrito em uma sequência de passos no arquivo de análise e posteriormente informado pelo usuário no Biomon app juntamente com as matrizes de atributos(B) e espécies(W). O que se espera destes passos é a transposição da matriz de atributos(B) em uma nova matriz que deve ser multiplicada pela matriz de espécies(W), ao final a nova matriz (T) será apresentada como resultado e um arquivo `New_Factor_T.csv` gerado no repositório de matrizes do ambiente.

A elaboração do arquivo de análise como apresentado na tabela 5 considerou a **versão 3.1.2 do R** e será descrita a seguir.

Tabela 5: `Sample_Matrix_T.csv`

read_attr	read.table	\$	gbl.mAttributes	logical	TRUE	character	“,”	null	null	character	.	character	allow.loss	numeric	1
read_spec	read.table	\$	gbl.mSpecies	logical	TRUE	character	“,”	null	null	character	.	character	allow.loss	numeric	1
matrix1	as.matrix	@	1												
matrix2	as.matrix	@	2												
transp_mtx	t	@	3												
fact_mtx	tcrossprod	@	5	@	4										
path	paste0	\$	gbl.matrices	character	New_Factor_T.csv										
write	write.csv	@	6	@	7	logical	FALSE	logical	FALSE						
read_spec	read.table	@	7	logical	TRUE	character	“,”	null	null	character	.	character	allow.loss	numeric	1
matrix	as.matrix	@	9												

Fonte: Elaborada pelo autor

1. **Leitura da matriz de atributos:** o primeiro passo a ser executado pelo interpretador do Biomon é a leitura dos dados da matriz de atributos informada pelo pesquisador. Para tanto o arquivo de análise deve, em sua primeira linha, descrever este procedimento. Como já foi descrito anteriormente a primeira instrução de um passo é a declaração de uma identificação, em seguida informamos o comando R “`read.table`”. O primeiro parâmetro “`file`” representa o caminho e o arquivo a ser importado, neste caso a matriz de atributos, para isso deve-se informar o *token* \$ como tipo de parâmetro e `gbl.mAttributes` como valor para o parâmetro. Desta forma em tempo de execução o interpretador envia para a função R o caminho e o arquivo informado pelo pesquisador na interface. Os demais parâmetros são definidos na sequência como:

- Tipo *logical* valor *TRUE* (parâmetro “header”);
- Tipo *character* valor “,” (parâmetro “sep”);
- Tipo *null* valor *null* (parâmetro “quote”);
- Tipo *character* valor “.” (parâmetro “dec”);
- Tipo *character* valor *allow.loss* (parâmetro “numerals”);
- Tipo *numeric* valor 1 (parâmetro “row.names”).

O resultado da execução deste passo é armazenada na primeira posição da tabela interna de resultados do Biomon app, e pode ser acessada com a utilização to *token* @ no tipo de uma variável e 1 no seu valor.

2. **Leitura da matriz de espécies:** o próximo passo é a leitura da matriz de espécies informada pelo pesquisador. O procedimento é basicamente o mesmo seguido no passo anterior sendo informada apenas uma nova identificação, e trocada a variável global que aponta para a matriz, agora de espécies, para `gbl.mSpecies`, o restante se mantém o inalterado. A execução deste passo tem seu resultado armazenado na segunda posição da tabela interna de resultados, este procedimento se repete para os demais passos, da mesma forma que o acesso a estes resultados.
3. **Dados dos atributos como tipo de dado matriz:** agora que as duas matrizes necessárias para o cálculo da matriz T já foram lidas é necessário assegurar que ambas sejam preparadas para os próximos passos. Para isso, este terceiro passo vai converter o resultado obtido com o passo 1 (leitura da matriz de atributos) em dado do tipo matriz. A descrição deste passo segue o mesmo procedimento adotado anteriormente, informa-se uma identificação, um comando R que neste caso será “`as.matrix`” e um único parâmetro necessário, neste caso o resultado da linha 1, ou seja, tipo de parâmetro @ valor 1.

Desta forma o interpretador executa o comando “`as.matrix`” no R enviando como parâmetro o valor lido no passo 1 e armazena este novo resultado na terceira posição da tabela interna de resultados.
4. **Dados das espécies como tipo de dado matriz:** Assim como no passo anterior, a leitura da matriz das espécies precisa ter seu tipo de dado convertido para matriz. O procedimento é basicamente o mesmo do passo anterior apenas trocando-se a linha da tabela de resultados a ser lida para 2 (resultado do segundo passo).
5. **Transpor matriz de atributos:** Agora que a matriz de atributos foi devidamente lida e teve seu tipo de dado convertido para matriz, é possível executar a sua transposição. Para tanto será utilizada a função “`t`” e o único parâmetro a ser informado é o resultado da linha três. Desta forma teremos uma nova matriz na quinta posição da tabela interna de resultados.
6. **Multiplicar matrizes:** De posse de todas as informações necessárias chega o momento de executar o cálculo da matriz T. Para isso será usada a função “`tcrossprod`”, como primeiro parâmetro será informada a posição na tabela interna de resultados da matriz de atributos transposta (posição cinco) e como segundo parâmetro a matriz de espécies (posição dois).
7. **Caminho e arquivo da nova matriz:** Como proposta para este arquivo de análise ficou definido que além do cálculo, um novo arquivo com a matriz T recém calculada seria disponibilizada no repositório de matrizes do Biomon app para fu-

turas análises. O primeiro passo para este objetivo é definir um caminho e nome de arquivo para que os dados sejam escritos fisicamente no disco do computador do pesquisador. A função R a ser chamada nesta situação será “paste0” que concatenará seus parâmetros em uma única *string*. O primeiro parâmetro será o caminho do repositório onde o Biomon app busca suas matrizes de dados, ou seja, tipo de parâmetro \$ valor gbl.matrices. Para o nome do arquivo será definido o tipo de parâmetro *character* e o valor New_Factor_T.csv.

8. **Escrita do novo arquivo:** A escrita física da nova matriz no repositório de matrizes será executada pela função “write.csv” com os seguintes parâmetros:
 - Tipo @ valor 6 (parâmetro “X”);
 - Tipo @ valor 7 (parâmetro “file”);
 - Tipo *logical* valor *FALSE* (parâmetro “append”);
 - Tipo *logical* valor *FALSE* (parâmetro “quote”).

Desta forma a matriz T que acabou de ser calculada (posição 6 na tabela interna de resultados) será escrita em um novo arquivo no repositório de matrizes do Biomon app (posição 7 da tabela interna de resultados) e estará disponível para outras análises que o pesquisador venha a executar bastando apenas informá-la na interface.

9. **Leitura da nova matriz:** Como já foi mencionado anteriormente ao final da execução de uma análise é apreseto ao usuário o resultado da execução do último passo de uma análise. Por este motivo e para garantirmos que o arquivo da matriz T foi corretamente criado no disco do computador do pesquisador, este passo vai ler o arquivo criado para que o próximo passo o apresente como resultado final da execução da análise atual. Para isso, este passo seguirá os mesmos procedimentos descritos nos passos um e dois apenas substituindo a referência às variáveis globais (tipo \$ valor gbl.mAttribute ou tipo \$ valor gbl.mSpecies) por tipo @ valor 7, desta forma importando os dados da matriz T.
10. **Apresentação do resultado:** Com os dados da matriz T corretamente lidos a partir do novo arquivo físico, é possível executar um último passo que fará com que o Biomon app apresente ao pesquisador o resultado final da execução do arquivo de análise escolhido sobre as matrizes de dados informadas. Este passo executa a função “as.matrix” tendo como parâmetro único o resultado da execução do passo 9, ou seja, tipo @ valor 9.

5.4 O Pesquisador

O pesquisador tem o papel de usuário do software. O principal objetivo do Biomon app é auxiliar o pesquisador na análise dos dados coletados em campo. Considerando que a maior parte do trabalho complexo já foi coberto tanto pelo estatístico com seu arquivo de análise

quanto pelo software com sua lógica interna, resta ao pesquisador interagir com a interface informando os dados e a análise a ser executada.

Há diferentes técnicas que podem ser utilizadas para coletar, organizar e armazenar os dados de campo. Para o Biomon considerou-se o que está descrito em (PILLAR et al., 2009).

5.4.1 Matrizes de Dados

A tabela 6 apresenta as matrizes de dados suportadas pelo software, o pesquisador deve considerar esta distribuição quando informar seus dados de campo no Biomon app.

Tabela 6: Matrizes de Dados de Campo

Matriz	Linha	Coluna	Conteúdo
	Observação	Variáveis	
Atributos (B)	Espécies	Atributos	Valores quantitativos ou Presença/Ausência
Espécies (W)	Unidade Amostral	Espécies	Valores quantitativos ou Presença/Ausência
Espaço (S)	Unidade Amostral	Coordenadas Geográficas	Valores reais representando uma coordenada geográfica
Fator (T)	Atributos	Espécies	$T = B' * W$ (B transposta multiplicada por W)
Ambiente (E)	Unidade Amostral	Variáveis Ambientais	Valores quantitativos

Fonte: Elaborada pelo autor

Com o propósito de estabelecer um protocolo entre o que o pesquisador informar e o que o estatístico espera como matriz de dados defini-se que a primeira linha descreve cada coluna e primeira coluna descreve cada linha, como na figura 13.

Figura 13: Modelo de Matriz de Dados

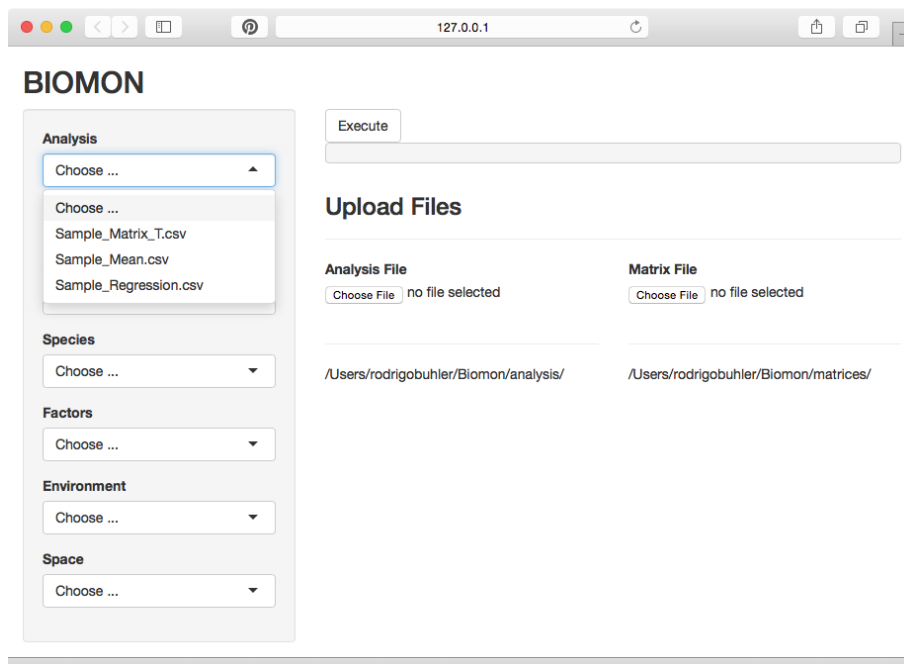
	A	B	C	D	E	F	G	H	I	J
1	Species	Trait_A	Trait_B	Trait_C	Trait_D	Trait_E	Trait_F	Trait_G	Trait_H	Trait_I
2	pos2	1	1	0	0	0	0	0	0	0
3	magy	1	1	0	0	0	0	0	0	1
4	phpa	1	0	1	0	0	0	0	0	1
5	lecy	0	0	0	1	0	1	0	1	0
6	bsp4	1	0	1	0	0	0	0	0	1
7	brle	1	1	0	0	0	0	0	0	1
8	phbr	1	0	1	0	0	0	0	0	1
9	lep3	1	1	0	0	0	1	1	0	0
10	gdin	1	1	0	0	0	0	0	0	0
11										
12										

Fonte: Elaborada pelo autor

5.4.2 Interface Gráfica

Levando em conta a distribuição de responsabilidades proposta, que considera a área de conhecimento de cada um envolvido na operação do Biomon app é esperada baixa complexidade tecnológica e poucas interações para a operação do software pelo pesquisador. É tarefa da interface do Biomon app atender a este requisito, como apresentado na figura 14.

Figura 14: Interface do Biomon app



Fonte: Elaborada pelo autor

5.4.3 Caso de Uso

Com o objetivo de apresentar uma execução real do Biomon app a seguir será apresentado um caso de uso com uma descrição detalhada de cada etapa.

- **Interface com o usuário:** Para a instalação e operação do software é recomendável algum conhecimento em R, pois a operação e mensagens de alerta ou erro ao usuário seguem o padrão da linguagem. A partir de um ambiente R com acesso a internet, o pesquisador deve instalar os três pacotes que compõe o software, para isso algumas dependências precisam ser atendidas previamente.

Para a instalação das dependências o pesquisador deve executar os seguintes comandos R:

```
install.packages("devtools")
```

```
install.packages("shiny")
```

Com as dependências devidamente instaladas ou atualizadas a instalação do Biomon app é feita a partir dos seguintes comandos:

```
devtools::install_github("rbuhler/analyz")
devtools::install_github("rbuhler/biomonCore")
devtools::install_github("rbuhler/biomon")
```

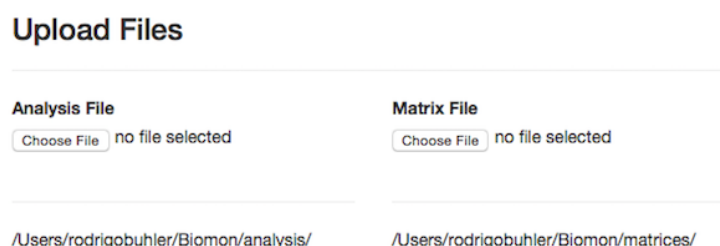
Após a instalação dos pacotes e suas dependências o software está pronto para ser executado, para isso o pesquisador deve executar os seguintes comandos:

```
library("biomon")
biomon()
```

Com a execução dos passos descritos acima, a interface do Biomon app é apresentada em um navegador de internet, como na figura 14. A interface do usuário é composta pelos seguintes objetos:

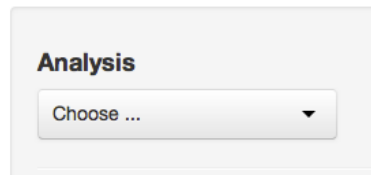
- Botões de *Upload* figura 15, usados para carregar tanto os arquivos de análise quanto as matrizes de dados para o contexto do software, desta forma estando disponíveis para a execução de uma análise;

Figura 15: Botões de *Upload*

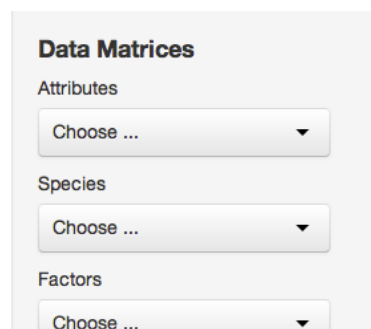


Fonte: Elaborada pelo autor

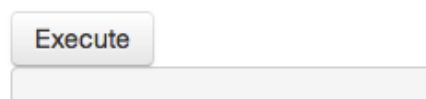
- Drop-down menus do grupo *Analysis* figura 16, usados para o pesquisador informar qual a análise será executada;
- Drop-down menus do grupo *Data Matrices* figura 17, usados para o pesquisador informar qua(is) matriz(es) poderão ser utilizada(s) na análise;
- Botão *Execute* figura 18, usado para o pesquisador executar a análise escolhida sobre as matrizes de dados informadas. Na área logo abaixo do botão será apresentado o resultado do último passo executado pelo interpretador ou alguma mensagem informando o usuário que algo não aconteceu como esperado.

Figura 16: Drop-down menu grupo *Analysis*

Fonte: Elaborada pelo autor

Figura 17: Drop-down menu grupo *Data Matrices*

Fonte: Elaborada pelo autor

Figura 18: Botão *Execute*

Fonte: Elaborada pelo autor

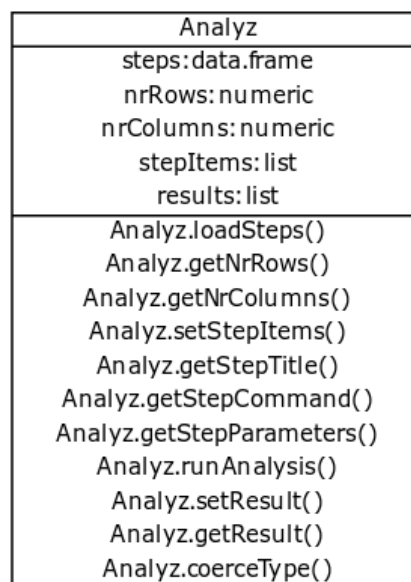
5.5 Discussão

A descrição do protótipo reforça a importância da participação do estatístico na usabilidade do Biomon app. O quanto mais funcional e bem formulado o arquivo de análise for mais valor este agrega à tarefa de análise de dados. É possível observar com isso, que a complexidade está mais concentrada na criação do arquivo de análise, tarefa do estatístico, do que na operação do software feito pelo pesquisador. Desta forma, como estabelecido pelo presente trabalho, o pesquisador não precisa investir tempo e esforço lidando com complexidades tecnológicas ou elaborando metodologias para execução de análises e pode focar no que é mais importante, a interpretação dos resultados.

5.6 Código

A implementação do Biomon app em R fez uso de dois pacotes que estendem as funcionalidades da linguagem R. Para a camada do Modelo foi utilizado o pacote **method** que permite o desenvolvimento do código orientado a objetos. A classe *Analyz* (figura 19) define e implementa todos os métodos responsáveis pela leitura e interpretação do *script* de análise bem como o acesso as matrizes informadas pelo usuário na interface. Bem mais simples, a camada de Controladores possui apenas uma rotina que responde ao clique no botão *Execute* da interface. Esta rotina instancia a classe *Analyz* e executa a chamada de seus métodos de forma a executar um procedimento de análise. Para a camada de Visões o pacote **shiny** proporciona um ambiente de desenvolvimento de interface gráfica HTML no R. Desta forma com pouco esforço de desenvolvimento é possível ter um *look and feel* intuitivo para a utilização do Biomon app.

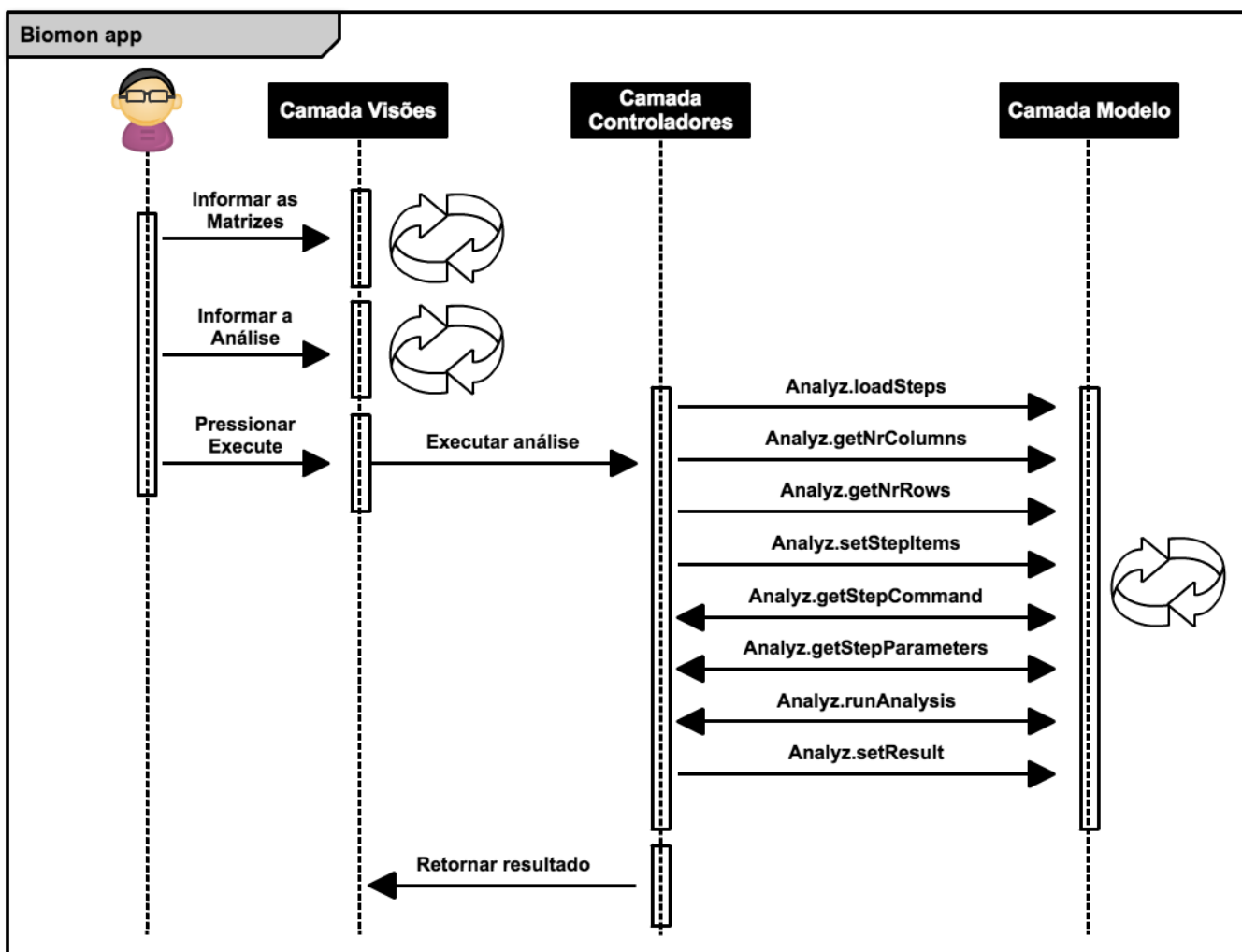
Figura 19: Diagrama da Classe *Analyz*



Fonte: Elaborada pelo autor

O caso de uso do pesquisador descrito no capítulo anterior é representado pelo diagrama de sequência abaixo. Com ele é possível observar a interação que usuário tem com a interface e o processo executado com o clique no botão *Execute* da interface.

Figura 20: Diagrama de Sequência



Fonte: Elaborada pelo autor

6 AVALIAÇÃO

Como descrito em (DAVIS, 1993) foi observado que a falta de testes de aceite de software tem sido fator determinante para o fracasso de projetos de sistemas de informação. Geralmente o desenvolvimento focado em desempenho e performance acaba por ser afetado pela rejeição dos usuários em utilizar estes sistemas. Considerando a premissa de que o teste de aceite tem papel fundamental na identificação de fatores de sucesso ou fracasso de sistemas de informação o presente modelo de software Biomon será avaliado quanto sua aceitação por pesquisadores da área do biomonitoramento.

6.1 Metodologia

A metodologia de avaliação aplicada ao modelo proposto foi a execução de um protótipo e avaliação por pesquisadores. Aos pesquisadores convidados foram enviados um documento descrevendo um caso de teste (Apêndice A) e o formulário de avaliação (Apêndice B). Os pesquisadores de posse dos passos a serem seguidos puderam avaliar todas as etapas de um caso real, desde a instalação do Biomon no ambiente R até sua execução tanto como descrito no documento fornecido como executando testes exploratórios de maneira livre. Ao final os pesquisadores eram convidados a informar seu grau de concordância em relação a dez afirmações.

Para o teste de aceitação foram convidados seis pesquisadores com diferentes perfis acadêmicos como descrito na tabela 7.

Tabela 7: Pesquisadores Entrevistados

Identificação	Perfil acadêmico	Atuação
Entrevistado 1	Doutoranda em Ecologia	Doutoranda bolsista na UFRGS
Entrevistado 2	Mestranda em Qualidade Ambiental	Mestranda em Qualidade Ambiental
Entrevistado 3	Doutoranda Ecologia	Doutoranda bolsista na UFRGS
Entrevistado 4	Mestre em Ecologia	
Entrevistado 5	Doutora em Ecologia	Professora na UFRGS
Entrevistado 7	Doutora em Ciências Ambientais	Pesquisadora da Fundação Zoobotânica do RS

Fonte: Elaborada pelo autor

6.2 Caso de Teste

Todos os pesquisadores foram convidados a participar do teste de aceitação de software e receberam o mesmo conjunto de documentos por e-mail. Durante a execução do caso de teste o pesquisador não foi acompanhado *in loco*, eventuais dúvidas técnicas foram resolvidas por meio eletrônico (e-mail). Desta forma o entrevistado teve condições de avaliar a interface e usabilidade do software procurando por respostas no próprio software para eventuais impedimentos

ou dúvidas.

O caso de teste (Apêndice A) aplicado no teste de aceitação descreve ao entrevistado todos os passos necessários para a operação do Biomon, começando com uma breve descrição a respeito das matrizes de dados, passando pela instalação do software no ambiente R e finalmente a descrição de um caso de teste.

6.3 Questionário

Para a elaboração do questionário (Apêndice B) foi adotado o Modelo de Aceitação de Tecnologia (*TAM - Technology Acceptance Model*) construído sobre os princípios descritos em (FISHBEIN; AJZEN, 1975) e publicado em (DAVIS, 1989). O modelo proposto procura prever e explicar a aceitação de sistemas de informação tomando como base dois aspectos considerados fundamentais para sua definição, a percepção do usuário em relação a facilidade de uso e utilidade da solução sob avaliação.

Procurando explicar os fatores que fazem um usuário adotar um software, foram identificados alguns aspectos dos quais dois se destacam e são adotados pelo TAM. A percepção do usuário de que o sistema em questão agrega algum benefício ao seu trabalho caracteriza o aspecto utilidade percebida. Já outra característica a ser considerada, e que tem impacto na aceitação do usuário, é a percepção do esforço necessário na operação do sistema ou software, ou facilidade de uso, atributo este que apresenta ter peso suficiente para sobrepor a percepção de utilidade percebida na decisão de aceite do software pelo usuário.

Para a avaliação do Biomon foram elaboradas afirmações (tabela 8) que os pesquisadores classificaram usando a escala Lickert (LIKERT, 1932), onde são oferecidas 5 pontos de avaliação para as afirmações que variavam de **1. Concordo totalmente até 5. Discordo totalmente**.

Tabela 8: Afirmações de ítems Likert

Grupo	Questão	Afirmação
Facilidade de Uso	1	Compreendi a função de todos os elementos da tela sem ajuda externa.
	2	A interface disponibiliza todas as funções necessárias para a operação do software.
	3	Pude seguir o caso de teste sem dificuldades.
	4	De maneira geral achei o Biomon fácil de utilizar.
	5	Achei a interface simples e agradável.
Utilidade Percebida	6	Considerando a forma como analiso meus dados de campo considero que o Biomon pode tornar meu trabalho mais fácil.
	7	Eu substituiria a maneira como faço minhas análises para utilizar apenas o Biomon.
	8	Gostei da maneira como a proposta separa os papeis do estatístico e do pesquisador.
	9	Considero importante a possibilidade de facilmente compartilhar os arquivos de análise.
	10	Com o Biomon eu posso economizar tempo executando as análises e investir mais na interpretação dos resultados.

Fonte: Elaborada pelo autor

Depois de executados os testes e respondidas as avaliações, os resultados foram compilados e são apresentados a seguir.

6.4 Considerações sobre a Avaliação

Embora não tenha ficado explícito aos pesquisadores convidados as perguntas foram formuladas em dois grupos com objetivos distintos. O primeiro grupo procurou avaliar a experiência do entrevistado quanto a utilização do software e sua impressão quanto a interface do protótipo.

Já o segundo grupo de afirmações tratou de questões mais práticas com por exemplo a substituição dos procedimentos atuais de análise de dados pela utilização do software representado pelo protótipo.

Em ambos os grupos de afirmações a avaliação foi bastante positiva, apresentando claramente pontos de sucesso e áreas a serem melhoradas como forma de se aproximarem as necessidades dos pesquisadores.

6.5 Facilidade de Uso

Por facilidade de uso entende-se (MARANGUNIC; GRANIC, 2014) a impressão que o usuário tem, quando usando o software, de que sua operação será livre de esforço, que acontecerá de forma mais natural e intuitiva possível. Procurando avaliar esta impressão dos pesquisadores convidados, o grupo de perguntas deste aspecto procurou questionar o usuário quanto a sua experiência seguindo os passos do caso de teste.

A figura 21 apresenta a compilação das respostas da avaliação feita pelos pesquisadores quando da execução dos passos do caso de teste. É possível observar que aproximadamente dois terços (62%) das respostas corroboram para o fato de que a percepção de facilidade é clara. Outro aspecto que aponta para um resultado bastante positivo é o de que no total apenas 7% das respostas não representaram a realidade do teste, mas mesmo assim parcialmente, não houveram rejeições totais.

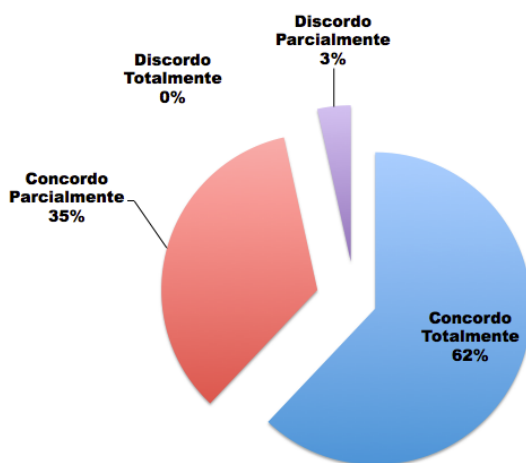
De maneira mais detalhada a figura 22 destaca as questões dois e três como pontos de atenção pois apresentam índices muito altos quando comparados com os demais. Estas questões se referem a informações de ajuda na interface e facilidade na execução dos passos do caso de teste.

As avaliações destas questões apontam para a necessidade de um maior refino na elaboração da interface do software para que na sua execução as situações onde possam haver dúvidas o usuário tenha condições de resolvê-las com a ajuda do próprio software.

6.6 Utilidade Percebida

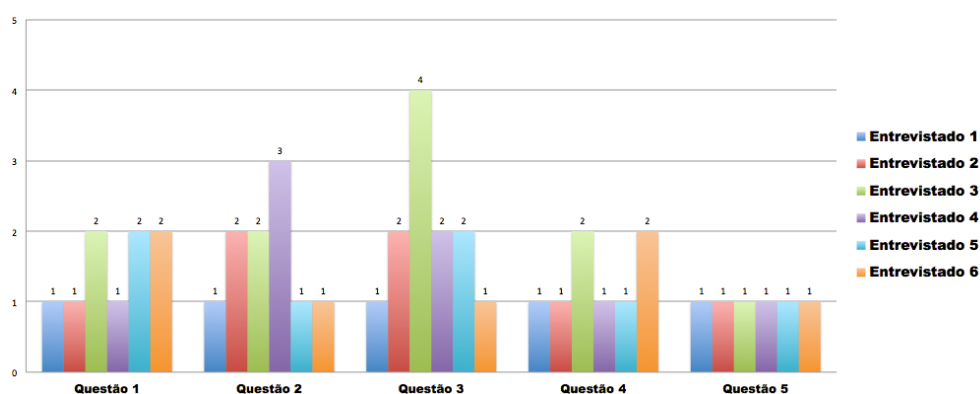
A utilidade percebida se caracteriza (MARANGUNIC; GRANIC, 2014) pelo nível de confiança que o usuário tem de que o software em uso pode efetivamente agregar valor ao seu trabalho. Desta forma as afirmações deste grupo procuraram colocar o protótipo e sua execução no contexto de análise de dados do pesquisador convidado, para que este possa avaliar de

Figura 21: Facilidade de Uso Avaliação



Fonte: Elaborada pelo autor

Figura 22: Facilidade de Uso Questões



Fonte: Elaborada pelo autor

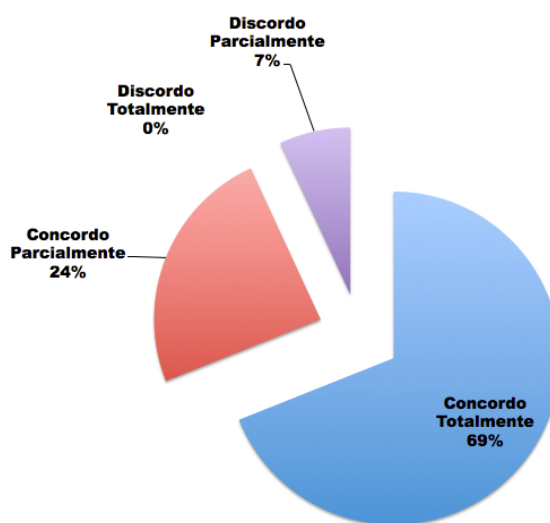
maneira fiel o cenário executado.

É possível observar na figura 23 que mais de dois terços (69%) das afirmações foram consideradas totalmente alinhadas com a realidade do entrevistado. Já os resultados de discordância em relação as afirmações somam apenas 7% do total geral.

De maneira mais detalhada observa-se na figura 24 que as questões sete e oito apresentam índices discrepantes em relação ao restante do conjunto.

A questão sete trata da possibilidade de substituição total da maneira como o pesquisador faz suas análises de dados de campo pela proposta apresentada. É de se esperar que esta questão receba índices mais elevados por se tratar da avaliação de um protótipo com poucas funcionalidades, mas ainda assim não houve qualquer indicação de rejeição absoluta a ideia.

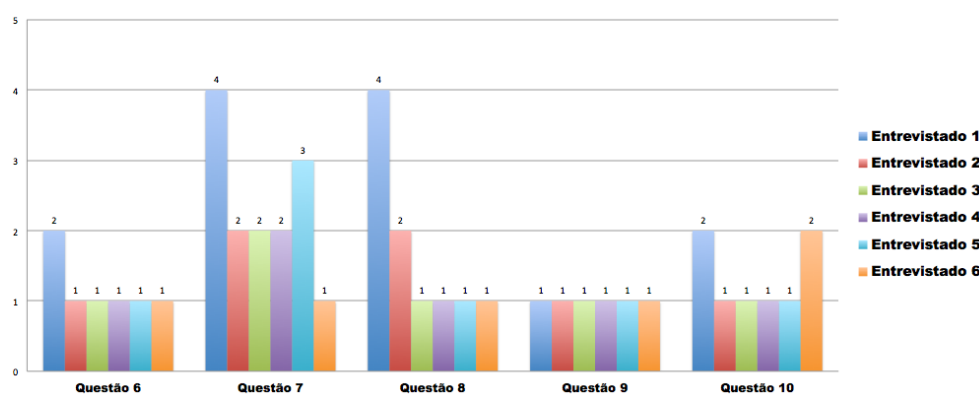
Figura 23: Avaliação da Utilidade Percebida



Fonte: Elaborada pelo autor

Já a questão oito se apresentou contraditória e isso pode ser reflexo do fato de que a relação pesquisador e estatístico não esteja clara, pois no que foi possível apurar durante o presente trabalho ambas funções são desempenhadas pelo mesmo indivíduo.

Figura 24: Utilidade Percebida Questões



Fonte: Elaborada pelo autor

Ao final da avaliação os pesquisadores eram convidados a deixarem algum comentário textual a respeito da experiência. Algumas das observações feitas pelos entrevistados são reproduzidas a seguir.

- “Considero a experiência promissora e interessante. A possibilidade de interagir mais e melhor com um estatístico profissional é realmente animadora. A cada dia novas e desafiadoras metodologias surgem e a análise dos dados, ou melhor, a simples compreensão

estatística demanda muito tempo. Isto, muitas vezes, nos faz desistir de uma análise nova, ficando com aquelas que já conhecemos. Neste sentido, uma relação mais próxima com um estatístico e, ao mesmo tempo, com uma interface amigável de análise pode de fato contribuir muito com o progresso da pesquisa e com o aumento e a qualidade das publicações geradas com dados coletados em campo. Acredito que, com um prazo maior de contato entre ambos (estatístico e o pesquisador) possa de fato conduzir a interessantes resultados e análises.”

- *“Acho que o Biomon facilitará em muito minhas análises, para isso preciso dominar melhor o programa. Até aqui achei prático e de fácil utilização.”*

7 CONSIDERAÇÕES FINAIS

O Biomon é uma proposta de modelo de software para automação de análise de dados de campo em biomonitoramento. O modelo apresentado procura contemplar ambos apoio ao pesquisador na execução de suas análises e uma arquitetura para o fomento ao desenvolvimento de novas instâncias de software com pouco esforço e impacto limitado.

Para sua avaliação um protótipo foi implementado nos moldes da arquitetura proposta o que demonstrou sua viabilidade e permitiu uma apresentação prática da aplicação do conceito de desenvolvimento em camadas MVC (*Model-Views-Controllers*). Também como parte da avaliação do modelo proposto foram executados casos de teste com seis pesquisadores convidados.

7.1 Conclusão

Para o apoio ao pesquisador o modelo propõe uma divisão de responsabilidades quando da formatação de um processo de análise de dados. Esta análise é composta por passos que descrevem a execução de técnicas estatísticas sobre os dados de campo disponibilizados pelo pesquisador. O Biomon estabelece que esta definição seja feita por um estatístico, considerando que esta é a sua área de *expertise*. Como resultado desta atribuição será gerado um arquivo *script* com instruções que serão posteriormente interpretadas e executadas pelo software.

Ao pesquisador é atribuída a responsabilidade de informar ao software suas matrizes de dados de campo juntamente com o *script* elaborado pelo estatístico. Desta forma o software é capaz de executar os passos descritos pelo estatístico sobre os dados de campo do pesquisador. Para a operação do software se espera uma baixa complexidade tecnológica, ou seja, a interface deve fornecer ao usuário final poucas opções de objetos de interface, limitando as ações a informação dos requisitos mínimos necessário para a execução do arquivo de análise, este sim cobrindo boa parte da complexidade.

O modelo de software apresentado é dividido em camadas que organizam o código de maneira contextual. Todo o tratamento da interface é organizado na camada de Visões, desta forma implementações neste escopo ficam, em grande parte, limitadas a este contexto. Já o tratamento das ações e comportamentos implementados na interface são tratados na camada de Controladores, que se encarrega de atender as requisições da interação humana com o software. As funcionalidades disponibilizadas pelo software são modeladas e implementadas na camada de Modelo. Todo seu comportamento e funcionalidades são disponibilizadas para as demais camadas a partir desta camada.

Com esta divisão é possível aplicar alterações na interface, procurando atender a uma nova área da pesquisa sem que as demais camadas sejam afetadas. Da mesma forma que um comportamento pode ser alterado agregando ao software novas possibilidades sem que necessariamente sua interface seja afetada.

A partir do protótipo foi possível aplicar testes de aceitação de software com pesquisadores

da área de biomonitoramento. A organização destes testes teve como estrutura o convite de seis testadores com perfil de pesquisadores em biomonitoramento e a elaboração de um caso de teste descrevendo os requisitos necessários, instalação e execução de uma análise no protótipo. O material com o passo a passo foi enviado para o pesquisador juntamente com um questionário para que este pudesse relatar sua experiência. Ao final dos testes foi possível observar pontos de melhoria e áreas onde as expectativas foram atendidas.

O que se pode concluir é que comprovadamente os objetivos foram alcançados, talvez não em sua totalidade, mas o que fica claro é que as necessidades observadas são reais. Quando da avaliação do critério de facilidade de uso, ficou evidente a questão de que a interface se apresenta como fator determinante para o sucesso do software e este é um assunto tratado pelo modelo proposto. Em relações as questões de utilidade percebida foi possível observar durante e logo após os testes de que a contribuição é mais evidente para usuários menos desenvoltura no tratamento com a tecnologia.

7.2 Contribuições

A proposta, implementação e teste do modelo de software Biomon colocou em discussão a forma como são feitas as análises estatísticas no âmbito do biomonitoramento e propôs melhorias na relação dos pesquisadores com a tecnologia. Estas propostas de melhorias foram baseadas em percepções que se provaram reais a partir da avaliação feita pelos próprios pesquisadores no questionário de avaliação do caso de teste.

O que foi proposto não se limita apenas a propor uma solução tecnológica, com a apresentação de uma interface com baixa complexidade tecnológica e mecanismos otimizados de software, mas também uma nova relação no procedimento de elaboração destas análises, atribuído responsabilidades ao estatístico na descrição dos passos a serem seguidos e elaboração de um *script* para posterior utilização pelo pesquisador, que por sua vez deve informar ao software esta informação juntamente com seus dados de campo.

Para o fomento ao desenvolvimentos de novas instâncias do Biomon como forma de atender as mais variadas áreas da pesquisa em biomonitoramento, o que foi proposto foi implementado em um protótipo e se apresentou flexível como esperado. Durante o desenvolvimento a manutenção e eventuais ajustes no código ficaram isolados nas camadas propostas gerando pouco ou nenhum efeito colateral nas demais camadas. Desta forma tanto o desenvolvimento de novos softwares quanto sua manutenção se beneficiam desta distribuição.

A tabela 9 apresenta novamente os trabalhos relacionados, bem como os critérios de comparação adotados. Ao final da tabela é possível observar que o modelo proposto, o Biomon, possui a combinação considerada ideal aliando a interface gráfica e cobertura de um contexto de pesquisa. Estas duas características com impacto direto na usabilidade por parte do pesquisador que com isso terá acesso a todas funcionalidades em um mesmo ambiente para o atendimento das suas necessidades. Aliado ao fato de seu código ser de domínio público, atendendo ao

objetivo de fomentar o desenvolvimento de novas instâncias.

A tabela 9 apresenta novamente os trabalhos relacionados e os critérios utilizados para comparação entre eles e o Biomon. O modelo proposto oferece soluções para as características que mais agregam valor aos seus usuários. Desta forma o investimento em uma interface gráfica se apresenta vital para aspectos de usabilidade e redução na curva de aprendizado de operação do software.

Outro critério coberto pelo Biomon é o suporte ao contexto de pesquisa, fornecendo ao pesquisador meios para executar sua análise em um ambiente único sem ter que recorrer a diferentes soluções de software, com isso evitando a troca de contextos em uma mesma análise de dados de campo. Ainda neste mesmo critério o modelo proporciona ao pesquisador a construção de conjuntos de rotinas estatísticas de forma dinâmica com a inserção de novos *scripts* de análise elaborados pelo estatístico sem a necessidade de alteração no software.

O critério de extensibilidade tem impacto indireto para a área da pesquisa e proporciona a comunidade de desenvolvimento um ambiente preparado para a criação de novas instâncias de software ou mesmo seu acoplamento em outras soluções de software já existentes. Sua arquitetura aberta de domínio público e organizada em camadas especializadas fomenta sua aplicação em projetos de desenvolvimento para atendimento as mais variadas necessidades da área da análise dados de forma controlada e com um mínimo de esforço.

Tabela 9: Contribuição do modelo Biomon

Software	Interface	Suporte	Extensibilidade
Multiv & SYNCSA	Linha de comando	Método estatístico	Código aberto
Pct.SYNCSA para R	Linha de comando	Método estatístico	Código aberto
Past	Interface gráfica	Contexto de análise	Código proprietário
PC-ORD	Interface gráfica	Método estatístico	Código proprietário
CANOCO	Interface gráfica	Método estatístico	Código proprietário
IBM SPSS	Interface gráfica	Método estatístico	Código proprietário
Biomon	Interface gráfica	Contexto de análise	Código aberto

Fonte: Elaborada pelo autor

7.3 Trabalhos Futuros

Como trabalhos futuros se propõe inicialmente atenção para os itens apontados pelos pesquisadores como críticos na proposta avaliada. Como já foi discutido anteriormente, a proposta carece de melhorias quanto ao tratamento da interface com o usuário, fornecendo a este mais informações a respeito de sua utilização. Uma melhor distribuição dos objetos de interface e mais informação visual seriam pontos a serem melhorados em futuras versões do Biomon.

Um outro aspecto observado na pesquisa de aceitação de software é a relação pesquisador e estatístico que parece não estar clara ou bem definida. Uma solução possível seria a proposta,

nos mesmos moldes do Biomon, de uma ferramenta para o estatístico. Desta forma pesquisadores que atuem em ambos papéis possam se beneficiar das mesmas características da proposta de software para execução das análises.

Durante os preparativos para as sessões de teste houve a oportunidade de receber dos pesquisadores convidados relatos a respeito de suas dificuldades quando da execução de análises estatísticas. Fatores que não foram cobertos pelo modelo proposto se apresentaram e podem ser atendidos por trabalhos futuros de forma a cobrir um espectro mais amplo do biomonitoramento. Um bom exemplo disso é a edição de gráficos, onde sua customização não é possível em determinadas soluções de software ou demasiada complexa em outras tantas. Por isso uma proposta de software nos moldes do Biomon para edição e geração de gráficos com componentes dinâmicos suportado por uma interface gráfica e intuitiva se apresenta como uma área com relativa demanda.

REFERÊNCIAS

- ALCARAZ, C.; CAIOLA, N.; IBÁÑEZ, C. Bioaccumulation of pollutants in the zebra mussel from hazardous industrial waste and evaluation of spatial distribution using GAMs. **Science of the total environment**, [S.l.], v. 409, n. 5, p. 898–904, 2011.
- ARHONDITSIS, G.; PAERL, H.; VALDES-WEAVER, L.; STOW, C.; STEINBERG, L.; RECKHOW, K. Application of Bayesian structural equation modeling for examining phytoplankton dynamics in the Neuse River Estuary (North Carolina, USA). **Estuarine, Coastal and Shelf Science**, [S.l.], v. 72, n. 1, p. 63–80, 2007.
- AYUKE, F. O.; PULLEMAN, M.; VANLAUWE, B.; GOEDE, R. de; SIX, J.; CSUZDI, C.; BRUSSAARD, L. Agricultural management affects earthworm and termite diversity across humid to semi-arid tropical zones. **Agriculture, ecosystems and environment**, [S.l.], v. 140, n. 1, p. 148–154, 2011.
- BLASIUS, J.; GREENACRE, M. **Visualization and Verbalization of Data**. [S.l.]: CRC Press, 2014.
- BONNET, E.; PEER, Y. Van de. zt: a software tool for simple and partial mantel tests. **Journal of Statistical software**, [S.l.], v. 7, n. 10, p. 1–12, 2002.
- CANOCO. Disponível em:
<<http://www.wageningenur.nl/en/show/Canoco-for-visualization-of-multivariate-data.htm>>.
Acesso em: 15 de Maio 2014 .
- CARBONI, M.; ACOSTA, A. T.; RICOTTA, C. Are differences in functional diversity among plant communities on Mediterranean coastal dunes driven by their phylogenetic history? **Journal of Vegetation Science**, [S.l.], v. 24, n. 5, p. 932–941, 2013.
- CONTI, M.; CECCHETTI, G. Biological monitoring: lichens as bioindicators of air pollution assessment—a review. **Environmental Pollution**, [S.l.], v. 114, n. 3, p. 471–492, 2001.
- CRUZ, P.; DE QUADROS, F. L. F.; THEAU, J. P.; FRIZZO, A.; JOUANY, C.; DURU, M.; CARVALHO, P. C. F. Leaf traits as functional descriptors of the intensity of continuous grazing in native grasslands in the south of Brazil. **Rangeland Ecology & Management**, [S.l.], v. 63, n. 3, p. 350–358, 2010.
- DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS quarterly**, [S.l.], p. 319–340, 1989.
- DAVIS, F. D. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. **International journal of man-machine studies**, [S.l.], v. 38, n. 3, p. 475–487, 1993.
- DEBASTIANI, V. J.; PILLAR, V. D. SYNCOSA—R tool for analysis of metacommunities based on functional traits and phylogeny of the community components. **Bioinformatics**, [S.l.], v. 28, n. 15, p. 2067–2068, 2012.
- ELLENBERG, H.; WOODS-SCHANK, G.; ENTWICKLUNGSTECHNOLOGIEN, D. Z. für. **Biological monitoring**: signals from the environment. [S.l.]: Vieweg Braunschweig, 1991.

FISHBEIN, M.; AJZEN, I. **Belief, attitude, intention and behavior**: an introduction to theory and research. [S.l.: s.n.], 1975.

FREE Software Foundation. Disponível em: <<http://www.gnu.org/>>. Acesso em: 15 de Maio 2014 .

GENERAL Public Lincense. Disponível em: <<http://www.r-project.org/COPYING>>. Acesso em: 15 de Maio 2014 .

GENTLEMAN, R. C.; CAREY, V. J.; BATES, D. M.; BOLSTAD, B.; DETTLING, M.; DUDOIT, S.; ELLIS, B.; GAUTIER, L.; GE, Y.; GENTRY, J. et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome biology**, [S.l.], v. 5, n. 10, p. R80, 2004.

GOTELLI, N. J. et al. **A primer of ecology**. [S.l.]: Sinauer Associates Incorporated, 1995.

HAMMER, Ø.; HARPER, D.; RYAN, P. Past: paleontological statistics software package for education and data analysis. **Paleontología Electrónica**, [S.l.], v. 4, n. 1, p. 4:9, 2001.

HEPP, L. U.; MILESI, S. V.; BIASI, C.; RESTELLO, R. M. Effects of agricultural and urban impacts on macroinvertebrates assemblages in streams (Rio Grande do Sul, Brazil). **Zoologia (Curitiba)**, [S.l.], v. 27, n. 1, p. 106–113, 2010.

JEON, J.; KIM, J. H.; LEE, B. C.; KIM, S. D. Development of a new biomonitoring method to detect the abnormal activity of *Daphnia magna* using automated Grid Counter device. **Science of the total environment**, [S.l.], v. 389, n. 2, p. 545–556, 2008.

KÄFFER, M. I.; GANADE, G.; MARCELLI, M. P. Lichen diversity and composition in Araucaria forests and tree monocultures in southern Brazil. **Biodiversity and conservation**, [S.l.], v. 18, n. 13, p. 3543–3561, 2009.

KNEITEL, J. M.; LESSIN, C. L. Ecosystem-phase interactions: aquatic eutrophication decreases terrestrial plant diversity in california vernal pools. **Oecologia**, [S.l.], v. 163, n. 2, p. 461–469, 2010.

KOCH, N. M. **Dinâmica da sucessão líquênica**: padrões estruturais e funcionais como indicadores de regeneração florestal. 2012. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal do Rio Grande do Sul, 2012.

KOSICKI, J. Z.; CHYLARECKI, P. Effect of climate, topography and habitat on species-richness of breeding birds in Poland. **Basic and Applied Ecology**, [S.l.], v. 13, n. 5, p. 475–483, 2012.

KRUCHTEN, P. B. The 4+ 1 view model of architecture. **Software, IEEE**, [S.l.], v. 12, n. 6, p. 42–50, 1995.

LAVOREL, S.; STORKEY, J.; BARDGETT, R. D.; BELLO, F.; BERG, M. P.; ROUX, X.; MORETTI, M.; MULDER, C.; PAKEMAN, R. J.; DÍAZ, S. et al. A novel framework for linking functional diversity of plants with other trophic levels for the quantification of ecosystem services. **Journal of Vegetation Science**, [S.l.], v. 24, n. 5, p. 942–948, 2013.

LEGENDRE, P.; LEGENDRE, L. **Numerical ecology**. Segunda Edição. ed. Quebec, Canada: Elsevier, 2012. v. 20.

- LIKERT, R. A technique for the measurement of attitudes. **Archives of psychology**, [S.l.], 1932.
- MARANGUNIĆ, N.; GRANIĆ, A. Technology acceptance model: a literature review from 1986 to 2013. **Universal Access in the Information Society**, [S.l.], p. 1–15, 2014.
- MARCELLI, M. P. **Ecologia líquênica nos manguezais do sul-sudeste brasileiro**. [S.l.]: Gebruder Borntraeger Verlagsbuchhandlung, 1992. v. 47.
- MARKERT, B.; BREURE, A.; ZECHMEISTER, H. **Bioindicators and Biomonitoring Principles, Concepts and Applications**. [S.l.]: Elsevier Science, 2003. (Trace Metals and other Contaminants in the Environment, v. 6).
- MAZANCOURT, C.; ISBELL, F.; LAROCQUE, A.; BERENDSE, F.; LUCA, E.; GRACE, J. B.; HAEGEMAN, B.; WAYNE POLLEY, H.; ROSCHER, C.; SCHMID, B. et al. Predicting ecosystem stability from community composition and biodiversity. **Ecology letters**, [S.l.], v. 16, n. 5, p. 617–625, 2013.
- MCCUNE, B.; GRACE, J. B.; URBAN, D. L. **Analysis of ecological communities**. [S.l.]: MjM software design Gleneden Beach, Oregon, 2002. v. 28.
- MCCUNE, B.; MEFFORD, M. J. **PC-ORD: multivariate analysis of ecological data**. version 6. Gleneden Beach, Oregon, U.S.A.: MjM Software Design, 2011.
- MULTIV. Disponível em: <<http://ecoqua.ecologia.ufrgs.br/ecoqua/MULTIV.html>>. Acesso em: 15 Maio 2014 .
- PAST. Disponível em: <<http://nhm2.uio.no/norlex/past/download.htm>>. Acesso em: 15 de Maio 2014 .
- PC-ORD. Disponível em: <<http://home.centurytel.net/mjm/pcordwin.htm>>. Acesso em: 15 de Maio 2014 .
- PILLAR, V. D.; DUARTE, L. d. S.; SOSINSKI, E. E.; JONER, F. Discriminating trait-convergence and trait-divergence assembly patterns in ecological community gradients. **Journal of Vegetation Science**, [S.l.], v. 20, n. 2, p. 334–348, 2009.
- PILLAR, V. D. P.; ORLÓCI, L. On randomization testing in vegetation science: multifactor comparisons of relevé groups. **Journal of Vegetation Science**, [S.l.], v. 7, n. 4, p. 585–592, 1996.
- PLATA, E. R.; LÜCKING, R.; LUMBSCH, H. T. When family matters: an analysis of thelotremataceae (lichenized ascomycota: ostropales) as bioindicators of ecological continuity in tropical forests. **Biodiversity and Conservation**, [S.l.], v. 17, n. 6, p. 1319–1351, 2008.
- POLZEHL, J.; TABELOW, K. **Adaptive smoothing of digital images: the r package adimpro**. [S.l.]: WIAS, 2006.
- PUGESEK, B. H.; TOMER, A.; VON EYE, A. **Structural equation modeling: applications in ecological and evolutionary biology**. [S.l.]: Cambridge University Press, 2003.
- REENSKAUG, T. Thing-model-view-editor—An example from a planning system. **technical note, Xerox Parc**, [S.l.], 1979.

REENSKAUG, T. Models-views-controllers. **Technical note, Xerox PARC**, [S.l.], v. 32, p. 55, 1979.

RFC 4180. Disponível em: <<http://tools.ietf.org/html/rfc4180>>. Acesso em: 30 de Janeiro 2015 .

ROELSTRAETE, B.; ROSSEEL, Y. FIAR: an r package for analyzing functional integration in the brain. **J. Stat. Soft.**, [S.l.], v. 44, p. 1–32, 2011.

ROSENBLUETH, A.; WIENER, N.; BIGELOW, J. Behavior, purpose and teleology. **Philosophy of science**, [S.l.], v. 10, n. 1, p. 18–24, 1943.

RSTUDIO. Disponível em: <<https://www.rstudio.com>>. Acesso em: 15 de Maio 2014 .

SHIPLEY, B. Confirmatory path analysis in a generalized multilevel context. **Ecology**, [S.l.], v. 90, n. 2, p. 363–368, 2009.

SILVA, J. da; FREITAS, T. R. de; HEUSER, V.; MARINHO, J. R.; ERDTMANN, B. Genotoxicity biomonitoring in coal regions using wild rodent *Ctenomys torquatus* by comet assay and micronucleus test. **Environmental and molecular mutagenesis**, [S.l.], v. 35, n. 4, p. 270–278, 2000.

SMART PLS. Disponível em: <<http://www.smartpls.de>>. Acesso em: 15 de Maio 2014 .

SPSS. Disponível em: <<http://www-01.ibm.com/software/analytics/spss/>>. Acesso em: 15 de Maio 2014 .

STATISTICA. Disponível em: <<http://www.statsoft.com/products/statistica/high-performance>>. Acesso em: 15 de Maio 2014 .

SYNCSA package for R. Disponível em: <<http://cran.r-project.org/web/packages/SYNCSA/index.html>>. Acesso em: 15 de Maio 2014 .

SYNCSA. Disponível em: <<http://ecoqua.ecologia.ufrgs.br/ecoqua/SYNCSA.html>>. Acesso em: 15 de Maio 2014 .

TER BRAAK, C. J. Permutation versus bootstrap significance tests in multiple regression and ANOVA. In: **Bootstrapping and related techniques**. [S.l.]: Springer, 1992. p. 79–85.

THE R Project for Statistical Computing. Disponível em: <<http://www.r-project.org/>>. Acesso em: 15 de Maio 2014 .

THE R Project. Disponível em: <<http://www.R-project.org>>. Acesso em: 15 de Maio 2014 .

THE S System. Disponível em: <<http://stat.bell-labs.com/S/>>. Acesso em: 15 de Maio 2014 .

WALKER, L. R.; SHIELS, A. B.; BELLINGHAM, P. J.; SPARROW, A. D.; FETCHER, N.; LANDAU, F. H.; LODGE, D. J. Changes in abiotic influences on seed plants and ferns during 18 years of primary succession on Puerto Rican landslides. **Journal of Ecology**, [S.l.], v. 101, n. 3, p. 650–661, 2013.

WILSON, J. L.; AYRES, D. R.; STEINMAUS, S.; BAAD, M. Vegetation and flora of a biodiversity hotspot: pine hill, el dorado county, california, usa. **Madroño**, [S.l.], v. 56, n. 4, p. 246–278, 2009.

APÊNDICE A CASO DE TESTE

CASO DE TESTE PARA AVALIAÇÃO DE ACEITE DO SOFTWARE BIOMON

1. DESCRIÇÃO

O presente caso de teste descreve os passos necessários para a execução de um cenário de análise de dados de campo em biomonitoramento fazendo uso do protótipo Biomon com o objetivo de realizar sua avaliação.

2. PREPARAÇÃO

Dados de Campo

Os dados de campo devem estar organizados em planilhas eletrônicas onde:

- A primeira linha descreve cada coluna;
- A primeira coluna descreve cada linha;
- Ao final, a planilha deve ser salva como arquivo CSV.

Estas planilhas devem representar as seguintes matrizes:

- Atributos (B) - espécies descritas por atributos;
- Espécies (W) - unidade amostral descrita pelas espécies;
- Ambiente (E) - unidade amostral descrita por variáveis ambientais;
- Espaço (S) - unidade amostral descrita por unidades de localização;
- Fator (T) - resultado da multiplicação da matriz W pela transposta B.

Ambiente R

Para execução do Biomon o testador necessita:

- Acesso a uma computador com o ambiente R/RStudio e internet;
- Para a instalação, alguns pacotes suplementares são necessários:


```
install.packages("devtools")
install.packages("shiny")
```
- O Biomon é instalado com os comandos abaixo:


```
devtools::install_github("rbuhler/analyz")
devtools::install_github("rbuhler/biomonCore")
devtools::install_github("rbuhler/biomon")
```
- Depois de instalado sempre que precisar acessar o Biomon execute os comandos abaixo:


```
library("biomon")
```

biomon()

- O Biomon será aberto em uma janela de seu navegador de internet;
- Execute um teste simples, a partir do quadro a esquerda escolha o arquivo de análise *Sample_Mean.csv* e pressione o botão *Execute*, você deve obter como resposta o número 5;
- Se tudo funcionou bem o software está pronto para ser utilizado, quando não quiser mais usá-lo simplesmente feche o navegador e tecle “ESC” no ambiente R para encerrar a execução.

3. REALIZAÇÃO

- Execute o software Biomon;
- + Preparação
 - Faça o *upload* de um arquivo CSV de análise a partir da interface;
 - Da mesma forma faça o *upload* das matrizes de dados de campo;
 - Desta forma estes arquivos ficam disponíveis para análise no ambiente Biomon;
 - Atualize a página do navegador (isso será corrigido em uma nova versão).
- + Execução
 - A partir do quadro a esquerda (*Analysis*) escolha o arquivo de análise a ser executado;
 - Em seguida escolha a(s) matriz(es) (*Data Matrices*) necessária(s) para a análise indicada;
 - + Por exemplo:
 - Para o cálculo da matriz T são necessárias apenas as matrizes de atributos B e espécies W;
 - Esta informação deve ser combinada entre o estatístico e o pesquisador previamente.
 - Selecione o botão *Execute*.
- + Resultado
 - Confira se a execução trouxe os resultados esperados.

4. OUTRAS INFORMAÇÕES

Final do Teste

- Repita este processo quantas vezes julgar necessário com outras matrizes ou solicite novos arquivos de análise;

- Ao final responda ao questionário de avaliação de sua experiência em:
<http://goo.gl/forms/OSqClNaWwl>.

Fevereiro de 2015.

Muito obrigado,

Rodrigo Bühler

APÊNDICE B AVALIAÇÃO

Questionário de Avaliação de Aceite de Software

Caro(a) pesquisador(a),

Como forma de avaliar a proposta peço que responda o questionário abaixo com base em sua opinião pessoal a respeito da sua experiência executando os passos do caso de teste fornecido. Não existem respostas certas ou erradas o que se procura com esta avaliação é realmente validar a proposta como ferramenta de apoio à análise de dados de campo. Ao final você é convidado(a) a fornecer um relato textual que julgar conveniente para, com seus comentários, contribuir ainda mais para a presente avaliação.

Desde já agradeço a sua participação.

Instrução

Considerando sua experiência quando utilizando o Biomon e levando em conta de se tratar de um protótipo para avaliação de um conceito, como você avaliaria as afirmações abaixo:

- **Compreendi a função de todos os elementos da tela sem ajuda externa.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **A interface disponibiliza todas as funções necessárias para a operação do software.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **Pude seguir o caso de teste sem dificuldades.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **De maneira geral achei o Biomon fácil de utilizar.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **Achei a interface simples e agradável.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **Considerando a forma como analiso meus dados de campo considero que o Biomon pode tornar meu trabalho mais fácil.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **Eu substituiria a maneira como faço minhas análises para utilizar apenas o Biomon.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **Gostei da maneira como a proposta separa os papéis do estatístico e do pesquisador.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente
- **Considero importante a possibilidade de facilmente compartilhar os arquivos de análise.**
Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente

- **Com o Biomon eu posso economizar tempo executando as análises e investir mais na interpretação dos resultados.**

Concordo totalmente ☐¹ ☐² ☐³ ☐⁴ ☐⁵ Discordo totalmente

- **O espaço abaixo é reservado para seus comentários a respeito da experiência.**