



Maria d'Amaral Ferreira

Licenciada em Engenharia Informática

Visualização Interativa de Dados para Projetos de Monitorização Fitossanitária

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática

Orientador: João Moura Pires, Professor Associado,
Universidade Nova de Lisboa

Co-orientador: Carlos Viegas Damásio, Professor Associado,
Universidade Nova de Lisboa

Júri

Presidente: Carla Maria Gonçalves Ferreira
Arguente: Daniel Jorge Viegas Gonçalves
Vogal: João Moura Pires



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro, 2021

Visualização Interativa de Dados para Projetos de Monitorização Fitossanitária

Copyright © Maria d’Amaral Ferreira, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Ao meu pai e pela minha mãe.

AGRADECIMENTOS

Ao meu orientador, Prof. João Moura Pires, agradeço toda a disponibilidade, apoio e motivação ao longo desta dissertação, por sempre acreditar no meu potencial e insistir até que eu alcance a melhor versão de tudo o faço. Agradeço tudo o que me ensinou e tudo o que fez e contribuiu para o meu crescimento pessoal e profissional.

Ao meu co-orientador, Prof. Carlos Damásio, agradeço por ter reconhecido o meu potencial e me introduzido a oportunidades nesta área, por todo o conhecimento que me transmitiu e pelas ideias e opiniões que expressou ao longo desta dissertação.

Este trabalho foi parcialmente financiado pelo PDR2020 Operação 1.0.1-FEADER-031678 - FITOAGRO NIFAP 9026322A, sendo que agradeço a todos os membros do grupo operacional por toda a colaboração e disponibilidade que apresentaram ao longo de todo o trabalho desenvolvido. Em especial, à Sra. Carmo Martins e Carolina Lucas do COTHN por terem acreditado na qualidade do meu trabalho e, por isso, selecionado-me para o apresentar em painel no Congresso Frutos 2020, em representação de todo o grupo.

Ao meu pai e irmãos por todo o valor que deram ao meu trabalho e pela compreensão que demonstraram perante a falta de disponibilidade que por vezes ocorreu. Ao meu pai agradeço todos os esforços que fez para que eu aqui chegasse e por sempre acreditar e lutar por mim e pelo meu futuro. À minha mãe, que embora não esteja comigo, sempre aspirou para que eu lutasse pelos meus estudos e pelo meu futuro. Pai e mãe, sem vocês eu não teria conseguido chegar até aqui, estarei para sempre grata por tudo o que fizeram por mim.

Ao meu colega e amigo Miguel Correia por todo o companheirismo e amizade, nos altos e baixos tivemos sempre as costas um do outro, sem ti este projeto e dissertação não seria igual. À Salúquia Marreiros pela paciência que teve em ouvir os meus desabafos e frustrações, por me apoiar e motivar a não desistir. Aos dois, um grande obrigado.

Por fim, aos meus amigos pelo apoio, por me obrigarem a espaiar de vez em quando e pelo interesse que mostraram no meu trabalho, mesmo que nem sempre o percebessem na totalidade. Em especial, à família João por me ter recebido de braços abertos e, principalmente, por valorizarem o meu trabalho, disponibilizando todas as condições necessárias para que eu o terminasse com sucesso.

A todos vós, o meu eterno obrigado.

Visualization gives you answers to questions you didn't know you had.

— Ben Schneiderman

RESUMO

Ao longo dos tempos, as pragas têm sido responsáveis por causar diversos prejuízos económicos e sociais para os produtores agrícolas. Atualmente, a produção de pomóideas, como as pereiras e macieiras, na região do Oeste confronta-se com diversos problemas fitossanitários devido ao aparecimento de novas pragas na região e ao aumento do impacto económico das pragas já existentes. Deste contexto surgiu o projeto onde se enquadra esta dissertação, o projeto FitoAgro, cujo objetivo é delinear uma estratégia de controlo regional para a produção de pomóideas nesta região. Assim, o grupo operacional deste projeto visa monitorizar e estudar estas pragas e os seus ciclos de vida de forma a elaborar planos de risco fitossanitário e especificar estratégias de prevenção e controlo de base biológica e biotecnológica. No entanto, para analisar os dados recolhidos é necessário uma componente de visualização de dados que permita identificar os fatores influentes no desenvolvimento das pragas.

Deste modo, o principal objetivo desta dissertação é o desenvolvimento de uma componente de visualização para o grupo operacional FitoAgro que não só responda às necessidades analíticas do mesmo, mas que também promova a descoberta de correlações entre os diversos tipos de dados recolhidos e o desenvolvimento das pragas. Para isso, esta componente irá acomodar dados recolhidos em observações de campo e em estações meteorológicas, permitindo a sua análise ao longo de diferentes séries temporais e a observação do ajuste de modelos preditivos aos mesmos. Para além da sua aplicabilidade ao projeto FitoAgo, esta componente de visualização terá algum carácter genérico de modo a seja facilmente adaptada no futuro para outros projetos e áreas relevantes no contexto da monitorização fitossanitária.

Palavras-chave: Visualização interativa de dados, dados meteorológicos, dados recolhidos em campo, pragas de culturas, ajuste de modelos preditivos, fitossanidade.

ABSTRACT

Throughout times, crop pests have been responsible for causing various economical and social losses to agricultural producers. Nowadays, the pome fruit production, like pear and apple trees, in the West region is facing various phytosanitary problems due to new pests emerging in the region and to the growing economical impact of the existing ones. From this context surfaced the project where this dissertation lies, the FitoAgro project, whose goal is to outline a regional control strategy for the pome fruit production in this region. Thus, the operational group of this project aims to monitor and study these pests and their life cycles, in order to elaborate phytosanitary risc plans and specify prevention and control strategies of biological and biothechnological nature. To analyse the data gathered, a data visualization component that allows the identification of influential factors in the crop pests development is needed.

Therefore, the main goal of this dissertation is to develop a visualization component for the FitoAgro operational group that, not only responds to the analytical requirements of the project, but also promotes the correlation discovery between the diferent type of data collected and the crop pests development. To achieve this, the component will acommodate data from field observations and meteorological stations, allowing their analysis using different time series and the observation of the ajustment of the predictive models to this data. Besides its aplication to the FitoAgro project, this visualization component will have some degree of generalization in order to be easily adapted in the future to other projects and relevant areas in the phytosanitary monitorization context.

Keywords: Interactive data visualization, meteorological data, field data, crop pests, ajustment of prediction models, phytosanitary.

ÍNDICE

Lista de Figuras	xix
Lista de Tabelas	xxv
Glossário	xxvii
Siglas	xxxii
1 Introdução	1
1.1 Contexto e Motivação	1
1.2 Grupo Operacional FitoAgro	2
1.3 Descrição do Problema	3
1.4 Contributos	4
1.5 Estrutura do Documento	4
2 Trabalho Relacionado	7
2.1 Projetos e Aplicações Relacionadas	7
2.1.1 Cropio	7
2.1.2 FieldClimate	8
2.1.3 ProTomate	9
2.1.4 SafeBrócolo	9
2.1.5 WiseCrop	11
2.1.6 Conclusão	12
2.2 Visualização Interativa de Dados	13
2.2.1 Fundamentos	13
2.2.2 Técnicas	17
2.2.3 Ferramentas	23
2.2.4 Domo	25
3 Grupo Operacional FitoAgro	29
3.1 Problema nas Pomóideas	29
3.1.1 Cecidómia	30
3.1.2 Cochonilha-Algodão	31
3.1.3 Filoxera	32

3.1.4	Novo Lepidóptero	32
3.2	Recolha de Dados	33
3.2.1	Dados Fitossanitários	33
3.2.2	Dados Meteorológicos	36
3.3	Modelos Preditivos	37
3.4	<i>Stakeholders</i>	38
4	Requisitos Analíticos	41
4.1	Monitorização e Controlo de Qualidade	41
4.1.1	Estações Meteorológicas	41
4.1.2	Visitas de Campo	42
4.2	Analítica dos Dados	43
4.2.1	Sensores Meteorológicos	43
4.2.2	Observações Fitossanitárias	43
4.2.3	Correlação da Meteorologia com a Fitossanidade	44
4.2.4	Ajuste de Modelos Preditivos	44
4.3	Resumo	45
5	Abordagem e Metodologia	47
5.1	Definição de Boas Práticas de Desenho	47
5.1.1	Modelação de Visualizações Interativas de Dados	47
5.1.2	Recomendações Gerais	48
5.2	Processo Iterativo de Avaliação	50
5.3	Modelação da Base de Dados Analítica	51
6	Meteorologia - Desenho e Implementação	55
6.1	Monitorização do Estado e Qualidade das Estações	55
6.2	Analítica dos Dados Meteorológicos	59
6.2.1	Sensores Meteorológicos	59
6.2.2	Variáveis Derivadas	66
7	Fitossanidade - Desenho e Implementação	71
7.1	Monitorização e Controlo das Visitas de Campo	71
7.2	Analítica dos Dados Fitossanitários	75
7.2.1	Evolução das Observações Fitossanitárias	75
7.2.2	Comparação entre anos e POBs	87
7.3	Impacto da Meteorologia na Fitossanidade	89
7.3.1	Importância dos Graus-dia Acumulados	89
7.3.2	Ajuste de Modelos Preditivos	92
8	Avaliação e Discussão	97
8.1	Metodologia de Avaliação	97

8.2	Questionário	97
8.2.1	Análise de Resultados	99
8.2.2	Satisfação dos Requisitos Analíticos	103
8.3	Adaptação a Outros Contextos	104
8.4	Conclusão	107
9	Conclusões e Trabalho Futuro	109
9.1	Conclusão	109
9.2	Trabalho Futuro	110
9.2.1	Melhoria e extensão da linguagem IVML	110
9.2.2	Ferramenta para a construção de visualizações	110
9.2.3	Visualizar a aplicação de modelos preditivos	111
9.2.4	Reaplicabilidade do ajuste de componentes temporais ao longo de funções monótonas não-decrescentes temporalmente dependentes	111
	Bibliografia	113
A	Resumo dos Requisitos Analíticos	117
B	Descrição Técnica do IVML	119
B.1	Descrição da Linguagem	119
B.1.1	Componentes de Visualização	121
B.1.2	Componentes de <i>Layout</i>	125
B.1.3	Componentes de Interação	127
C	Variantes do Dashboard Generalista	133
C.1	Aranhinho Vermelho	133
C.2	Broca	134
C.3	Cochonilha São José	134
C.4	Estenfiliose	135
C.5	Hoplocampa	135
C.6	Lagarta Mineira	136
C.7	Mosca Mediterrânica	136
C.8	Oídio	137
C.9	Pedrado	137
C.10	Psila	138
C.11	Pulgão Lanígero	138
D	Questionário para Avaliação da Componente	139
I	Levantamento de Momentos Críticos	145

LISTA DE FIGURAS

2.1	Exemplo da vista geral do sistema Cropio. [4]	8
2.2	Exemplo de interface da plataforma FieldClimate. [5]	8
2.3	Mapa de risco da plataforma <i>web</i> do projeto ProTomate. [7]	9
2.4	Exemplo de páginas da plataforma web do projeto SafeBrócolo. [9]	10
2.5	Exemplo de alguns módulos relevantes do sistema da WiseCrop. [10]	11
2.6	Exemplo de formas disponíveis na plataforma Tableau [13]	14
2.7	Exemplos de orientações de uma forma [14]	15
2.8	Gráfico de dispersão que representa a esperança média de vida em 2018 por rendimento do país, onde o tamanho representa a população do país e a cor o continente da mesma (Verde - América, Azul - África, Amarelo - Europa, Vermelho - Ásia) [15].	16
2.9	Canais ordenados por eficácia de acordo com o tipo de dados e de canal [16]	17
2.10	Exemplos de gráficos que utilizam técnicas baseadas em linhas [17] [18] [19]	18
2.11	Exemplos de gráficos que utilizam técnicas baseadas em regiões [17]	19
2.12	Algumas características de dados temporais [20]	20
2.13	Exemplos de técnicas de visualização para dados temporais multivariados [21]	21
2.14	Exemplos de mapas de pontos	22
2.15	Exemplo de um mapa de linhas [24].	22
2.16	Exemplos de mapas de áreas	23
3.1	Ciclo de vida de <i>Dasineura pyri Bouché</i> [39]	30
3.2	Folhas enroladas com larvas de <i>Dasineura pyri Bouché</i> [40]	30
3.3	Ciclo de vida de uma cochonilha-algodão: 1 - ovos; 2 - ninfa de 1º instar; 3 - ninfa de 2º instar; 4 - ninfa de 3º instar; 5 - fêmea adulta; 6- pré-pupa e pupa macho; 7 - macho adulto. [41]	31
3.4	fêmea adulta e danos no fruto de <i>Pseudococcus viburni</i>	32
3.5	Colónias de filoxera na primavera (fêmeas, ovos e larvas) [42]	32
3.6	Larva e traça adulta de <i>Grapholita molesta</i>	33
3.7	Mapa da região do Oeste onde se localizam os treze pontos de observação biológica (POB)s do projeto FitoAgro, identificados por cultura (dados de 2021).	34
4.1	Exemplo de falhas e inconsistências nos valores diários do sensor temperatura do ar da estação Apas04 (APAS).	42

5.2	Comparação visual pode induzir em erro dependendo do tipo de eixo utilizado na visualização.	49
5.3	Fluxograma do processo iterativo de avaliação que deverá ser seguido para visualização de dados.	50
5.4	Camada de armazenamento do sistema de informação, cuja base de dados analítica contém o <i>data warehouse</i> e o <i>publish</i>	51
5.5	Modelo de dados da base de dados analítica, com indicação das tabelas que pertencem a cada vista materializada (<i>data warehouse</i> ou <i>publish</i>) e as que estão em ambas (duplicadas).	52
6.1	Visualização que apresenta o estado e tipo das estações meteorológicas para 4 de fevereiro de 2021.	56
6.2	<i>Tooltip</i> da visualização na figura 6.1 com informação acerca da estação Leiria (Aeródromo) do IPMA e do estado diário dos seus sensores nos últimos 30 dias, referentes a 4 de fevereiro de 2021.	57
6.3	Diagrama de modelação IVML que representa o <i>dashboard</i> da figura 6.1 . . .	57
6.4	Visualização que apresenta a qualidade diária das estações nos últimos 30 dias (para 8 de fevereiro de 2021), com uma <i>tooltip</i> visível.	58
6.5	Diagrama que apresenta um resumo de todos os <i>dashboards</i> propostos para a monitorização e analítica dos dados provenientes de sensores meteorológicos, bem como formas de navegação que existam entre os mesmos.	60
6.6	<i>Dashboard</i> que apresenta os dados meteorológicos em bruto provenientes da estação Apas05Do, ao longo do dia 30 de novembro de 2020.	61
6.7	Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a temperatura do ar e ponto de orvalho na estação Apas05Do.	61
6.8	Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a humidade relativa na estação Apas05Do.	62
6.9	Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a folha molhada, na estação Apas05Do, e o cálculo da depressão do ponto de orvalho.	62
6.10	Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a precipitação acumulada, na estação Apas05Do, e o cálculo do máximo por hora e da média do dia, numa escala logarítmica <i>versus</i> linear.	63
6.11	Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a velocidade do vento na estação Apas05Do.	63
6.12	Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a radiação solar na estação Apas05Do.	63
6.13	<i>Dashboard</i> que apresenta os dados meteorológicos da estação Apas05Do, ao longo de um período de 3 meses centrado em 12 de novembro de 2020. . . .	64

6.14	Uma <i>tooltip</i> da visualização da figura 6.13, onde se apresenta os valores agregados para o dia selecionado (29 de novembro de 2020) e uma visualização da evolução desses sensores ao longo desse dia.	64
6.15	<i>Dashboard</i> que apresenta os dados meteorológicos da estação Apas05Do, horariamente ao longo do dia 29 de novembro de 2020.	65
6.16	Replicação do <i>dashboard</i> da figura 6.13 que apresenta os dados meteorológicos da estação virtual do POB Termas, ao longo de um período de 3 meses centrado em 12 de novembro de 2020.	66
6.17	Evolução da temperatura média por hora, entre outubro de 2019 e março de 2020, no POB Quinta do Rato, com indicação de horas de frio.	67
6.18	Visualização das horas de frio acumuladas, entre outubro de 2019 e março de 2020, no POB Quinta do Rato.	68
6.19	Visualização dos graus-dia, entre janeiro e outubro de 2020, no POB Quinta do Rato.	69
6.20	Visualização das diferentes fases de desenvolvimento do bichado conforme os graus-dia acumulados, entre janeiro e outubro de 2020, no POB Quinta do Rato.	70
7.2	Página inicial (após <i>login</i>) do <i>website</i> complementar à aplicação de recolha de dados fitossanitários para o projeto FitoAgro, com e sem agrupar por parcela.	72
7.4	Página da visita ao POB de São Lourenço (SL) no dia 7 de abril de 2021.	74
7.5	Página <i>web</i> "Estatísticas" que contém um conjunto de visualizações com estatísticas básicas sobre a recolha de dados.	75
7.6	Visualização das observações registadas para a cecidómia ao longo de 2020, no POB Quinta do Brejo (QB).	77
7.7	Visualização que apresenta uma visão holística das visitas ao longo de 2020, no POB Quinta do Brejo.	77
7.8	Visualização que apresenta os estados fenológicos registados ao longo de 2020, no POB Quinta do Brejo.	78
7.9	Visualização que apresenta os eventos que ocorreram nas visitas de 2020, no POB Quinta do Brejo.	78
7.10	Visualização que apresenta os tratamentos efetuados ao longo de 2020, no POB Vale Leite.	79
7.11	<i>Dashboard</i> com as observações fitossanitárias de cecidómia e outros fatores relevantes que ocorreram no POB Quinta do Brejo, ao longo do ano de 2020.	79
7.12	<i>Dashboard</i> com as observações fitossanitárias de cochonilha-algodão e outros fatores relevantes que ocorreram no POB Vale Leite, ao longo do ano de 2020.	80
7.13	Visualizações para as fases I e fase II, onde se analisa a extensão e intensidade do ataque de filoxera ao longo de 2020, no POB Encarnação.	82

7.14	Visualização para a fase III e IV, onde se analisa os sintomas dos frutos e o número de indivíduos que contêm de cada estágio, para o POB Encarnação em 2019.	83
7.15	<i>Dashboard</i> com a extensão do ataque de filoxera e outros fatores relevantes que ocorreram no POB Encarnação, ao longo de 2020.	83
7.16	<i>Dashboard</i> com as observações fitossanitárias das diferentes espécies de lepidópteros, bem como outros fatores relevantes que ocorreram no POB Termas, em 2020.	84
7.17	<i>Dashboard</i> com as observações fitossanitárias das diferentes espécies de lepidópteros, bem como outros fatores relevantes que ocorreram no POB Vale do Leite, ao longo de 2020.	85
7.18	<i>Dashboard</i> generalista com as observações fitossanitárias da sésia, bem como outros fatores relevantes que ocorreram no POB João Boi, ao longo de 2020. .	86
7.19	Estrutura genérica dos <i>dashboards</i> para observação dos dados fitossanitários, com indicação da visualização que varia por inimigo e da variante para a filoxera.	87
7.20	<i>Dashboard</i> que permite comparar toda a informação recolhida para a cecidómia nos POBs Quinta do Brejo e Casal Serrano, ao longo do ano de 2020. . .	88
7.21	<i>Dashboard</i> que permite comparar toda a informação recolhida para a cecidómia ao longo de 2020 e 2021, no POB Quinta do Brejo.	88
7.22	Graus-dia acumulados no POB Quinta do Brejo, ao longo de 2020.	90
7.23	Faixa de graus-dia acumulados no POB Quinta do Brejo, ao longo de 2020. .	90
7.24	<i>Dashboard</i> que permite comparar toda a informação de cecidómia recolhida nos POBs Quinta do Brejo e Casal Serrano, ao longo dos graus-dia acumulados em 2020.	91
7.25	<i>Dashboard</i> que permite comparar toda a informação de cecidómia recolhida ao longo dos graus-dia acumulados em 2020 e 2021, no POB Quinta do Brejo.	91
7.26	<i>Dashboard</i> da figura 7.25 com o modelo de desenvolvimento da praga visível sob as observações fitossanitárias registadas.	93
7.27	Visualização para observar o ajuste dos modelos preditivos de cecidómia às observações fitossanitárias registadas num POB (Quinta do Brejo) em 2020. .	94
7.28	Visualização para observar o ajuste de um dos modelos preditivos de cecidómia às observações fitossanitárias registadas em todos os POBs.	95
8.1	Número de respostas corretas/incorretas por pergunta.	99
8.2	Percentagem de respostas corretas/incorretas por pergunta.	100
8.3	Respostas obtidas nas perguntas com mais de 20% de respostas incorretas. .	100
8.4	Respostas obtidas por tipo e pela ação que requer.	102
8.5	Percentagem de respostas obtidas por tipo de <i>stakeholder</i>	102
8.6	Percentagem obtida por requisito analítico, em cada nível da escala de <i>Likert</i> .	103

8.7	Porcentagem obtida por requisito analítico, em cada nível da escala de <i>Likert</i> e ordenada de forma ascendente pela percentagem de avaliações de nível 4 e categorizado por tipo de informação e abordagem utilizada.	104
8.8	Dados Meteorológicos do IPMA referentes a 2018 para Lisboa, apresentados com a componente desenvolvida.	105
8.9	Evolução do número de casos positivos em Portugal em relação aos restantes países, ao longo do número de dias a partir do 1º caso positivo confirmado em cada país. Fonte: <i>Our World in Data</i> [58].	106
8.10	Evolução temporal do número de pacientes nos cuidados intensivos e do acumulado de casos positivos em Portugal e em Itália (por 1 milhão de habitantes).	106
8.11	Evolução do número de pacientes nos cuidados intensivos ao longo do acumulado de casos positivos em Portugal e em Itália (por 1 milhão de habitantes).	107
B.1	Exemplo do esqueleto de um diagrama de interações, com três interações.	120
B.2	Exemplo do esqueleto de um diagrama de <i>tooltips</i> , com três <i>tooltips</i>	120
B.3	Componente <i>Dashboard</i>	121
B.4	Componente Visualização.	121
B.5	Componente <i>Tooltip</i>	122
B.6	Componente Legenda.	122
B.8	Ícones para identificar variáveis visuais.	123
B.10	Ícones para identificar o tipo de gráfico.	123
B.11	Exemplo de diagramas <i>interactive visualization modeling language</i> (IVML) apenas com componentes de visualização.	124
B.13	Componente visualização com eixo X partilhado	125
B.15	Componente visualização com eixo Y partilhado	125
B.16	Componente visualização com eixos X e Y partilhados.	126
B.17	Componente <i>Placeholder</i>	126
B.18	Exemplo do diagrama de interações IVML incompleto, contendo apenas componentes de visualização e <i>layout</i>	127
B.19	Componente Botão.	127
B.20	Componente Filtro.	127
B.21	Componente Parâmetro.	128
B.22	Componente Ação de Dados.	128
B.23	Componente <i>Parameter Binding</i>	129
B.25	Ícones para os tipos de interação.	129
B.27	Ícones para os tipos de resultados das ações.	130
B.29	Ícones para os tipos de navegação.	130
B.31	Tipos de setas utilizadas para conexão de componentes.	130
B.32	Exemplo de diagrama de interações IVML completo.	131
B.33	<i>Dashboard</i> de exemplo do <i>Tableau</i> para o conjunto de dados " <i>Superstore</i> ", com algumas alterações e uma <i>tooltip</i> ativa.	131

LISTA DE FIGURAS

C.1	<i>Dashboard</i> generalista com as observações fitossanitárias de aranhaço vermelho, bem como outros fatores relevantes que ocorreram no POB João Boi, em 2020.	133
C.2	<i>Dashboard</i> generalista com as observações fitossanitárias de broca, bem como outros fatores relevantes que ocorreram no POB Termas, ao longo de 2020.	134
C.3	<i>Dashboard</i> generalista com as observações fitossanitárias de cochonilha São José (sem registros), bem como outros fatores relevantes que ocorreram no POB Casal Serrano, ao longo de 2020.	134
C.4	<i>Dashboard</i> generalista com as observações fitossanitárias de estenfiliose, bem como outros fatores relevantes que ocorreram no POB Boiça do Louro, ao longo de 2020.	135
C.5	<i>Dashboard</i> generalista com as observações fitossanitárias de hoplocampa (sem registros), bem como outros fatores relevantes que ocorreram no POB Quinta do Brejo, ao longo de 2020.	135
C.6	<i>Dashboard</i> generalista com as observações fitossanitárias de lagarta mineira, bem como outros fatores relevantes que ocorreram no POB Termas, ao longo de 2020.	136
C.7	<i>Dashboard</i> generalista com as observações fitossanitárias de mosca mediterrânica, bem como outros fatores relevantes que ocorreram no POB Encarnação, em 2020.	136
C.8	<i>Dashboard</i> generalista com as observações fitossanitárias de oídio, bem como outros fatores relevantes que ocorreram no POB Termas, ao longo de 2020.	137
C.9	<i>Dashboard</i> generalista com as observações fitossanitárias de pedrado, bem como outros fatores relevantes que ocorreram no POB João Boi, ao longo de 2020.	137
C.10	<i>Dashboard</i> generalista com as observações fitossanitárias de psila, bem como outros fatores relevantes que ocorreram no POB Casal Serrano, ao longo de 2020.	138
C.11	<i>Dashboard</i> generalista com as observações fitossanitárias de pulgão Lanígero, bem como outros fatores relevantes que ocorreram no POB João Boi, ao longo de 2020.	138

LISTA DE TABELAS

2.1	Resumo das funcionalidades-chave das ferramentas apresentadas.	26
3.1	Cultura e inimigo principal de cada POB (dados de 2021).	34
3.2	Estados fenológicos por cultura.	35
3.3	Sensores selecionados como relevantes e a sua unidade correspondente. . . .	37
3.4	Sensores selecionados como relevantes e a sua respetiva unidade.	37
4.1	Mapeamento dos requisitos analíticos aos stakeholders a que se associam. . .	45
6.1	Valores do somatório de temperaturas médias diárias determinadas para as várias fases de desenvolvimento do bichado (<i>Cydia pomonella</i>), na região da Beira Interior de Portugal [54].	69
8.1	Nº Questões por secção e enumeração dos requisitos analíticos avaliados. . .	98
A.1	Resumo dos requisitos analíticos levantados, classificados por tipo.	118

GLOSSÁRIO

armadilhas sexuais	Armadilha que contém feromonas adequadas à praga da qual se pretende proteger. Estas feromonas correspondem às que são geralmente emitidas pela fêmea e que tem como função atrair sexualmente o macho que consegue detetar a vários quilómetros de distância.
armadilhas cromotrópicas	Armadilhas compostas por placas de cor amarela ou azul, que atraem visualmente os insetos voadores. Estes são capturados na sua superfície, que se encontra coberta por uma cola não-secante fortemente adesiva.
cintas	Armadilhas com cola que se aplicam ao redor dos troncos das árvores de fruto e ornamentais, protegendo as mesmas contra insectos rastejantes que possam causar danos nas copas ou nos frutos.
dashboard	Painéis que mostram visualmente métricas e indicadores importantes para alcançar objetivos e metas traçadas, facilitando a compreensão das informações geradas.
diapausa	Suspensão temporária do desenvolvimento de um embrião, larva ou pupa, ou redução da atividade de um inseto adulto, geralmente provocadas por condições ambientais adversas.
díptero	Insetos que pertencem ao grupo <i>Diptera</i> , cuja principal característica é a presença de duas asas.
estado fenológico	Etapas de desenvolvimento da planta que são facilmente identificáveis por alterações da morfologia externa da mesma (e.g. floração).

fitossanitário	Que diz respeito ao estado de saúde das espécies vegetais.
framework	Software que tem como principal objetivo resolver problemas recorrentes com uma abordagem genérica, como por exemplo o desenvolvimento de aplicações web.
homóptero	Insetos de pequenas dimensões que pertencem à subordem <i>Homoptera</i> , cuja característica determinante é a presença de asas anteriores membranosas.
hortofrutícola	Que produz ou comercializa produtos hortícolas e frutos simultaneamente.
lepidópteros	Insetos da ordem <i>Lepidoptera</i> , tipicamente com dois pares de asas escamosas, armadura bucal sugadora e com metamorfoses completas. Normalmente denominados de borboletas.
nível económico de ataque	Intensidade de ataque de um inimigo da cultura a que se devem aplicar medidas limitativas ou de combate para impedir que a cultura corra riscos de prejuízos superiores ao custo das medidas de luta a adotar, acrescido dos efeitos indesejáveis que estas últimas podem provocar.
overhead	Excesso de tempo de computação, memória ou outros recursos necessários para executar uma tarefa.
parcela	Constitui a porção contínua de terreno homogéneo com limites estáveis agrónomica e geograficamente, com uma identificação única conforme registado no Sistema de Identificação Parcelar.
pomóideas	Designação das árvores e arbustos cujos frutos comestíveis são pomos como a maçã e a pêra.
pontos de observação biológica	Parcelas que têm o objetivo de monitorizar um ou mais problemas fitossanitários.

praga	Surto de insectos nocivos ao desenvolvimento agrícola.
produtos fitofarmacêuticos	Produtos químicos utilizados para combater e evitar pragas e doenças agrícolas, protegendo as plantas e as culturas.
região do Oeste	Região delimitada a sul pela Grande Lisboa, a leste pela Lezíria do Tejo, a norte pelo Pinhal Litoral e a oeste pelo Oceano Atlântico.
stakeholder	Indivíduo, grupo ou organização que pode afetar ou ser afetada pelas decisões, atividades ou resultados de um projeto.
tooltip	Vista que aparece quando o cursor é posicionado sobre um elemento da interface gráfica.

SIGLAS

DOM *document object model.*

EOI *element of interest.*

IFML *interactive flow modeling language.*

IVML *interactive visualization modeling language.*

NEA *nível económico de ataque.*

POB *pontos de observação biológica.*

INTRODUÇÃO

Neste capítulo é descrito o contexto e motivação para esta dissertação (1.1), bem como o projeto onde esta se enquadra (1.2). Além disso, é apresentada uma descrição do problema (1.3), os objetivos que se pretende alcançar e os contributos esperados desta dissertação (1.4). Por fim, explica-se a estrutura deste documento com uma enumeração dos seus capítulos, seguidos de uma breve descrição (1.5).

1.1 Contexto e Motivação

Ao longo da história da agricultura e da humanidade, as **pragas** têm sido responsáveis por prejuízos económicos e sociais nefastos, como a extinção de recursos naturais e períodos de fome. Atualmente as **pragas** não apresentam consequências tão adversas na sociedade como antes, mas ainda trazem problemas para os agricultores que podem resultar em graves perdas económicas. Para combaterem as **pragas**, os produtores agrícolas adoptam técnicas de controlo e prevenção e, quando estas falham, recorrem à aplicação de **produtos fitofarmacêuticos**. No entanto, é sabido que a aplicação destes produtos químicos pode ter consequências prejudiciais para a saúde pública e para o ambiente, diminuindo não só a qualidade do produto como a confiança do consumidor. Assim, a maior aposta dos produtores está na prevenção e monitorização destas **pragas**, evitando perdas das culturas e de uma forma menos dispendiosa e mais segura do que a aplicação de produtos químicos.

A adoção de estratégias de prevenção e controlo das **pragas** apesar de ser mais vantajoso que o combate químico, apresenta mais desafios de utilização para que seja eficiente. Perante isto, surgiram diversos projetos de investigação com o objetivo de estudar esses inimigos e auxiliar os produtores agrícolas a aumentar a eficácia e eficiência do seu combate. Em Portugal surgiram também projetos de investigação neste contexto, como

o ProTomate, Safebrócolo e o FitoAgro. Esta dissertação integra-se no grupo operacional FitoAgro, cujo objetivo é definir uma estratégia de controlo regional para a produção de **pomóideas** na **região do Oeste**. A região tem vindo a ser afetada por diversos problemas de caráter **fitossanitário** e pelo aparecimento de novas **pragas**. Esta situação apresenta uma crescente importância económica para os produtores, pelo que tencionam mitigar os danos causados pelas **pragas** com o auxílio do projeto FitoAgro. Para isso, o grupo operacional pretende estudar o ciclo de vida das pragas ao efetuar a monitorização de pontos de observação biológica (**POBs**), i.e. pomares, e recolher dados meteorológicos e de observação de campo. Para que as entidades parceiras do projeto possam analisar os dados recolhidos, existe a necessidade de criar uma componente de visualização de dados que permita identificar fatores influentes no desenvolvimento das pragas e momentos de risco. A identificação destes irá, por sua vez, auxiliar as entidades a elaborar planos de risco **fitossanitário** e a delinear estratégias eficientes no controlo das **pragas**.

Deste modo, o objetivo principal desta dissertação é desenhar e implementar uma componente de visualização interativa de dados para o Projeto FitoAgro. No entanto, considera-se importante que esta contenha algum caráter genérico, de forma a que as soluções desenvolvidas possam ser reutilizadas ou adaptadas no futuro para outros projetos de monitorização fitossanitária, bem como para a visualização interativa de dados de três tipos: meteorológicos, de recolha *in situ* e modelos de desenvolvimento e/ou previsão.

1.2 Grupo Operacional FitoAgro

O FitoAgro é um projeto nacional cujo objetivo é estudar as **pragas** emergentes na **região do Oeste** que afetam as **pomóideas**, como as pereiras e macieiras. O seu grupo operacional visa compreender o ciclo de vida destes inimigos, validar modelos de desenvolvimento dos mesmos, construir mapas de risco, definir metodologias de estimativa de risco e indicadores de ataque. Assim, pretende-se com este projeto elaborar planos de risco **fitossanitário** e especificar estratégias de controlo de base biológica e biotecnológica, diminuindo a utilização de **produtos fitofarmacêuticos** químicos e, conseqüentemente, reduzindo a perigosidade toxicológica, ecotoxicológica e ambiental. [1]

Este projeto é financiado através do "PDR2020 - Programa de Desenvolvimento Rural" no âmbito do acordo Portugal 2020, sendo constituído por membros de diversas instituições e entidades, nomeadamente:

- **Instituições de Ensino Superior**
 - Escola Superior Agrária de Castelo Branco (ESACB)
 - Escola Superior Agrária de Santarém (ESAS)
 - Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa (FCT/UNL)
 - Instituto Superior de Agronomia (ISA)

• **Entidades Agrícolas e de Hortofrutícola**

- Associação dos Produtores Agrícolas da Sobrena (APAS)
- Cooperativa Agrícola do Bombarral, CRL (CAB)
- Cooperativa Agrícola dos Fruticultores do Cadaval, CRL (COOPVAL)
- Centro de Produção e Comercialização Hortofrutícola, Lda (CPF)
- Centro Operativo e Tecnológico Hortofrutícola Nacional (COTHN)
- Estação Fruteira da Estremadura, Lda (Ecofrutas)
- Cooperativa de Hortofruticultores, CRL (Frubaça)
- Cooperativa Agrícola de Hortofruticultores do Oeste, CRL (Frutoeste)
- Estação Fruteira do Montejunto, CRL (Frutus)
- Produtores de Frutas, CRL (Granfer)

As entidades produtoras desta parceria disponibilizaram treze **parcelas** de pera e maçã que servem de pontos de observação biológica (**POBs**) e nos quais foram instalados vários equipamentos de monitorização como **armadilhas cromotrópicas**, **armadilhas sexuais**, **cintas** para árvores, entre outros. A monitorização destes equipamentos juntamente com a observação dos técnicos no campo, permitirá o estudo e a validação dos ciclos de vida das **pragas**, de forma a que se possa delinear a estratégia biológica de controlo regional anteriormente referida e cumprir a obrigações impostas pela Lei nº26/2013 [2].

1.3 Descrição do Problema

O objetivo principal da visualização de dados é permitir a análise e compreensão dos dados recolhidos. No caso do projeto FitoAgro, requiere-se uma componente de visualização que seja capaz de atender aos requisitos analíticos elicitados pelos seus diferentes elementos, para que depois se possa delinear estratégias de controlo biológico para os diferentes tipos de **pragas** na **região do Oeste**.

De forma a responder a esses requisitos, a componente de visualização deverá procurar soluções apropriadas para a análise de dados meteorológicos e das observações de campo, bem como promover a descoberta de correlações entre estes dois tipos de dados. Para além disso, um dos maiores problemas surge na comparação entre dados de campo, uma vez que a utilização de eixos temporais implica a comparação do comportamento de pragas em condições de desenvolvimento diferentes. Finalmente, o último requisito prende-se na observação do ajuste de modelos preditivos adaptativos para pragas, sendo que deverá ser possível comparar diferentes tipos de ajustes de modelos e de parametrizações, de forma a aumentar a eficácia da sua previsão e, conseqüentemente, a confiança dos produtores nesses modelos.

Para além da sua aplicabilidade ao Projeto FitoAgro, a componente de visualização deverá ter algum carácter genérico para que se possa ser facilmente adaptada no futuro. Deste

modo, deverá ser capaz de contribuir com estruturas e regras gerais para a construção de visualizações interativas de dados para outros projetos de monitorização fitossanitária, bem como para áreas envolventes destes, a meteorologia, recolha *in situ* e modelos de desenvolvimento e/ou previsão de pragas.

1.4 Contributos

O objetivo desta dissertação, como referido anteriormente, é desenhar e implementar uma componente de visualização para o Projeto FitoAgro, sobre um sistema de informação já existente [3]. No entanto, para além disso, aspira-se que esta dissertação contribua para o desenho e modelação de visualizações interativa de dados no geral e apropriadas no âmbito da monitorização fitossanitária e áreas envolventes. Assim, espera-se que o trabalho desta dissertação tenha os seguintes contributos:

- Proposta de linguagem de modelação específica para visualização interativa de dados;
- Recomendações gerais de desenho e implementação de visualizações interativas de dados para projetos de monitorização fitossanitária;
- Componente de visualização fácil de utilizar, intuitiva, interativa e flexível que responde às necessidades analíticas do projeto FitoAgro;
- Soluções originais no âmbito da monitorização fitossanitária que promovam a descoberta de correlações entre diferentes variáveis (como a meteorologia) e o desenvolvimento das diferentes pragas;
- Generalização da componente de visualização como uma estrutura base para a construção de visualizações no âmbito da monitorização fitossanitária e dados envolventes (meteorologia, recolha *in situ* e modelos de desenvolvimento/preditivos), permitindo a incorporação de novos dados nesses contextos.

1.5 Estrutura do Documento

Neste capítulo (1) foi introduzido o problema das pragas na produção de pomóideas em Portugal e em especial na região do Oeste. Foi ainda descrito o grupo operacional FitoAgro, nomeadamente os seus objetivos, parceiros e abordagens. Finalmente, foi apresentado o problema que se pretende resolver e quais os objetivos e contributos desta dissertação face a esse problema.

O capítulo seguinte (2) descreve a realização de uma pesquisa alargada acerca dos trabalhos e áreas de conhecimentos no domínio do problema descrito. É efetuado assim o levantamento de projetos, aplicações, ferramentas e tecnologias relevantes e discutida a pertinência de cada uma delas para o trabalho desta dissertação.

No capítulo 3 é especificado o projeto FitoAgro, nomeadamente o tipo, formato e propósito dos dados recolhidos, os modelos preditivos e o seu funcionamento e, por fim, os *stakeholders* do projeto.

De seguida, no capítulo 4, são elicitados os requisitos analíticos do projeto e feito o seu mapeamento para os *stakeholders* apropriados.

No capítulo 5 é descrita abordagem e metodologia utilizada, explicando a importância da modelação e especificando algumas recomendações de desenho e o fluxo iterativo de avaliação utilizados para na construção de visualizações.

Por sua vez, os capítulos 6 e 7 apresentam o desenho e implementação da componente de visualização, sendo que no primeiro o foco incide sobre a meteorologia e no segundo incide sobre a fitossanidade, nomeadamente a análise dos dados de campo, bem como a descoberta de correlações entre estes e destes com os dados meteorológicos. Note-se que os *dashboards* e visualizações apresentadas nestes capítulos estão limitadas no documento pela formatação imposta no tamanho das figuras. No entanto, quando visualizadas na plataforma apropriada têm um tamanho adequado.

No penúltimo capítulo (8) apresenta-se a abordagem utilizada para a avaliação da componente e uma discussão dos resultados obtidos.

Finalmente, o último capítulo (9) conclui o trabalho elaborado ao longo desta dissertação e enumera propostas de trabalho futuro.

TRABALHO RELACIONADO

Neste capítulo são resumidos projetos e aplicações que possam estar relacionados com o objetivo desta dissertação (2.1). Além disso, são apresentados conceitos relativamente à área de conhecimento de visualização interativa de dados (2.2), nomeadamente os fundamentos da área (2.2.1) e as principais técnicas (2.2.2) e ferramentas (2.2.3) relevantes.

2.1 Projetos e Aplicações Relacionadas

O objetivo desta primeira secção é fazer o levantamento de projetos e aplicações relevantes para a camada de visualização a desenvolver no âmbito do projeto FitoAgro. Uma vez que existe uma grande variedade de projetos e aplicações no domínio agrícola cujo objetivo é auxiliar os agricultores ao minimizar os danos e maximizar a sua produção, foram apenas selecionados os que melhor abrangem os objetivos e requisitos analíticos impostos pelo grupo operacional FitoAgro.

2.1.1 Cropio

O Cropio [4] é um sistema de gestão agrícola por satélite que facilita a monitorização remota de parcelas, permitindo que os produtores planeiem as suas operações de uma forma mais eficiente. Oferece informação relativamente ao estado da **parcela** (e.g. cultura, variedade e data da colheita) e visualizações do nível de vegetação, da temperatura diária no ar e solo, da precipitação diária e acumulada, da humidade do solo em diferentes profundidades, entre outros (figura 2.1). Adicionalmente, disponibiliza imagens de satélite em tempo real da **parcela**, do relevo, da vegetação, entre outros. Além disso, gera alertas relativamente à vegetação e a eventos ou previsões meteorológicas, bem como relatórios semanais com um sumário dos dados relevantes e da condição e produtividade das **parcelas** monitorizadas.

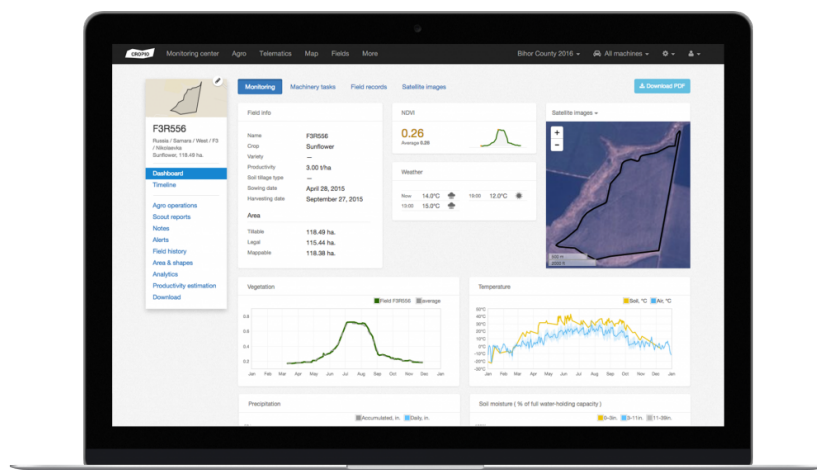


Figura 2.1: Exemplo da vista geral do sistema Cropio. [4]

Este sistema está disponível mundialmente, sendo que não requer a instalação de nenhum equipamento. No entanto, o seu acesso é pago, sendo que o preço varia consoante o número de hectares da *parcela*.

2.1.2 FieldClimate

O FieldClimate [5] é uma plataforma *web*, criada em 2005, que coleta e disponibiliza visualizações de dados agro meteorológicos recolhidos em diversos sensores e estações instaladas por todo o mundo. Oferece diferentes visualizações gráficas e tabulares das diversas medições efetuadas e previstas pelas estações meteorológicas, nomeadamente, a temperatura do ar, humidade relativa no ar, velocidade do vento, folha molhada, entre outros. Além disso, disponibiliza também visualizações para as medições dos sensores de solo como a precipitação e a humidade no solo. Para qualquer uma destas visualizações, é possível agrupar os dados de hora em hora, diariamente ou mensalmente.



Figura 2.2: Exemplo de interface da plataforma FieldClimate. [5]

Para além dos dados meteorológicos, esta plataforma também fornece serviços de monitorização de captura de insetos através da instalação de armadilhas integradas com componentes eletrónicos e com uma câmara de alta resolução para monitorizar remotamente através da *web*. Finalmente, esta plataforma oferece diversos modelos de desenvolvimento de doenças e de algumas pragas em mais de 40 culturas. Todos os serviços oferecidos por esta plataforma bem com a instalação/utilização de sensores meteorológicos é pago, sendo que o custo associado varia entre produtos.

2.1.3 ProTomate

O projeto ProTomate surgiu para dar resposta ao aparecimento da *Tuta absoluta* (traça do tomateiro), que em 2011 causou perdas elevadíssimas nas searas de tomate, com situações de perda total. O objetivo do projeto era o «desenvolvimento de um sistema de apoio à decisão com base no acompanhamento técnico dos inimigos da cultura e nos mapas de risco de ocorrência, nas principais regiões produtoras, contribuindo para o uso sustentável dos produtos fitofarmacêuticos» [6]. Para isso, foi desenvolvida uma plataforma *web* que utiliza dados de satélite para calcular os graus-dia acumulados para a *Tuta absoluta* e produzir uma carta para o mínimo de gerações esperadas desde o dia um de Janeiro de cada ano (figura 2.3). [7]

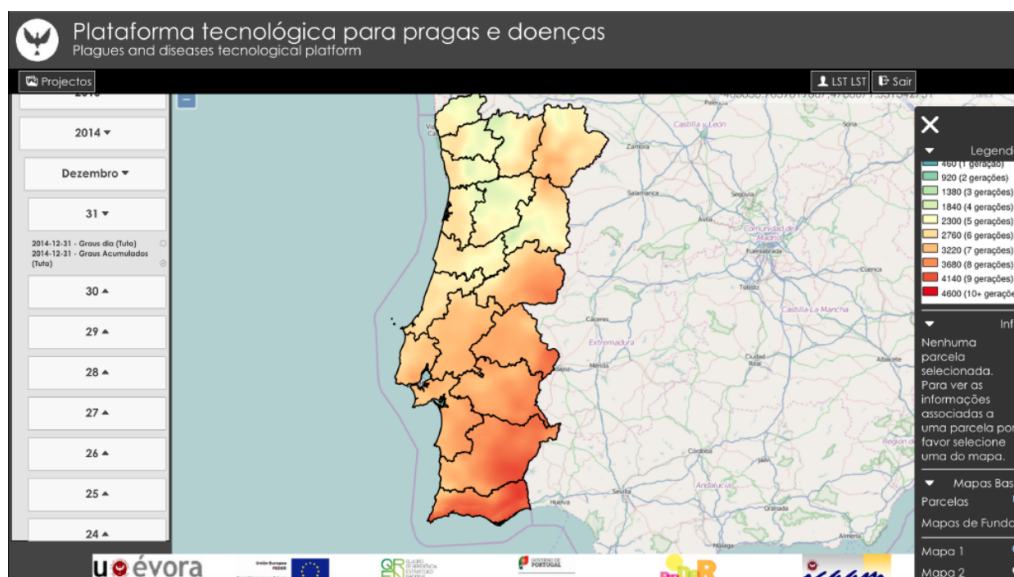


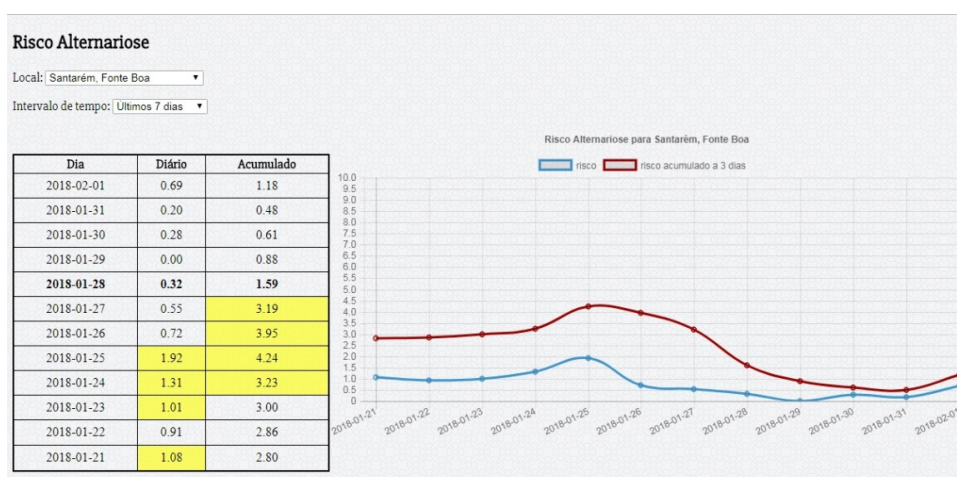
Figura 2.3: Mapa de risco da plataforma *web* do projeto ProTomate. [7]

2.1.4 SafeBrócolo

“SafeBrócolo – Melhoria do Processo Produtivo com base em modelos de risco para *alternaria* e mosca da couve” é um projeto maioritariamente português cujo objetivo é o desenvolvimento de modelos de previsão de ocorrência de *alternariose* e ataques de mosca

da couve [8]. Ao longo deste projeto foram desenvolvidas uma plataforma *web* e uma aplicação móvel.

Por um lado, a aplicação móvel permite aos utilizadores registarem as suas parcelas e registem as observações efetuadas no campo. Além disso, disponibiliza e mapas de risco de Portugal Continental e indicadores de risco para cada uma das *parcelas* do utilizador, com alertas personalizáveis para o mesmo. Por outro lado, a plataforma *web* tem por objetivo apresentar os dados georreferenciados recolhidos pela aplicação móvel bem como permitir a adição e complementação de informação registada. Permite também o registo de *parcelas* e observações, mas, ao contrário da aplicação móvel, é possível definir cabeceiras e bordaduras e fazer divisões de áreas de plantação dentro de uma *parcela*. Esta plataforma apresenta não só visualizações para os dados georreferenciados, mas também para estimativas de risco por *parcela* ou estação (2.4a) e mapas de risco (2.4b).



(a) Visualização do risco de *Alternariose* por estação.

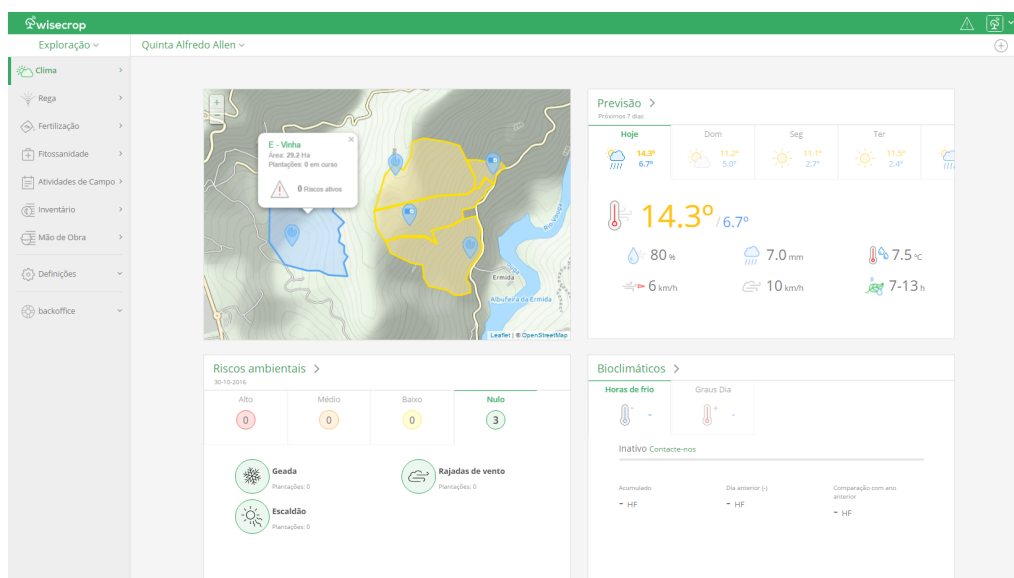


(b) Mapa de risco de *Alternariose* para Portugal Continental.

Figura 2.4: Exemplo de páginas da plataforma web do projeto SafeBrócolo. [9]

2.1.5 WiseCrop

A WiseCrop [10] é uma *startup* portuguesa que disponibiliza diversos sistemas de auxílio ao agricultor. De entre estes sistemas destacam-se as 7 aplicações que oferecem diferentes funcionalidades ao produtor: clima, rega, fertilização, atividades de campo, fitossanidade, custos e mão de obra. Estas podem ser utilizadas em conjunto ou em módulos separados, sendo que estão integradas de forma a acomodar as diferentes necessidades dos produtores. No âmbito desta dissertação, destacam-se os módulos clima, atividades de campo e fitossanidade, apresentados nas figuras 2.5a, 2.5b e 2.5c, respetivamente.



(a) Módulo para clima



(b) Módulo atividades de campo



(c) Módulo para fitossanidade

Figura 2.5: Exemplo de alguns módulos relevantes do sistema da WiseCrop. [10]

O módulo do clima permite analisar dados meteorológicos quase em tempo real (de 15 em 15 minutos), através das medições efetuadas por diversos sensores colocados nas **parcelas** de monitorização, enquanto que o módulo atividades de campo permite registar e visualizar vários tipos de observações e eventos que ocorreram nesses locais.

Finalmente, o módulo fitossanitário, este utiliza modelos de previsão avançados, juntamente com dados recolhidos no campo, para estimar o risco relativo da cultura a **pragas** ou doenças. Caso os valores registados excedam o limite de segurança estabelecido, o sistema lança alertas ao produtor, sendo estes alertas personalizáveis. Desta forma, pretendem aumentar a eficácia dos tratamentos, diminuindo a probabilidade de se desenvolverem resistências. A WiseCrop acredita que o investimento neste módulo resulta numa redução de pelo menos 20% dos custos de tratamentos **fitossanitários**. Todo o sistema da WiseCrop é pago por módulos consoante as necessidades de cada utilizador.

2.1.6 Conclusão

Os projetos e plataformas apresentadas anteriormente contém componentes de visualização por norma focadas numa temática, sendo que nenhuma por si só parece abranger todas as vertentes analíticas dos requisitos do grupo operacional FitoAgro nem a diversidade de **pragas** em estudo.

Os sistemas Cropio e FieldClimate possibilitam a visualização de dados meteorológicos relevantes para a gestão agrícola, através de satélite e estações meteorológicas, respetivamente. Por outro lado, não permitem a observação de dados recolhidos *in situ* e apenas a FieldClimate é que fornece serviços de monitorização de capturas em armadilhas e de análise de modelos de desenvolvimento de doenças e **pragas**. No entanto, esses modelos não se ajustam consoante as capturas observadas, para além de que, sendo na sua maioria modelos para doenças, não abrangem a diversidade de **pragas** do FitoAgro.

Por outro lado, os projetos ProTomate e Safebrócolo incorporaram recolha de dados de campo, mas estes apenas serviram para o desenvolvimento de modelos preditivos, tendo sido também este o foco das componentes de visualização desses projetos, nomeadamente através da criação de mapas e estimativas de risco.

Finalmente, os sistemas fornecidos pela WiseCrop são aqueles que apresentam ser mais abrangentes e completos, uma vez que permitem a monitorização e analítica de dados meteorológicos, a recolha e visualização de dados de campo e, por fim, a observação de estimativas de risco através de modelos preditivos de **pragas**. No entanto, este sistema está dividido em módulos com diferentes funcionalidades, sendo que para abordar os requisitos analíticos do projeto FitoAgro seria preciso um conjunto de pelo menos três módulos: clima, fitossanidade e actividades de campo. Esse conjunto iria permitir a análise das três componentes modularmente, o que poderá não promover a descoberta de correlações entre os dados e a confiança dos utilizadores nos resultados e estimativas de risco obtidas. Ademais, não se sabe se este sistemas e respetivos modelos abrangeriam a diversidade de **pragas** estudadas pelo grupo operacional.

Deste modo, conclui-se que as soluções existentes não permitem satisfazer todas as vertentes analíticas do projeto FitoAgro. No entanto, é importante realçar que a semelhança entre soluções, de diferentes sistemas, contribuiu para a generalização de uma abordagem para visualizações nesses contextos. Por exemplo, notou-se que é comum na visualização de dados meteorológicos utilizar barras para valores acumulados e linhas para valores de instante, bem como permitir que os utilizadores visualizem os dados em diferentes agregações e intervalos de tempo. Essas conclusões em conjunto com críticas construtivas às visualizações desses sistemas, contribuiu para a construção dos diferentes elementos da componente de visualização interativa de dados proposta.

2.2 Visualização Interativa de Dados

Uma vez que a área de conhecimento em foco nesta dissertação é a visualização interativa de dados, foi efetuada um estudo acerca dos fundamentos, técnicas e aplicações da mesma. O livro de referência para o estudo destes conceitos foi *Interactive Data Visualization: Foundations, Techniques, and Applications, 2nd Edition* (Ward et al., 2015) [11].

2.2.1 Fundamentos

Segundo John C. Hart [12], a visualização de dados não serve apenas para ver os dados, mas sim para os compreender e ser capaz de tomar decisões conforme os mesmos. Fundamentalmente, existem dois modos de visualização: interativa e apresentação. Por um lado, a visualização de apresentação é utilizada para comunicar com grupos grandes ou audiências, sem interação/contribuição desta. Por outro lado, a interativa é utilizada para explorar e descobrir correlações nos dados e tem como alvo audiências mais pequenas ou até mesmo um único investigador/colaborador que pode interagir com esta dinamicamente. Para além disso, existe um tipo híbrido, a narrativa interativa, que permite comunicar uma apresentação, mas de forma interativa com o utilizador.

A visualização de dados é importante quando existe um público alvo que necessita de compreender os dados. A informação tem de ser comunicada de uma forma clara, eficiente e eficaz. Para isso, os dados são representados no ecrã através de um mapeamento para variáveis visuais. O utilizador poderá então interagir com essas visualizações de forma a satisfazer os seus requisitos, perguntas e/ou tarefas, contribuindo para a sua compreensão dos dados iniciais. Existem dois principais tipos de dados: nominal e ordinal. Os dados ordinais tomam valores numéricos que podem ser binários, discretos ou contínuos. Contrariamente, os nominais tomam valores não numéricos que podem ser categóricos (lista finita de valores), classificativos (variável categórica com ordem implícita) ou arbitrários. Para além disso, os dados podem ter diferentes tipos de estruturas, nomeadamente escalares (valor individual), vetoriais (múltiplos valores que representam uma única variável, e.g. componentes RGB), geométricas, temporais, entre outras.

Tal como referido anteriormente, um dos principais passos do processo de visualização é o mapeamento dos dados para variáveis visuais. Existem oito variáveis visuais diferentes, mas apenas as duas primeiras são realmente necessárias para definir uma visualização. As restantes seis variáveis apenas afetam a forma como as representações individuais das duas primeiras são expostas. Seguidamente, apresenta-se as oito variáveis visuais e uma breve descrição acerca de cada uma destas.

- **Posição**

Refere-se ao posicionamento de visualizações dentro de um ecrã, quer seja em uma, duas ou três dimensões. Distribuir as visualizações de forma a maximizar a utilização do espaço disponível no ecrã, maximiza também a quantidade de informação comunicada. Para além disso, a representação de visualizações utilizando diferentes escalas pode ter influência na descoberta de relações entre diferentes variáveis, estejam estas verdadeiramente presentes nos dados ou não. Por este motivo, é importante averiguar, nos dados iniciais, a informação transmitida nas visualizações.

- **Forma**

A forma ou marca representa a utilização de pontos, linhas, áreas, volumes ou composições destas para representar graficamente dados. Quando se utilizam várias formas em simultâneo é importante que estas se distingam facilmente entre si.



Figura 2.6: Exemplo de formas disponíveis na plataforma Tableau [13]

- **Tamanho**

O tamanho mapeia-se facilmente para variáveis contínuas ou com intervalos, porque permitem incrementações graduais ao longo de uma gama de valores. Uma vez que não é trivial distinguir formas de tamanhos semelhantes, esta variável visual apenas suporta categorias de pequena cardinalidade. Para além disso, um dos maiores problemas da sua utilização é o tipo de forma utilizada. O tamanho é eficaz quando utilizado com pontos, linhas e curvas. No entanto, quando estas são representadas por gráficos que contêm uma área suficientemente grande, os aspetos quantitativos do tamanho perdem-se e as diferenças entre as formas tornam-se qualitativas.

- **Brilho/Luminância**

O brilho pode ser utilizado para proporcionar a diferença relativa entre grandes intervalos ou variáveis de dados contínuas. Além disso, pode ainda ser utilizado para a distinção de marcas, através de uma escala de valores de brilho reduzida. Embora seja possível utilizar o intervalo completo de valores de brilho, os humanos não são capazes de distinguir entre todos esses valores.

- **Cor**

A cor é normalmente selecionada consoante a natureza dos dados: sequencial, divergente ou qualitativa. Os mapas de cor (sequenciais ou divergentes) são úteis para intervalos e variáveis de dados contínuas, uma vez que são geralmente definidos por um intervalo contínuo de valores de matiz e saturação. Quando os dados são categóricos ou intervalos com muito pouca cardinalidade, é geralmente aceitável selecionar manualmente as cores para valores individuais de forma a otimizar a distinção entre os diferentes tipos de dados (natureza qualitativa). Sempre que possível, essa seleção deverá ser intuitiva com base na percepção da audiência e de associações culturais (e.g. água é azul e vegetação é verde).

- **Orientação**

A orientação é o principal componente gráfico dos desenhos iconográficos e está diretamente ligada à visão pré-atenta, isto é, informação visual que é detetada subconscientemente pelos humanos, de forma muito rápida e precisa. As formas mais adequadas para utilizar a orientação são aquelas que possuem um único eixo.



Figura 2.7: Exemplos de orientações de uma forma [14]

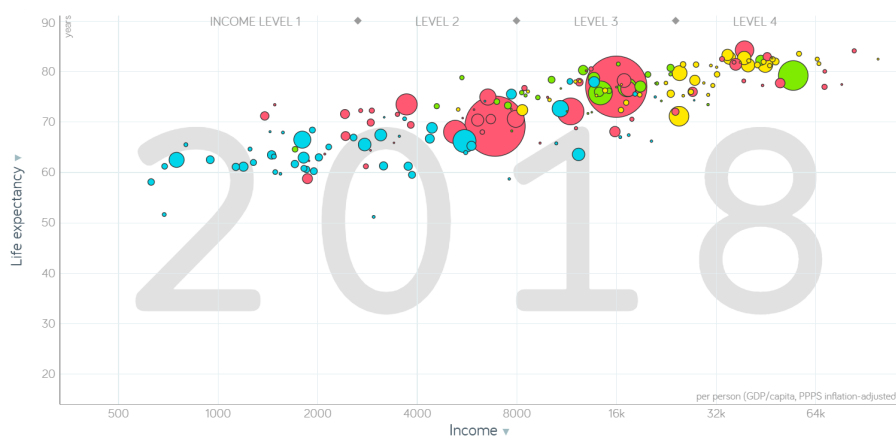
- **Textura**

A textura pode ser considerada uma combinação de outras variáveis visuais como a forma, cor e orientação. É geralmente associada a um polígono, região ou superfície.

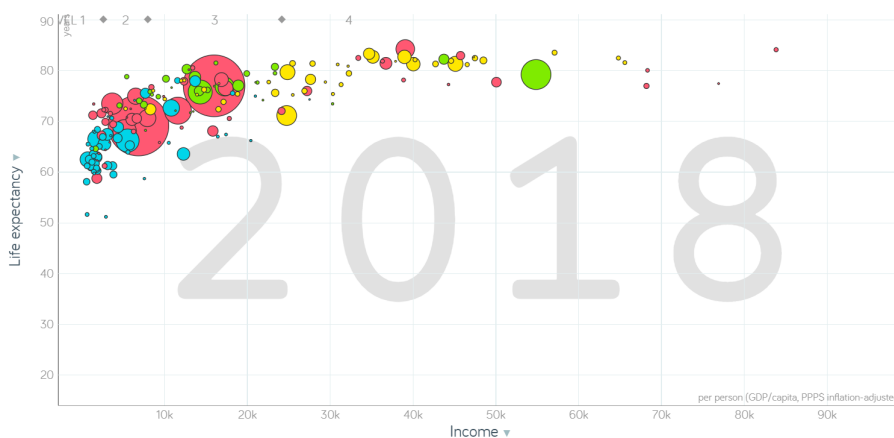
- **Movimento**

O movimento pode ser associado a qualquer uma das outras variáveis visuais uma vez que o modo como uma variável se altera ao longo do tempo poderá transmitir informação adicional. Uma utilização comum do movimento ocorre com a variação da velocidade, onde se pode observar alterações da posição dos valores ao longo do tempo e a direção dessa alteração.

Na figura 2.8 é possível observar um gráfico de dispersão que utiliza cinco variáveis visuais em simultâneo: posição, forma, tamanho, cor e movimento. A posição é representada pela utilização de dois eixos, onde é possível observar o efeito da utilização das diferentes escalas - linear (figura 2.8a) e logarítmica (figura 2.8b) - na percepção dos dados. Para além disso, o tamanho representa a população de cada país, a cor é utilizada qualitativamente para distinguir os continentes e a forma utilizada é o ponto, que é uma marca eficaz para a utilização conjunta com o tamanho. Finalmente, apesar de não ser observável na figura, em [15] é possível visualizar a movimentação das variáveis ao longo do tempo, desde 1800 até 2018.



(a) Escala logarítmica no eixo horizontal



(b) Escala linear no eixo horizontal

Figura 2.8: Gráfico de dispersão que representa a esperança média de vida em 2018 por rendimento do país, onde o tamanho representa a população do país e a cor o continente da mesma (Verde - América, Azul - África, Amarelo - Europa, Vermelho - Ásia) [15].

A percepção é o processo em que os humanos interpretam o mundo à sua volta, formando uma representação mental do ambiente. De forma a gerir a ambiguidade inerente na informação sensorial recebida, o cérebro assume factos sobre a mesma. Quando visualizamos dados, é necessário garantir que essas interferências não estejam presentes e não impeçam a compreensão do que está a ser transmitido nas visualizações. Um exemplo de interferência é o fenómeno perceptual chamado *change blindness*. Este fenómeno ocorre quando há interrupções (i.e. um movimento ocular, mudança de ecrã, etc) naquilo que está a ser observado, tornando-nos insensíveis às alterações que ocorreram durante a interrupção. Na visualização de dados, existem situações que podem levar a esse fenómeno como, por exemplo, quando o utilizador necessita de observar duas visualizações semelhantes em locais distintos e compará-las entre si.

Segundo Tamara Munzner [16], para maximizar a expressividade e percepção da informação visual, os canais utilizados devem corresponder às características dos dados. Dados com atributos ordenados devem utilizar canais de magnitude, enquanto que atributos categóricos deverão utilizar canais de identidade. Na figura 2.9 é possível observar

esses canais ordenados por eficácia, sendo que os atributos mais importantes deverão ser mapeados para os canais mais eficazes.

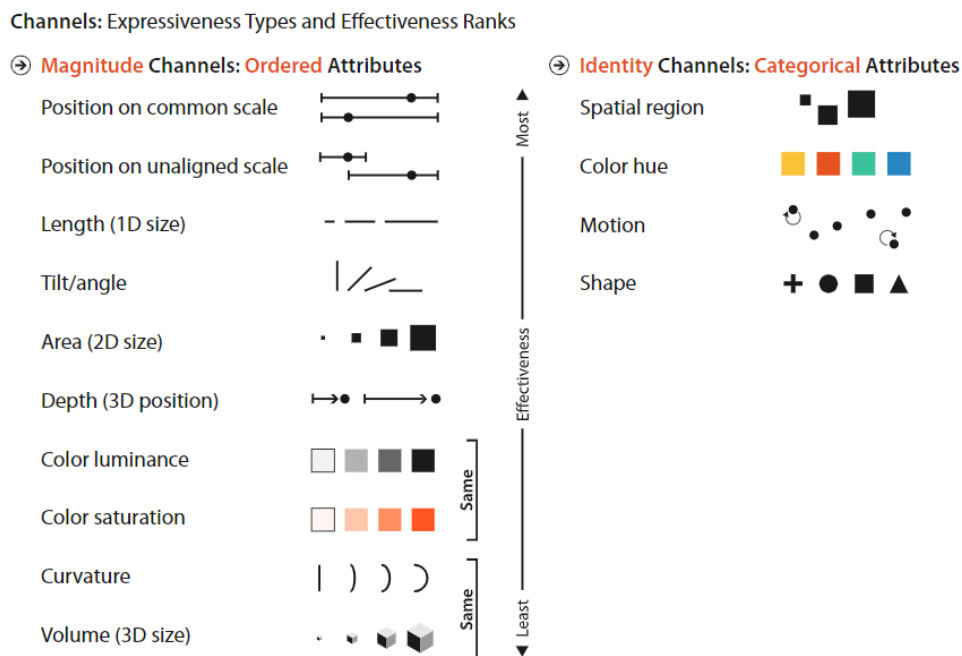


Figura 2.9: Canais ordenados por eficácia de acordo com o tipo de dados e de canal [16]

2.2.2 Técnicas

No caso desta dissertação e do projeto FitoAgro, os dados, descritos em 1.3, são variáveis multivariadas com foco na natureza temporal e geoespacial das mesmas. Deste modo, foi necessário realizar um estudo acerca das principais técnicas que se adequam a esse tipo de dados, apresentado de seguida.

2.2.2.1 Dados Multivariados

Os dados multivariados são aqueles que, geralmente, não têm atributos espaciais explícitos. Para este tipo de dados utilizam-se técnicas baseadas em pontos, em linhas e regiões. De seguida apresenta-se uma descrição mais detalhada acerca de cada um desses tipos de técnicas.

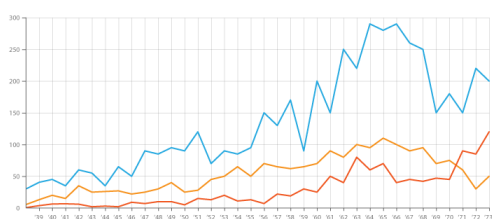
- **Técnicas baseadas em pontos**

As técnicas baseadas em pontos projetam os registos de dados de um espaço n-dimensional para um espaço no ecrã tal que estes se mapeiem para pontos k-dimensionais. Quando a dimensão dos dados excede a capacidade da técnica de visualização, torna-se necessário subdividir ou reduzir a dimensão, mapear para outras variáveis (e.g. cor e tamanho) ou utilizar vários ecrãs de visualização (e.g. matriz de dispersão). Os gráficos e matrizes de dispersão são exemplos da utilização das técnicas clássicas baseadas em pontos. O sucesso da sua utilização resulta da

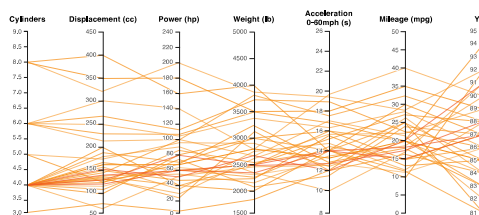
capacidade inata dos humanos de julgar posições relativas dentro de um espaço delimitado.

- **Técnicas baseadas em linhas**

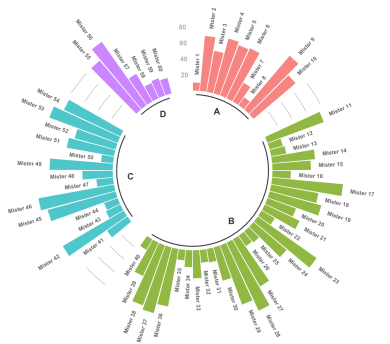
Nas técnicas baseadas em linhas, os pontos que correspondem a um registo de dados ou dimensão específica são ligados por linhas retas ou curvas. Os gráficos de linhas (2.10a) e de coordenadas paralelas (2.10b) são exemplos clássicos da utilização destas técnicas. Outras técnicas baseadas em linhas são as de eixo radial como gráficos circulares de linhas, de barras (2.10c) e de áreas (2.10d), gráficos polares (2.10e) e gráficos radiais de colunas (2.10f).



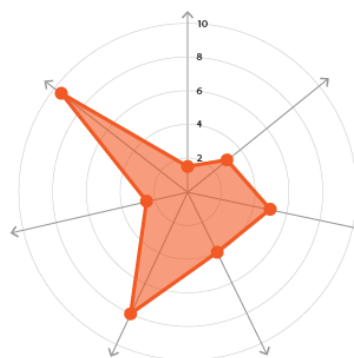
(a) Gráfico de linhas



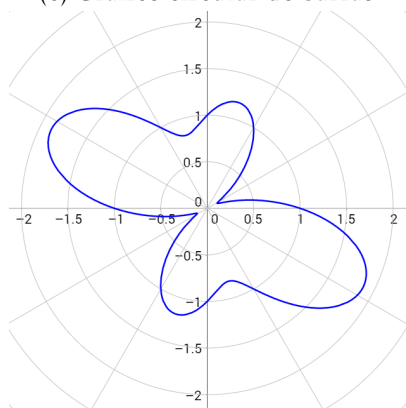
(b) Gráfico de coordenadas paralelas.



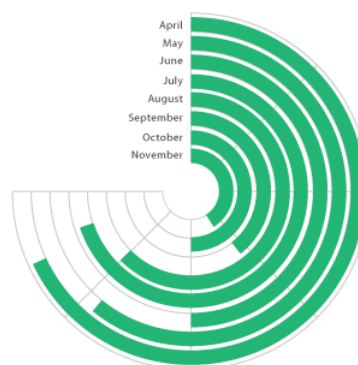
(c) Gráfico circular de barras



(d) Gráfico circular de área



(e) Gráfico circular de linhas com coordenadas polares (gráfico polar)



(f) Gráfico radial de colunas

Figura 2.10: Exemplos de gráficos que utilizam técnicas baseadas em linhas [17] [18] [19]

- **Técnicas baseadas em regiões**

As técnicas baseadas em regiões utilizam polígonos preenchidos para transmitir valores conforme a sua cor, tamanho, forma ou outros atributos. Os exemplos clássicos da utilização destas técnicas são os gráficos de barras (2.11a e 2.11b), histogramas (2.11c) e mapas de calor (2.11d).

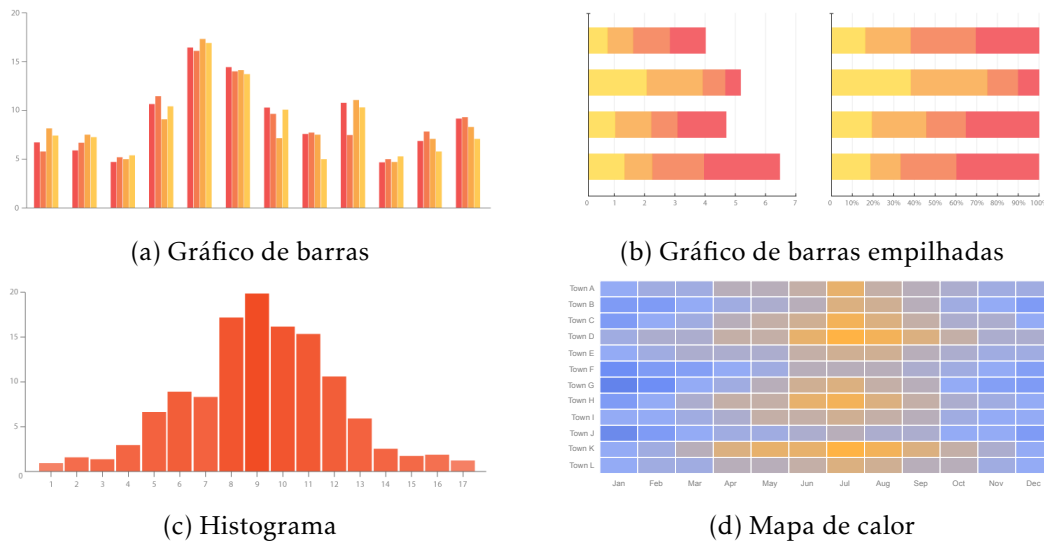


Figura 2.11: Exemplos de gráficos que utilizam técnicas baseadas em regiões [17]

2.2.2.2 Dados Temporais

Os elementos temporais podem ser abstraídos de diversas formas. A granularidade e o calendário são abstrações humanas do tempo com o intuito de facilitar a execução de tarefas. Descrevem o mapeamento de valores temporais para unidades conceituais maiores ou menores (e.g. minutos, horas, dias). Contrariamente, as primitivas do tempo podem ser vistas como uma camada intermédia entre os elementos dos dados e o domínio temporal. Estas primitivas podem ser absolutas ou relativas, onde as absolutas correspondem a instantes (i.e. pontos únicos no tempo) ou a intervalos (i.e. dois pontos no tempo) localizados numa posição fixa do domínio temporal e as relativas a durações no tempo sem posição absoluta. Finalmente, os elementos temporais podem ser abstraídos por determinação, sejam eles determinados, indeterminados ou incertos.

Assim, o tempo e os dados temporais têm características distintas que fazem com que seja relevante considerá-los como um tipo de dados distinto. Uma dessas características é a possibilidade de os representar linearmente ou ciclicamente, salientando diferentes aspetos dos dados. Por exemplo, no contexto agrícola, os dados podem ser representado linearmente ao longo do tempo ou segundo a ciclicidade do dia que influencia o desenvolvimento das culturas e das pragas. Além disso, podem ser representados pelo ciclo das estações do ano ou pelo ciclo cultural (e.g. campanhas). Observar este tipo de características específicas do tempo nos dados pode aumentar significativamente a expressividade

das representações visuais.

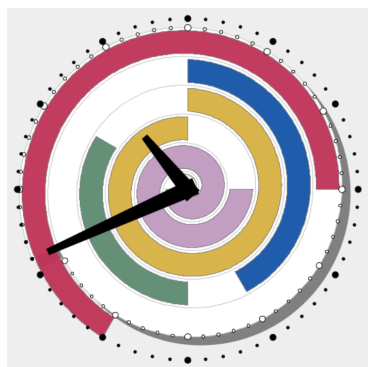
Deste modo, a escala, que pode ser qualitativa (ordinal ou nominal) ou quantitativa (discreta ou contínua), é uma das principais características deste tipo de dados. Escalas ordinais refletem ordens relativas ao presente (i.e. antes, durante, depois, etc.) enquanto que as nominais denotam conjuntos de elementos desordenados. Nas escalas contínuas os valores são mapeados para números reais enquanto que na escala discreta são mapeados para um conjunto de inteiros. A escala discreta é a mais utilizada e baseia-se na unidade mais pequena possível (e.g segundos, minutos), sendo que distâncias temporais também podem ser consideradas. Ademais, outras características deste tipo de dados é poderem ser uni ou multivariados e com contexto espacial ou abstrato. Finalmente, estes dados podem ser eventos ou estados, onde os eventos marcam alterações de estado e os estados representam as fases de continuidade entre eventos.

scale	3.14 3.27 4.88 quantitative	coconut banana apple qualitative
frame of reference	▼ abstract	🌐 spatial
kind of data	⌈ ⌋ events	— — states
number of variables	📈 univariate	📈 multivariate

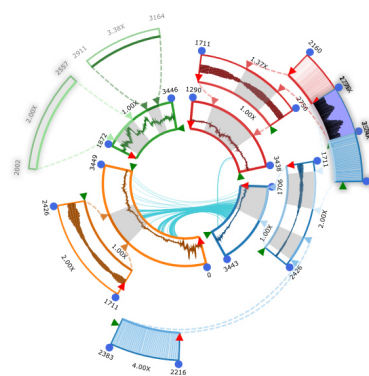
Figura 2.12: Algumas características de dados temporais [20]

Para visualizar dados temporais, o tempo deverá ser mapeado para o tempo (dinâmico) ou para o espaço (estático). Por um lado, o mapeamento temporal utiliza o tempo da audiência para representar a evolução dos dados através de uma sequência de imagens (filme), cujos valores alteram de posição ao longo do tempo. Este mapeamento não necessita de outras variáveis visuais além da posição, forma e movimento. No entanto, como referido em 2.2.1, a utilização do movimento não é eficaz quando os dados não sofrem alterações significativas de posição ao longo do tempo. Por outro lado, o mapeamento para o espaço refere-se à representação do tempo através de eixos e posições na visualização. Este é o mapeamento mais comum e suporta o mapeamento simultâneo de outras variáveis temporais para variáveis visuais não espaciais (e.g. cor e brilho). Para além disso, neste mapeamento, a dimensão do espaço de apresentação poderá ser em 2D, onde o eixo do tempo tem de ser enfatizado, ou em 3D, com uma dimensão dedicada para o mesmo.

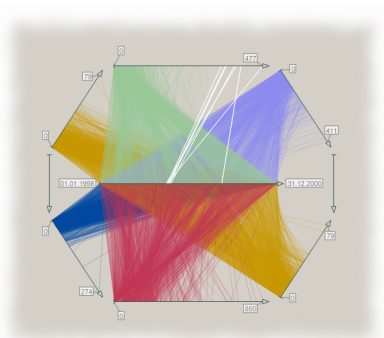
Assim, tendo em consideração todas as características dos dados temporais e as técnicas de mapeamento dos mesmos, apresentam-se na figura 2.13 exemplos da utilização de técnicas de visualização para dados temporais multivariados.



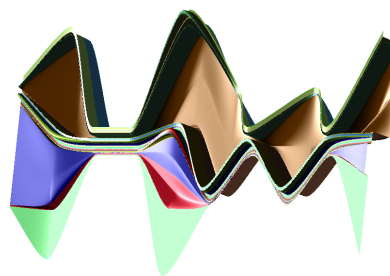
(a) SpiraClock - 2D, cíclico, intervalo, dinâmico



(b) Kronominer - 2D, cíclico, instante, estático



(c) TimeWheel - 2D, linear, instante, estático



(d) ThemeRiver - 3D, linear, instante, estático

Figura 2.13: Exemplos de técnicas de visualização para dados temporais multivariados [21]

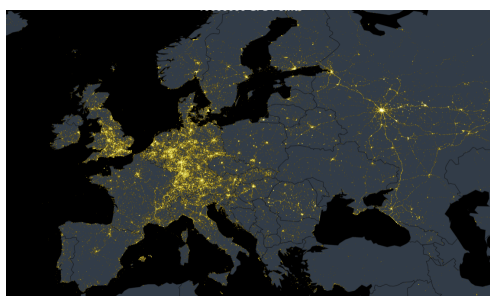
2.2.2.3 Dados Geoespaciais

Os dados geoespaciais descrevem objetos ou fenômenos com uma localização específica no mundo real, sendo frequentemente amostras discretas de um fenômeno contínuo. Estes fenômenos podem ocorrer em pontos, linhas, áreas ou superfícies. Devido a estas características, a estratégia de visualização base para estes dados é o mapeamento dos atributos espaciais diretamente para duas dimensões físicas do ecrã, resultando nos mapas. Estes representam o mundo através de pontos, linhas e áreas, sendo que os parâmetros de visualização (i.e. tamanho, cor, forma, etc.) apenas representam informação adicional acerca do objeto ou fenómeno em análise. Os mapas podem ser divididos em tipos consoante as propriedades dos dados (i.e. qualitativo ou quantitativo; discreto ou contínuo) e das variáveis gráficas (i.e. pontos, linhas, superfícies e volumes). As variáveis visuais para dados espaciais são o tamanho, forma, brilho, cor, orientação, espaço (textura), altura da perspetiva (3D) e disposição dos padrões. A utilização destas variáveis de forma eficaz varia consoante o tipo de fenómenos/dados em análise, como descrito de seguida.

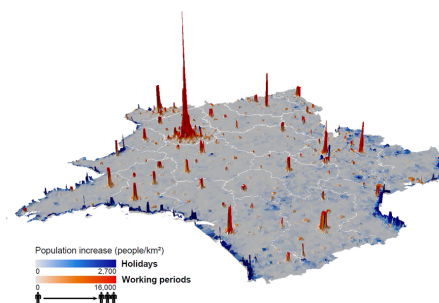
- **Pontos**

Os dados espaciais descritos por pontos são naturalmente discretos, mas podem

descrever um fenómeno contínuo. Dados discretos ocorrem, presumidamente, em locais distintos, enquanto que dados contínuos são definidos em todos os locais. Estes fenómenos podem ser visualizados ao colocar um símbolo ou pixel no local onde este ocorre. Parâmetros quantitativos podem ser mapeados para o tamanho ou cor do símbolo/pixel, sendo que os símbolos mais utilizados são os círculos, mas outras formas podem ser igualmente utilizadas. Quando são desenhados conjuntos de dados com grandes dimensões em mapas de pontos surgem alguns problemas. Nas áreas mais populadas há sobreposição de pontos enquanto que as áreas pouco populadas estão praticamente vazias (figura 2.14a). Para além disso, a resolução do ecrã poderá ter influência na quantidade de pontos observados. Uma das técnicas para lidar com estes problemas é a utilização de visualização 2.5D, isto é, agregar os pontos em regiões no mapa (figura 2.14b).



(a) Mapa de pontos tradicional [22]



(b) Mapa de pontos agrupados por barras verticais [23]

Figura 2.14: Exemplos de mapas de pontos

- **Linhas**

Fenómenos lineares são representados por segmentos de reta entre dois locais. Um mapeamento *standard* de uma linha permite que os parâmetros dos dados sejam mapeados para a espessura, padrão, cor e etiqueta desta. Adicionalmente, os pontos de início, interseção e fim da linha podem também ser mapeados para parâmetros visuais como o tamanho, forma, cor e etiqueta.

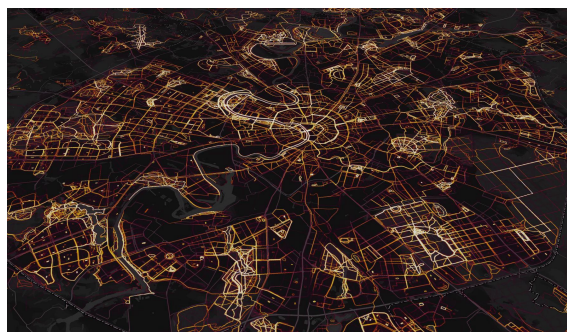


Figura 2.15: Exemplo de um mapa de linhas [24]

- **Áreas**

Fenómenos regionais são representados por mapas de áreas como mapas coropléticos, isarítmicos e cartogramas. Os mapas coropléticos mapeiam valores de um atributo para regiões coloridas ou sombreadas no mapa (figura 2.16a). O problema deste tipo de mapas é que assumem que o atributo está distribuído uniformemente dentro de cada região, causando perda de informação. Por outro lado, os cartogramas são generalizações que evitam esse problema ao distorcer a geografia para refletir o valor do atributo (figura 2.16c). Os cartogramas não contínuos são os mais utilizados uma vez que os polígonos escalados ao valor do atributo são desenhados dentro das regiões originais, não causando problemas de percepção (figura 2.16d). Finalmente, os mapas isarítmicos mostram os contornos de um fenômeno contínuo (figura 2.16b), sendo que estes podem ser determinados por valores específicos de um ponto (isométricos) ou medidos numa região (isopléticos).

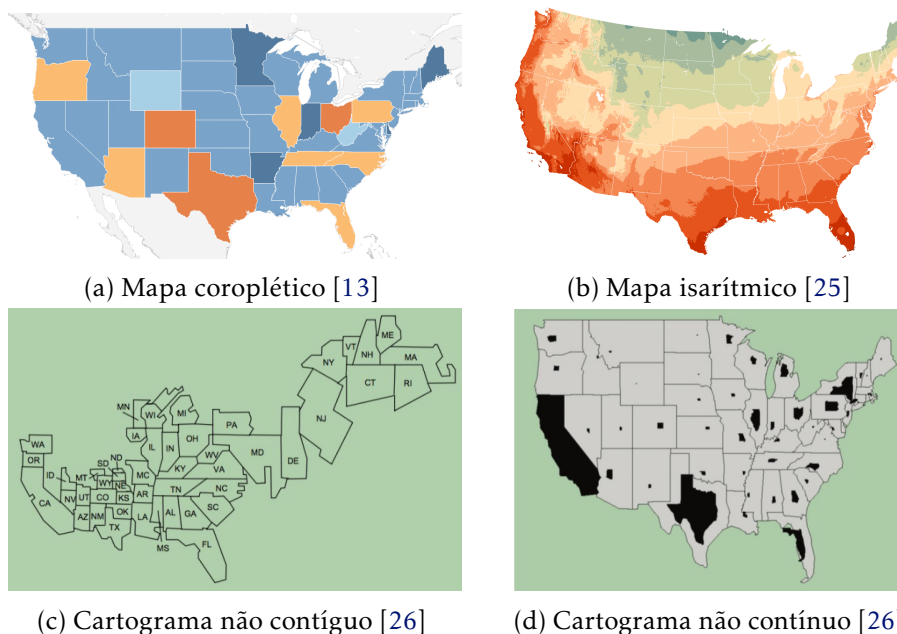


Figura 2.16: Exemplos de mapas de áreas

2.2.3 Ferramentas

Uma vez que existem diversas ferramentas de visualização de dados disponíveis atualmente, foi feita uma seleção tendo por base os requisitos do trabalho desta dissertação. Os critérios de seleção utilizados foram a necessidade de uma ferramenta analítica para a construção de *dashboards* e visualizações interativas de dados temporais e georreferenciados facilmente customizáveis e que cuja configuração e processamento de dados seja simples ou embutida, promovendo a exploração analítica dos mesmos. Note-se que a base de dados utilizada pelo sistema de informação do FitoAgro é o *PostgreSQL*, mas a ferramenta deverá ser capaz de integrar outros tipos de fontes de dados, principalmente ficheiros *Excel* ou similares.

2.2.3.1 Looker

O *Looker* [27] é uma ferramenta de inteligência de negócios (*Business Intelligence* - BI) e analítica de dados que oferece uma plataforma *browser* para a construção de visualizações e *dashboards* dinâmicos e interativos. Esta plataforma é considerada fácil de usar e permite a conexão com diferentes fontes de dados, nomeadamente com o *PostgreSQL*, sendo que este automaticamente gera, na sua linguagem nativa *LookML* (derivado do *SQL*), código editável para integração e criação de vistas, agregações, entre outros. No entanto, a conexão com dados em formatos *Excel* ou similares não é suportada, sendo que essa tem de ser feita manualmente através da criação na plataforma de uma tabela derivada baseada em *SQL* [28]. Para além disso, é possível customizar as visualizações para além dos tipos pré definidos na sua biblioteca (e.g. gráficos de linhas, mapas de calor, entre outros). Esta plataforma tem associado um custo mínimo mensal elevado, para até 10 utilizadores.

2.2.3.2 Sisense

A *Sisense* [29] é uma empresa de software derivado em inteligência artificial para inteligência de negócios, cuja plataforma analítica consiste numa aplicação *web*. Esta ferramenta permite a conexão de dados locais ou na nuvem, como ficheiros *Excel* e a base de dados *PostgreSQL*, respetivamente. Contudo a integração desses dados na plataforma requer um bom conhecimento de *SQL*, uma vez que os conectores entre tabelas são estabelecidos manualmente, através de uma ferramenta chamada *Elasticubes*. Apesar disso, a plataforma fornece a funcionalidade "*Data Discovery*" [30] que permite explorar os dados através da criação de visualizações através de uma interface *drag-and-drop* e sem necessitar de *SQL*. Ainda assim, a liberdade de customização das visualizações é reduzida, sendo que a plataforma apenas oferece tipos de gráficos limitados e não customizáveis. A ferramenta tem um preço mensal associado por utilizador.

2.2.3.3 Qlik Sense

A *Qlik Sense* [31] é uma plataforma analítica que permite que qualquer pessoa explore os seus dados e construa visualizações de forma autónoma, através de um *browser* ou da sua versão *desktop*. Permite a conexão com ficheiros locais e bases de dados, nomeadamente com o *PostgreSQL*, sendo que a plataforma auxilia no processamento dos dados e de criação de visualizações através de recomendações e sugestões baseadas em inteligência artificial. A criação de visualizações é bastante customizável, para além de permitir a utilização e adaptação de *widgets* e extensões desenvolvidos e publicados por outros utilizadores. Esta plataforma não suporta nativamente ficheiro espaciais ou geocodificação, pelo que para construir visualizações geoespaciais está dependente do conetor e servidor *Qlik GeoAnalytics* e de extensões. Esta plataforma tem um custo mensal associado por utilizador, porém oferece programas académicos para atribuição de licenças gratuitas com esse propósito.

2.2.3.4 Microsoft Power BI

O *Power BI* [32] é um software inicialmente lançado pela *Microsoft* como um complemento para o *Excel*, sendo que depois se tornou numa solução analítica independente. Facilmente integra-se com bases dados como o *PostgreSQL* e com dados locais, especialmente se forem outros produtos da *Microsoft*, como os do *Office 365*. O processamento de dados por sua vez assemelha-se ao do *Microsoft Excel*, acrescido de um editor de *queries* quando necessário. Esta ferramenta é intuitiva e fácil de usar, pois utiliza funcionalidades *drag-and-drop*, mas só é rápida e executa melhor quando o volume de dados é limitado. Embora forneça um grupo variado de tipos de gráficos, a customização de visualizações é realizada através do desenvolvimento ou importação de *Power BI Visuals* [33], isto é, pacotes que incluem código-fonte (normalmente em *TypeScript*) para renderizar os dados fornecidos numa visualização personalizada. Esta ferramenta tem um custo base mensal associado por utilizador e não oferece licenças académicas.

2.2.4 Domo

O *Domo* [34] é uma ferramenta analítica inteligência de negócios baseada na nuvem, sendo o acesso feito através do *browser*. Permite a integração com uma vasta lista de base de dados e conectores, facultando também carregamento de ficheiros locais ou a integração de outras fontes via API. Não requer um grande conhecimento de *SQL* sendo o processamento de dados simples e embutido na plataforma. Para além disso, o processo de construção de visualizações utiliza *drag-and-drop* e demonstra ser simples e intuitivo de usar. A plataforma oferece alguns tipos de gráficos, cuja lista está em continua expansão, mas a customização de visualizações é o ponto forte do *Domo*, pois oferece a capacidade de adicionar visualizações personalizadas cuja base pode ser definida através de formatos *SVG*, *KML*, *GeoJSON*, *TopoJSON* ou *Shape* [35].

2.2.4.1 Tableau

O *Tableau* [13] oferece software para construção de visualizações de dados, entre as quais uma plataforma *desktop*, *mobile*, online e diversas extensões. O *Tableau* destaca-se no mercado por ter um elevado nível de desempenho, ser extremamente robusto e ter uma grande capacidade personalização e exploração de visualizações. Permite conectar facilmente diferentes fontes de dados, efetuar cálculos derivados sobre os dados e construir rapidamente visualizações utilizando *drag and drop*, promovendo a exploração e análise da informação. Opcionalmente, se for necessário uma preparação prévia dos dados, a ferramenta *Tableau Prep* [36] permite fazê-lo de forma simples e transformando-os em ficheiros *hyper*, para que sejam mais rápidos de processar pelas ferramentas de analítica. Para além disso, a *Extensions API* [37] do *Tableau* permite integrar aplicações externas e construir visualizações customizadas de forma a que sejam incorporadas e utilizadas diretamente na plataforma. Por sua vez, esta disponibiliza diferentes tipos de gráficos e personalização

dos mesmos, incluindo mapas embutidos para visualizar os dados geoespaciais, aos quais se podem adicionar camadas (i.e. relevo, satélite, etc) através de extensões. Ademais, permite a construção de narrativas e *dashboards* com visualizações interativas entre si. Para além de ser fácil de utilizar, o *Tableau* tem uma grande comunidade e fórum de suporte. No entanto, apenas distribui licenças gratuitas limitadas para fins académicos e projetos de investigação, sendo que uma licença comercial tem um custo mensal por utilizador.

2.2.4.2 Conclusão

Em suma, qualquer uma das ferramentas e plataformas apresentadas oferece soluções robustas para a construção de visualizações interativas de dados para o contexto definido.

No entanto, o *Looker* não suporta nativamente ficheiros locais como *Excel*, o que poderá atrasar a integração dos dados, ademais o processamento de dados poderá requerer a aprendizagem e escrita de código em *LookML*, linguagem derivada do *SQL*. O *Sisense* também utiliza ferramentas denominadas *Elasticubes* para processar os dados em *SQL*, sendo que além disso tem um número limitado de gráficos, que por sua vez não são customizáveis. Similarmente, o *Power BI* requer a criação de pacotes *Power BI Visuals*, com escrita de código-fonte para a construção de visualizações customizadas. Por outro lado, o *Qlik Sense*, ainda que se conete ao *PostgreSQL*, suporta uma quantidade limitada de fontes de dados, que ainda está a ser estendida, e, para além disso, não suporta nativamente ficheiros espaciais ou geocodificação.

Na tabela 2.1 podemos observar as **vantagens**, **dificuldades** e **desvantagens** de cada ferramenta apresentada anteriormente, tendo em conta os critérios estabelecidos. Através desta podemos concluir que nenhuma destas plataformas cumpre os critérios definidos.

	Looker	Sisense	Qlik Sense	Power BI
Integração de ficheiros locais	Não suporta XLS/CSV	Sim	Sim	Sim
Conexão com diferentes fontes de dados	Sim	Sim	Limitado	Sim
Processamento de Dados Simples	LookML	Elasticubes	Sim	Sim
Visualizações de Dados Temporais e Geoespaciais	Sim	Sim	Qlik GeoAnalytics	Sim
Exploração e Customização de Visualizações	Sim	Limitado e não é customizável	Sim	Power BI Visuals

Tabela 2.1: Resumo das funcionalidades-chave das ferramentas apresentadas.

O Domo e o Tableau são as duas ferramentas que apresentam-se como melhores candidatas para cumprir os critérios definidos. Por um lado, o *Tableau* apenas tem como desvantagem o facto do seu processamento de dados avançado necessitar da utilização de outra ferramenta, o *Tableau Prep*. Apesar disso, o *Tableau Desktop* por si só fornece processamento de dados simples e editor de *queries* para situações onde for necessário. Posto isto, facilmente percebemos que o *Tableau* lidera no mercado como a plataforma para construção de visualizações de forma rápida e com qualidade. Por outro lado, cada vez mais o Domo parece estar a tomar grande importância no mercado, apresentando uma vasta capacidade de conexão e integração de dados aliada a uma interface simples e intuitiva de *drag-and-drop*. Ainda assim, a maior vantagem do Domo em relação ao *Tableau* está na customização. Ainda que o *Tableau* ofereça alguma customização através da *Extensions API*, sem esta a customização torna-se uma tarefa de engenharia reversa da ferramenta e limitando o resultado obtido. Ainda assim, na maioria dos casos a vasta quantidade de gráficos fornecido em conjunto com essa customização é suficiente.

Embora o Domo possibilite uma customização mais alargada, esta plataforma trabalha apenas com empresas, tendo um custo personalizado associado, enquanto o *Tableau* para além da licença comercial paga, fornece um programa académico com licenças gratuitas. Além disso, o facto do *Domo* ser totalmente na nuvem dificulta o processo de extração dos dados, o que não acontece com o *Tableau*. Por esses motivos e dado que já foi realizado trabalho prévio para o projeto FitoAgro em *Tableau Desktop*, utilizando uma licença gratuita de investigação, considerou-se que o *Tableau* seria a ferramenta adequada para realizar o trabalho desta dissertação.

GRUPO OPERACIONAL FITOAGRO

Neste capítulo é descrito o grupo operacional FitoAgro, nomeadamente o problema que este pretende combater, os dados recolhidos, os modelos preditivos adequados e os *stakeholders* envolvidos.

3.1 Problema nas Pomóideas

Como já foi referido anteriormente (1.2), a produção de **pomóideas** na **região do Oeste** tem vindo a ser afetada por diferentes espécies de insetos que podem ter efeitos económicos nefastos para os agricultores. De forma a que se possa combater esses inimigos de forma eficaz, é necessário estudar o seu desenvolvimento e ciclo de vida. Embora o projeto estude diversas **pragas**, o seu foco incidiu sobre quatro espécies:

- Cecidómia (*Dasineura pyri* Bouché)
- Cochonilha-Algodão (*Pseudococcus viburni*)
- Filoxera (*Aphanostigma pyri* Chol.)
- Novo Lepidóptero (espécie não identificada)

De seguida, é apresentada uma descrição de cada uma destas espécies e os danos que estas podem causar, clarificando a necessidade deste projeto para monitorização e combate das mesmas.

3.1.1 Cecidómia

O adulto é um pequeno **díptero** (entre 1.5 mm a 2 mm), com as patas mais longas que o corpo, sendo as suas asas membranosas e transparentes. Os ovos têm cor amarela-alaranjada e são depositados nas folhas mais novas. Na **região do Oeste**, os adultos emergem durante a primavera e as larvas eclodem após 3 a 12 dias. Estas larvas neonatas são transparentes, sendo que, com o desenvolvimento, atingem 2 mm de comprimento e coloração entre o branco e o amarelo. Esta espécie pode ter várias gerações por ano, sendo que em Portugal tem, normalmente, entre 2 a 4. A última geração acontece em Julho ou Agosto, após a qual as larvas se enterram no solo, onde permanecem em **diapausa** até à Primavera seguinte (figura 3.1). [38]

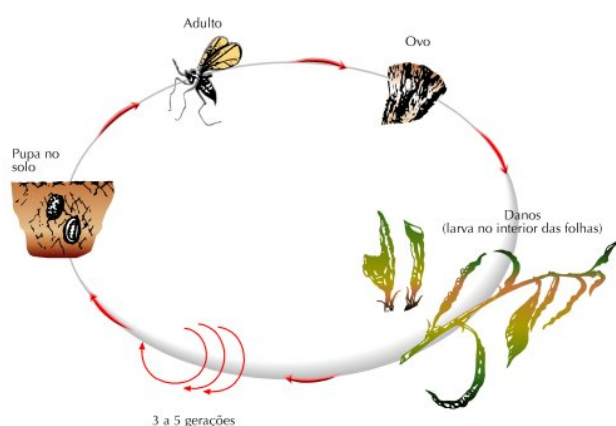


Figura 3.1: Ciclo de vida de *Dasineura pyri* Bouché [39]

Esta espécie causa danos ao nível das folhas, que se encontram enroladas longitudinalmente com larvas esbranquiçadas no seu interior (figura 3.2). Estas folhas tornam-se quebradiças por dessecação. Estes danos têm particular importância em pomares jovens, onde se verifica um ligeiro atraso na entrada em produção, podendo causar impacto económico para os produtores.



Figura 3.2: Folhas enroladas com larvas de *Dasineura pyri* Bouché [40]

3.1.2 Cochonilha-Algodão

A cochonilha-algodão é um pequeno **homóptero** da família *Pseudococcidae* que ataca vinhas e pomóideas. A origem da espécie *Pseudococcus viburni* encontra-se mal documentada devido a identificações incorretas da mesma no passado e à confusão, ainda existente, com uma espécie semelhante, *Pseudococcus maritimus*. No entanto, a *P. viburni*, além de algumas características morfológicas específicas, tem preferência por regiões costeiras e de clima mediterrânico. Além disso, apresenta 2 a 5 gerações anuais no Sul de França e entre 4 a 5 gerações anuais, no Sul de Itália. Estas diferenças devem-se à discrepância das temperaturas registadas entre estes países, sendo que a espécie atinge o máximo de fecundidade entre os 18 e 22°C. [41]

Este inseto tem um ciclo de vida dividido em 5 estados (figura 3.3): ovos, ninfas de primeiro, segundo e terceiro instar e adultos. Após a eclosão dos ovos, nos dois primeiros instares o desenvolvimento dos machos e das fêmeas é idêntico, sendo divergente nos estados seguintes. As fêmeas completam, ainda, um terceiro instar, enquanto que os machos segregam um invólucro (casulo) constituído por filamentos cerosos, no interior do qual irão pupar, correspondendo a mais dois instares (pré-pupa e pupa). Por fim, após estes estados, ambos atingem o estado de adulto (figura 3.4a), terminando o ciclo. [41]

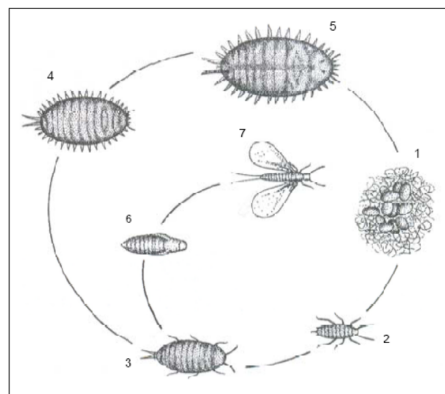


Figura 3.3: Ciclo de vida de uma cochonilha-algodão: 1 - ovos; 2 - ninfa de 1º instar; 3 - ninfa de 2º instar; 4 - ninfa de 3º instar; 5 - fêmea adulta; 6 - pré-pupa e pupa macho; 7 - macho adulto. [41]

Os danos causados por este inseto devem-se à sua ação alimentar ser diretamente nas folhas e ao facto de excretar grandes quantidades de melada para os frutos (figura 3.4b), o que os desvaloriza e pode causar prejuízos significativos para os produtores. Além disso, a melada constitui também um atrativo para outros organismos como alguns **lepidópteros**.

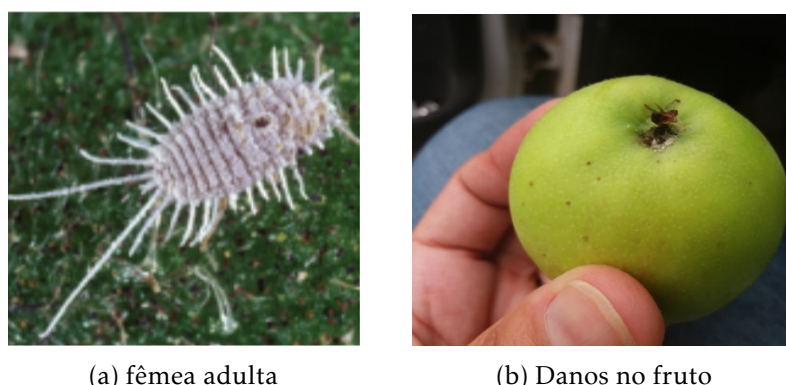


Figura 3.4: fêmea adulta e danos no fruto de *Pseudococcus viburni*.

3.1.3 Filoxera

A filoxera é um pequeno **homóptero** da família *Phylloxeridae* (entre 0,8 a 1 mm) de cor amarelada. Esta espécie hiberna sob a forma de ovos de Inverno, depositados nas saliências da casca dos troncos. A eclosão dá-se entre Março e Abril, sendo que estas larvas são muito móveis, mas ao evoluírem para adultos perdem a mobilidade. Entre 4 a 5 semanas, as larvas originadas sofrem 4 mudas, transformam-se nas fêmeas fundadoras, sendo que cada fêmea pode depositar entre 60 a 100 ovos (figura 3.5). A partir de meados de Junho, inicia-se a movimentação das larvas para os frutos. Após Setembro, podem surgir machos e fêmeas dos ovos. As fêmeas fecundadas põem então um único ovo em local abrigado, completando o ciclo.

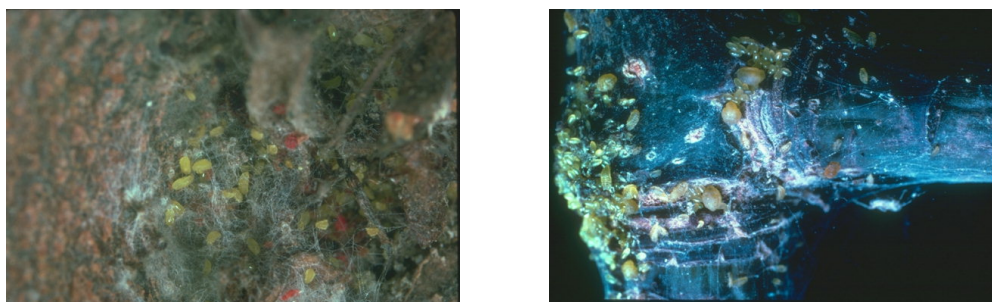


Figura 3.5: Colónias de filoxera na primavera (fêmeas, ovos e larvas) [42]

Por não existirem formas aladas (com asas), a praga não migra para outras plantas, sendo que a sua dispersão deve-se à acção do homem. Da acção alimentar da praga sobre os frutos surgem picadas no mesmo, que permitem a instalação de podridões, conduzindo à desvalorização comercial dos frutos. [42]

3.1.4 Novo Lepidóptero

Emergiu na região do oeste um novo tipo de lepidóptero, cuja espécie não é trivialmente identificável, sendo um dos objetivos do grupo operacional FitoAgro identificar a mesma para que se aplique metodologias de combate adequadas e eficazes. Foi proposto pelos

membros do projeto 3 espécies de lepidópteros para estudo: (*Cryptoblabes gnidiella*, *Cydia funebrana* e *Grapholita molesta* Busck). No entanto, com o desenvolvimento do projeto, a espécie principal em foco tornou-se a *Grapholita molesta* Buck.

Grapholita molesta Busck é uma espécie de insetos lepidópteros, mais especificamente de traças, da família *Tortricidae*. Este insecto tem como plantas hospedeiras o pessegueiro, mas também a pereira, macieira e ameixeira. Possui 4 estados de desenvolvimento no seu ciclo biológico: ovos, larvas, pupas e adultos. Os ovos são elípticos com cerca de 1 mm de comprimento, com uma coloração esbranquiçada (figura 3.6a), tornando-se posteriormente amarelada. As larvas apresentam 5 instares larvares podendo atingir 15 mm de comprimento no final do seu desenvolvimento, sendo que têm inicialmente o corpo branco tornando-se posteriormente cor-de-rosa, enquanto a cabeça apresenta uma coloração castanha clara. As pupas são castanhas escuras, com aproximadamente 6 mm de comprimento. Por fim, os adultos podem atingir até 16mm de envergadura, sendo que as asas dianteiras tem uma cor castanho escuro, enquanto que as posteriores apresentam uma coloração uniforme cinzenta acastanhado (figura 3.6b). [43]



(a) Larva



(b) Traça adulta

Figura 3.6: Larva e traça adulta de *Grapholita molesta*.

Após a eclosão, a larva ataca os rebentos jovens, onde o ponto de penetração situa-se, em geral, sobre a nervura da folha. Seguidamente, a larva vai formando uma galeria descendente, podendo esta atingir 3 para 5 cm. Deste modo, os maiores estragos situam-se nos frutos, causando diretamente estragos económicos para os produtores. [43]

3.2 Recolha de Dados

Ao longo do projeto FitoAgro foram recolhidos diversos tipos de dados que foram integrados automaticamente no sistema de informação do projeto, implementado previamente para esse efeito (A. Malafaia, 2018 [3]). Esses dados recaem essencialmente sobre duas categorias: dados fitossanitários recolhidos no campo e dados meteorológicos provenientes de estações meteorológicas internas e externas ao projeto.

3.2.1 Dados Fitossanitários

Os dados de campo são recolhidos *in situ* por técnicos que visitam semanalmente os POBs e registam as suas observações relativamente às árvores e equipamentos de monitorização

que lá se encontrem, denominados de *element of interest* (EOI). Existem atualmente no projeto treze POBs, sendo quatro de maçã e os restantes nove de pêra, como pode ser observado na figura 3.7.

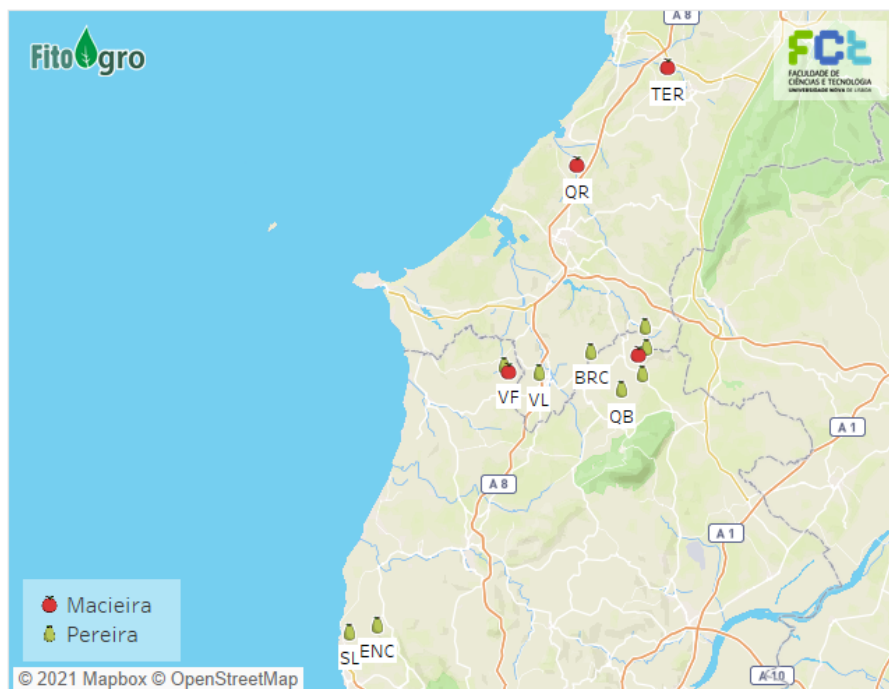


Figura 3.7: Mapa da região do Oeste onde se localizam os treze POBs do projeto FitoAgro, identificados por cultura (dados de 2021).

Para além da cultura, cada POB tem um inimigo principal associado (tabela 3.1) que, por sua vez, tem um protocolo de recolha de informação específico a ser seguido pelos técnicos durante das visitas. Estes protocolos foram elaborados pelos parceiros do projeto, tendo como objetivo o estudo detalhado dos inimigos principais (descritos em 3.1).

Código	Nome do POB	Cultura	Inimigo Principal
BB	Barreiras da Bica	Pereira	Cecidómia
BL	Boiça do Louro	Pereira	Cecidómia
BRC	Barrocalvo	Pereira	Filoxera
CS	Casal Serrano	Pereira	Cecidómia
ENC	Encarnação	Pereira	Filoxera
ISC	Iscarção	Pereira	Filoxera
JB	João Boi	Maceira	Cochonilha-Algodão
QB	Quinta do Brejo	Pereira	Cecidómia
QR	Quinta do Rato	Maceira	Novo Lepidóptero
SL	São Lourenço	Pereira	Cecidómia
TER	Termas	Maceira	Novo Lepidóptero
VF	Vale da Fonte	Maceira	Cochonilha-Algodão
VL	Vale Leite	Pereira	Cochonilha-Algodão

Tabela 3.1: Cultura e inimigo principal de cada POB (dados de 2021).

Existe ainda um protocolo generalista que deve ser executado em todas as visitas, independentemente do POB visitado, e para além do protocolo principal. Este protocolo generalista descreve as observações genéricas a serem realizadas para cada uma das pragas em estudo pelo projeto, sejam principais ou secundárias. O tipo de dados registado, dependendo do protocolo e da observação em execução, pode ser uma contagem, indicação de intervalos ou indicação de presença/ausência. Ademais, poderá ser necessário indicar o estágio de desenvolvimento do inimigo, o local onde se encontra e/ou a armadilha onde foi capturado.

Em adição aos dados recolhidos aquando da execução dos protocolos, os técnicos recolhem informação acerca do estado fenológico da planta, de ocorrências meteorológicas durante a visita, da realização de tratamentos por parte dos produtores do pomar e de outros detalhes que os técnicos considerem relevantes, sob a forma de anotações. Destes dados destaca-se o estado fenológico, uma vez que este indica a etapa de desenvolvimento em que a planta se encontra, o que também poderá ter influência no comportamento das diferentes pragas. Na tabela 3.2 apresenta-se os diferentes estados fenológicos para as culturas em estudo, a pera e a maçã, sendo que, como é possível observar, apenas o estado E difere entre culturas.

Estado Fenológico	Código	Pereira	Macieira
	A	Repouso Vegetativo	
	B	Pré-abrolhamento	
	C	Abrolhamento	
	C3	Escarchamento do Gomo	
	D	Ponta Verde	
	D3	Botão verde	
	E	Botão Rosa	Botão Branco
	F	Floração	
	F1	Desbrochamento (1ª Flor)	
	F2	Plena Floração	
	G	Queda das Pétalas	
	H	Queda Total das Pétalas	
	I	Vingamento	
	J	Frutos em Desenvolvimento	
K	Maturação		

Tabela 3.2: Estados fenológicos por cultura.

O registo destas informações adicionais tem como objetivo permitir a análise e descoberta *a posteriori* de fatores que poderão influenciar as observações registadas, sendo que, principalmente as anotações, podem explicar anomalias ou valores atípicos nos dados.

A recolha de dados é feita ao longo de determinados períodos de tempo denominados de campanhas. Esses períodos variam entre culturas, sendo que no caso do projeto FitoAgro, para que o período de observação seja sistemático, as campanhas são anuais ou até outubro/novembro de cada ano.

Anteriormente todos os registos das visitas eram feitos em papel e depois transcritos manualmente para uma folha *Excel* numa *Google Drive*, criada no âmbito do projeto. Atualmente, esse processo foi substituído por uma aplicação móvel, implementada no contexto da dissertação de M. Correia (2020) [44], que, após cada visita, valida e integra automaticamente os dados no sistema de informação do projeto.

Com a recolha e validação de todos estes dados fitossanitários, pretende-se calcular a percentagem de ataque para cada inimigo e estimar a sua variação com outros fatores influentes. A percentagem de ataque mede a intensidade do ataque de uma praga num determinado pomar, dependendo da forma como a praga o demonstra e da amostra recolhida. Por exemplo, a cecidómia, como referido em 3.1.1, ataca e causa danos ao nível das folhas pelo que a intensidade do seu ataque é medida pela percentagem de rebentos atacados em 100 observados. Por sua vez, o **nível económico de ataque (NEA)** de cada praga indica a partir (e/ou abaixo) de que intensidade de ataque deverão ser aplicadas medidas, limitativas ou de combate, para impedir correr riscos de prejuízos superiores ao custo das medidas de luta a adotar, acrescido dos efeitos indesejáveis que estas últimas podem provocar. Novamente no caso da cecidómia, o **NEA** está estabelecido em 15% para pomares jovens e em 50% para pomares adultos, pelo que em pomares com percentagens de ataque iguais ou superiores a esse valor, deverá ser considerada a tomada de medidas de combate [45]. Assim, o cálculo e visualização desses indicadores é de grande importância para a tomada de decisão por parte dos produtores.

Em suma, a visualização interativa de dados terá um papel importante na deteção de anomalias na recolha, no estudo do desenvolvimento das pragas e culturas e na tomada de decisão dos produtores na aplicação de medidas de combate.

3.2.2 Dados Meteorológicos

Uma vez que a meteorologia é um fator muito influente no desenvolvimento das pragas e culturas, foram recolhidos, ao longo do projeto, diversos dados meteorológicos provenientes de estações meteorológicas fornecidas por parceiros do projeto e entidades externas, nomeadamente a Associação dos Produtores Agrícolas da Sobrena (APAS) [46], o Instituto Português do Mar e da Atmosfera (IPMA) [47], a *Pessl Instruments* [48], a *TerraPro* [49] e a *WiseCrop* [10]. Na tabela 3.3 é apresentado o número de estações concedidas por cada fornecedor.

Fornecedor	Parceiro	Total de Estações
APAS	Sim	14
IPMA	Não	7
Pessl	Sim	10
TerraPro	Sim	3
WiseCrop	Sim	6

Tabela 3.3: Sensores selecionados como relevantes e a sua unidade correspondente.

Todas estes fornecedores e as suas respetivas estações já foram previamente configurados no sistema de informação [3], para que os dados meteorológicos gerados por estes sejam automaticamente carregados e integrados no sistema. Todavia, dada a diversidade de fornecedores e estações, existe também variações no formato dos dados, nomeadamente no fuso horário e frequência de amostragem utilizada. Deste modo, foi realizado, no contexto da dissertação de L. Correia (*unpublished*) [50], um processamento dos dados com o objetivo de uniformizar todas as estações para o fuso horário UTC (*Coordinated Universal Time*) e para uma frequência de amostragem horária.

Cada estação apresenta um conjunto variado de sensores meteorológicos, contudo, no contexto do projeto, foram selecionados apenas os sensores relevantes para o desenvolvimento das pragas. Na tabela 3.4 apresentam-se os sete sensores selecionados como relevantes e a sua respetiva unidade.

Sensor	Unidade
Temperatura do ar	Graus Celsius (°C)
Humidade relativa	Percentagem (%)
Precipitação acumulada	Milímetros (mm)
Ponto de orvalho	Graus Celsius (°C)
Indicador de folha molhada	Minutos (min)
Velocidade do vento	Minutos por segundo (m/s)
Radiação solar	Watts por metro quadrado (W/m ²)

Tabela 3.4: Sensores selecionados como relevantes e a sua respetiva unidade.

Espera-se que a visualização interativa de dados facilite a deteção de situações anómalas não previstas nos dados e permita a análise do impacto da meteorologia no desenvolvimento das diferentes pragas e culturas.

3.3 Modelos Preditivos

Como referido em 1.2, o grupo operacional FitoAgro pretende desenvolver modelos de previsão para todas as pragas em estudo, tanto para as principais (enunciadas em 3.1), como para as secundárias que estão a ganhar relevância na região.

Para perceber o conceito de modelos preditivos para **pragas** é necessário em primeiro lugar descrever modelos de desenvolvimento de **pragas**. Os modelos de desenvolvimento utilizam parametrizações para o cálculo de variáveis derivadas da meteorologia que, por sua vez, são utilizados para identificar as diferentes fases de desenvolvimento do inseto. As parametrizações utilizadas são normalmente obtidas num ambiente controlado em laboratório, onde são determinadas das condições ótimas para o desenvolvimento de cada inseto e fora das quais estes não se conseguem desenvolver. Já existem estudos que determinaram quais as condições ótimas para algumas das **pragas** em estudo e identificaram as suas diferentes fases de desenvolvimento utilizando essas condições como parametrização dos seus modelos de desenvolvimento. Posteriormente neste documento, serão apresentados e detalhados exemplos de estudos e modelos existentes.

Todavia, dado o microclima da **região do Oeste**, as alterações climáticas, e possivelmente outros fatores influentes, o momento em que essas fases ocorrem é variável. Por esse motivo, utilizam-se modelos preditivos para ajustar os modelos de desenvolvimento obtidos em laboratório às condições climáticas registadas em cada local e a outras variáveis que possam ser consideradas relevantes. Note-se que os modelos estimam as condições meteorológicas através da obtenção de estações na proximidade de cada local. Ao contrário dos modelos de desenvolvimento, os modelos preditivos retornam o(s) momento(s) em que é expectável que ocorram as diferentes fases de desenvolvimento já identificadas, através de datas ou outras variáveis temporais relevantes. Por sua vez, a previsão desses momentos permite auxiliar os produtores na tomada de decisão relativamente à aplicação de medidas de combate, tornando-o mais eficaz e eficiente.

A construção de modelos preditivos para **pragas** encontra-se fora do âmbito deste trabalho, tendo sido realizada no contexto da dissertação de L. Correia (*unpublished*) [50]. Contudo, a visualização de dados será necessária para observar o ajuste dos modelos gerados à informação observada localmente e, desta forma, adaptar a parametrização utilizada à realidade. Assim, a visualização interativa de dados possibilitará a validação iterativa dos modelos, gerando confiança nos mesmos por parte dos seus respetivos *stakeholders*, apresentados na próxima secção.

3.4 Stakeholders

Os parceiros do projeto FitoAgro provêm de diversas instituições (enumeradas em 1.2), com diferentes contributos e objetivos. Tendo em consideração a área de interesse de cada parceiro, é possível identificar cinco tipos de *stakeholders* para a visualização de dados:

- **Técnicos de Recolha de Dados** - Monitorização da recolha dos dados de campo.
- **Técnicos da Produção** - Gestão das culturas, **POBs** e campanhas.
- **Cientistas** - Estudo do desenvolvimento e ciclo de vida das **pragas** através da visualização de dados de campo e do ajuste de modelos preditivos.

- **Computer Science** - Manutenção do sistema de informação, da visualização de dados e dos modelos preditivos.
- **Gestão do Projeto** - Gestão do projeto de monitorização fitossanitária, tanto da parte de monitorização e controlo como da analítica dos dados.

De forma a responder a todas as necessidades analíticas de projetos de monitorização fitossanitária, a visualização interativa de dados deve considerar os requisitos analíticos necessários para cumprir os objetivos de cada tipo de *stakeholders* enumerado acima. No capítulo seguinte, será feito o levantamento e descrição desses requisitos.

REQUISITOS ANALÍTICOS

A componente analítica é o que permite que seja extraído conhecimento dos dados recolhidos. Neste capítulo é feito um levantamento dos requisitos analíticos necessários para analisar os dados meteorológicos e fitossanitários, bem como para promover a descoberta de eventuais correlações entre estes. Estes requisitos foram classificados consoante a sua finalidade, nomeadamente para monitorização e controlo de qualidade dos dados (secção 4.1) e para análise da sua evolução em diferentes locais e anos (secção 4.2).

4.1 Monitorização e Controlo de Qualidade

Antes de analisar os dados recolhidos é necessário garantir que estes estão corretos e atualizados. Embora os métodos computacionais de verificação e deteção de erros tenham o seu papel, a visualização de dados, se adequadamente construída, permite por um lado a deteção de anomalias não previstas, através de inspeção visual, e por outro lado apresentar aos *stakeholders* apropriados a qualidade dos dados, gerando assim confiança nos mesmos.

Nesta secção descreve-se os requisitos analíticos necessários para monitorização e controlo de qualidade dos dados, sejam estes provenientes de estações meteorológicas (secção 4.1.1) ou das visitas de campo realizadas pelos técnicos (secção 4.1.2). Note-se a utilização da denotação *MCQ* para identificar requisitos de **M**onitorização e **C**ontrolo de **Q**ualidade.

4.1.1 Estações Meteorológicas

As estações meteorológicas muitas vezes avariam ou apresentam falhas em alguns dos seus sensores, sendo que isto se reflete como períodos de ausência de dados. Deste tipo de situação surgem duas necessidades analíticas:

MCQ R1. Detetar estações avariadas e falhas de sensores para que se possa sinalizar a sua reparação.

MCQ R2. Monitorizar a qualidade das estações e respetivos sensores ao longo do tempo, de forma a identificar estações propícias a falhas *versus* estações fiáveis.

Para além de apenas detetar falhas e avarias é necessário **analisar os dados meteorológicos em bruto, identificando falhas através da ausência de dados**, i.e. quebras na série temporal (exemplo na figura 4.1a), **ou inconsistências nos mesmos** (exemplo na figura 4.1b).



(a) Ausência de dados no dia 26 de Dezembro de 2020.



(b) Temperatura média diária do ar a atingir os 64,5°C no dia 29 de Janeiro de 2021.

Figura 4.1: Exemplo de falhas e inconsistências nos valores diários do sensor temperatura do ar da estação Apas04 (APAS).

Finalmente, identifica-se assim um requisito adicional para a monitorização e controlo da qualidade das estações meteorológicas, descrito abaixo:

MCQ R3. Observar, ao longo do tempo, os dados em bruto dos diferentes sensores meteorológicos, para diferentes estações, identificando falhas através da ausência de dados ou inconsistências nos mesmos.

4.1.2 Visitas de Campo

Como descrito anteriormente em 3.2.1, os dados fitossanitários são recolhidos durante visitas de campo realizadas pelos técnicos aos POBs do projeto. Antes da recolha através da aplicação móvel, os técnicos estavam encarregues de copiar os dados para folhas *excel* na *Google Drive*, sendo que estas serviam também para **consulta e deteção a posteriori de erros e inconsistências**. Com a utilização da aplicação móvel, as folhas *excel* deixaram de ser utilizadas, apresentando-se a necessidade analítica de continuar a permitir essa verificação, como descrito no requisito abaixo:

MCQ R4. Verificar a correção e consistência os dados recolhidos em cada visita de campo.

4.2 Analítica dos Dados

Após garantir a monitorização e controlo da qualidade dos dados, é necessário analisar os seus valores e a sua evolução ao longo do tempo. Assim sendo, a visualização de dados deverá contemplar uma componente analítica dos dados para que os *stakeholders* possam analisar e compreender os dados de forma a cumprir os objetivos que lhes competem no projeto. Nesta secção apresentam-se os requisitos analíticos para essa componente, nomeadamente para analisar dados provenientes de sensores meteorológicos (secção 4.2.1) e de observações fitossanitárias (secção 4.2.2), analisar correlações entre a meteorologia e a fitossanidade (secção 4.2.3) e, finalmente, analisar o ajuste de modelos preditivos às observações fitossanitárias registadas (secção 4.2.4). Note-se a utilização da denotação *AD* para identificar requisitos de Analítica de Dados.

4.2.1 Sensores Meteorológicos

A evolução da meteorologia tem um grande impacto no desenvolvimento das pragas e culturas pelo que a **análise dos valores registados pelos sensores meteorológicos** é de grande importância para a monitorização fitossanitária. Contudo, o auge dessa importância acontece quando os **eventos meteorológicos ocorrem na proximidade do local onde se encontram os inimigos e culturas em estudo, neste caso dos POBs**. Para além dos valores registados pelos sensores meteorológicos, na agronomia utilizam-se **variáveis derivadas destes para prever o desenvolvimento vegetativo e das pragas**, das quais, para projetos de índole fitossanitárias, se destacam os valores acumulados das horas de frio e dos graus-dia. Tendo em consideração as necessidades analíticas levantadas acima relativamente aos sensores meteorológicos, é possível identificar três requisitos para a analítica de dados:

- AD R1.** Observar a evolução dos valores registados pelos diferentes sensores meteorológicos para diferentes estações, granularidades e intervalos de tempo.
- AD R2.** Observar a evolução dos valores registados pelos diferentes sensores meteorológicos nos locais de cada *POB*, para diferentes granularidades e intervalos de tempo
- AD R3.** Calcular e observar a evolução de variáveis derivadas da meteorologia relevantes para o desenvolvimento das pragas e culturas em estudo.

4.2.2 Observações Fitossanitárias

A análise das observações fitossanitárias e da sua evolução permite estudar o desenvolvimento das pragas e descobrir fatores influentes no seu ciclo de vida. Pretende-se, assim, conseguir delinear estratégias de combate adequadas para cada inimigo e momento, bem como auxiliar na tomada de decisão dos produtores.

Como referido em 3.2, o tipo de dados recolhido é variado, sendo necessário **observar a evolução de todos estes dados para promover a descoberta de correlações**. Ademais, essa **evolução pode variar consoante o local e ano em análise, pelo que é importante que se possa fazer a análise comparativa destes** de forma a descobrir quais os fatores influentes nessa variação, caso exista. Finalmente, também foi referida a necessidade de **calcular e apresentar, para cada inimigo, a percentagem de ataque e o respetivo nível económico de ataque (NEA)**, para que os produtores possam fazer uma estimativa de risco e tomar decisões conforme. Surgem assim três requisitos para a analítica de dados fitossanitários:

- AD R4.** Observar, para cada protocolo ou inimigo, a evolução das visitas, observações fitossanitárias, **estado fenológico**, tratamentos e ocorrências meteorológicas registadas, promovendo a descoberta de correlações nos mesmos.
- AD R5.** Comparar, para cada protocolo ou inimigo, a evolução de toda a informação fitossanitária em diferentes **POBs** e/ou campanhas.
- AD R6.** Calcular e apresentar, para cada inimigo, a respetiva percentagem de ataque e nível económico de ataque (**NEA**).

4.2.3 Correlação da Meteorologia com a Fitossanidade

Como referido anteriormente, os dados meteorológicos tem um grande impacto no desenvolvimento das **pragas** e culturas. Deste modo, para além da análise dos valores registados pelos sensores meteorológicos, é fundamental **promover a correlação da evolução destes valores ou de variáveis derivadas dos mesmos com as diferentes observações fitossanitárias registadas**. Espera-se, assim, compreender melhor quais as condições meteorológicas propícias para o desenvolvimento das diferentes **pragas** e o motivo da existência ou ausência de variações no seu desenvolvimento nos diferentes locais e anos. Esta necessidade analítica traduz-se no requisito expresso abaixo:

- AD R7.** Correlacionar ou promover a descoberta de correlações entre a meteorologia (ou variáveis derivadas da mesma) e a informação fitossanitária registada, para diferentes locais e/ou anos.

4.2.4 Ajuste de Modelos Preditivos

Como descrito em 3.3, os modelos preditivos foram gerados no contexto de outra dissertação, no entanto, requerem a **observação analítica do seu ajuste às observações fitossanitárias registadas**.

- AD R8.** Observar, para cada inimigo, o ajuste dos modelos preditivos existentes às observações fitossanitárias registadas.

4.3 Resumo

O levantamento dos requisitos analíticos é fundamental para a construção de visualizações interativas de dados que respondam a todas as necessidades dos seus *stakeholders*. Nesse sentido, apresenta-se na tabela 4.1 o mapeamento entre os *stakeholders*, descritos em 3.4, e os 12 requisitos analíticos levantados neste capítulo.

<i>Stakeholders</i>	Requisitos Analíticos											
	Monitorização e Controlo de Qualidade				Analítica dos Dados							
	1	2	3	4	1	2	3	4	5	6	7	8
Técnicos de Recolha de Dados				X				X	X	X		
Técnicos da Produção				X	X	X	X	X	X	X	X	X
Cientistas							X	X	X		X	X
<i>Computer Science</i>	X	X	X									X
Gestão do Projeto	X	X	X	X	X	X	X	X	X	X	X	X

Tabela 4.1: Mapeamento dos requisitos analíticos aos *stakeholders* a que se associam.

Para os técnicos de recolha de dados compete a monitorização e controlo de qualidade dos dados de campo (MCQ R4), bem como a análise preliminar da evolução desses dados ao longo do tempo e entre POBs e/ou campanhas (AD R4 e R5, respetivamente) e do cálculo de percentagens de ataque e NEAs em relação ao que foi observado (AD R6). Por outro lado, para os técnicos da produção, como o seu objetivo é a gestão das culturas e POBs, estão interessados na monitorização das observações fitossanitárias registadas MCQ R4 e em toda a análise dos dados sejam estes meteorológicos (AD R1 a R3), de observações fitossanitárias (AD R4 s R6), para a descoberta de correlações entre meteorologia e a fitossanidade (AD R7) para a visualização do ajuste de modelos preditivos para pragas (AD R8). A abrangência no escopo de requisitos deve-se ao facto dos técnicos de produção necessitarem de analisar toda a informação disponível para tomar decisões relativamente à aplicação (ou não) de medidas preventivas, limitativas ou de combate.

Contrariamente aos técnicos de produção, os cientistas apenas têm interesse em alguns dos requisitos para análise dos dados, nomeadamente os que os permitem estudar o desenvolvimento das pragas e os seus ciclos de vida. Destacam-se, assim, os requisitos que permitem observar a evolução de variáveis derivadas influentes no desenvolvimento das pragas e culturas (AD R3), de observações fitossanitárias ao longo do tempo e entre

POBs e/ou campanhas (**AD R4** e **R5**, respetivamente) e do ajuste de modelos preditivos (**AD R8**).

Do ponto de vista da *computer science* interessa maioritariamente a monitorização e o controlo da qualidade dos dados meteorológicos (**MCQ R1** a **R3**), bem como a observação do ajuste de modelos preditivos (**AD R8**) para ajustar as parametrizações utilizadas, se necessário.

Finalmente, para aqueles encarregues pela gestão do projeto é relevante todo o tipo de monitorização e controlo de qualidade dos dados (**MCQ R1** a **R4**), bem como de analítica dos mesmos (**AD R1** a **R8**), para que possam garantir o bom funcionamento do projeto.

Os 12 requisitos analíticos levantados neste capítulo encontram-se sintetizados na tabela [A.1](#) em apêndice, sendo que 4 dos requisitos têm por objetivo a monitorização e controlo de qualidade dos dados e 8 a analítica dos mesmos.

ABORDAGEM E METODOLOGIA

Neste capítulo será descrita a abordagem e metodologia seguida durante o processo de desenho e implementação das visualizações interativas dos dados da componente analítica desenvolvida. Começa-se por definir as boas práticas de desenho (secção 5.1), seguindo-se a especificação do fluxo iterativo de avaliação (secção 5.2) e, finalmente, a modelação de uma base de dados analítica como resposta aos requisitos analíticos elicitados.

5.1 Definição de Boas Práticas de Desenho

De forma a facilitar a compreensão das visualizações de dados, foi necessário definir um conjunto de boas práticas de desenho das mesmas. Em primeiro lugar, especificou-se como boa prática a modelação das visualizações de dados *a priori* da sua construção (secção 5.1.1). Em segundo lugar, estabeleceu-se um conjunto de recomendações gerais (secção 5.1.2) que vão além dos fundamentos básicos de visualização interativa de dados descritos na secção 2.2.1.

5.1.1 Modelação de Visualizações Interativas de Dados

O desenho e implementação de visualizações interativas de dados deve começar com a modelação de visualizações que possam responder às questões analíticas levantadas. Atualmente, as linguagens de modelação existentes abordam temas como a estruturação de sistemas, interfaces e/ou páginas web. No entanto, nenhuma destas linguagens aborda todos os componentes, conceitos e atributos que a área da visualização interativa de dados engloba. Por este motivo, surgiu a iniciativa de criar uma linguagem de modelação específica para esse âmbito, envolvendo vários professores e alunos do grupo de investigação de Sistemas Inteligentes para a Monitorização de Recursos Naturais (SI-MORENA). Este grupo é muito diverso, uma vez que engloba várias temáticas como a

monitorização agrícola, realidade virtual, aumentada e mista, deteção remota e problemas espaço-temporais. Apesar da sua diversidade, a visualização interativa de dados apresenta ser uma necessidade comum a maioria destas temáticas.

A *Interactive Visualization Modeling Language* (IVML) é uma linguagem de modelação para visualização interativa de dados que pretende responder às necessidades de modelação específicas da sua área. A sua construção teve como base a linguagem de modelação *interactive flow modeling language* (IFML), linguagem *standard* do *Object Management Group* (OMS) para desenho de interfaces *front-end* [51]. Embora o IFML seja específico para a modelação de interfaces, tem como objetivo a modelação de um conceito fundamental para a visualização interativa de dados: a interatividade dos utilizadores com os componentes que lhes são apresentados. No entanto, uma vez que o IFML modela apenas os componentes básicos da interação, o IVML estende esses componentes e o próprio fluxo de interação para responder ao nível de interatividade e especificidade que a visualização interativa de dados impõe. Esta linguagem de modelação é descrita em mais detalhe no apêndice B.

5.1.2 Recomendações Gerais

Algumas boas práticas foram seguidas para facilitar a compreensão dos utilizadores relativamente à visualização que lhe é apresentada. Seguem-se abaixo algumas recomendações gerais nesse sentido:

- **Título da visualização** - De forma a que os utilizadores possam ficar com uma noção do objetivo da visualização antes de a analisarem, é importante que haja um título que, de forma breve, explique o propósito da mesma e possíveis filtragens relevantes. Por exemplo, através apenas do título "Dados Meteorológicos no dia 30/11/2020 na estação Coopval01VV (APAS)" é possível obter informação acerca do que contém a visualização e qual o dia e estação meteorológica (filtros) a ser apresentada de momento. Preferencialmente, o título deverá ser colocado no topo da visualização e com um tamanho de fonte facilmente visível, para que este seja o primeiro elemento a ser observado pelo utilizador.
- **Filtros** - Uma vez que existem vários tipos de filtros (escolha única, seleção múltipla, intervalos, etc), é importante que o título destes refira o tipo de interação necessária, bem como a variável a ser filtrada. Por exemplo, através do título "Escolher estação" é possível inferir que se está a filtrar estações meteorológicas e que é um filtro de opção única, reforçado através do verbo "escolher" e do singular de "estação". Em alternativa, se a filtragem fosse sobre várias estações, o título deveria ser "Selecionar estações", sendo que a utilização do verbo "selecionar" e do plural de "estação" indicam seleção múltipla.
- **Legendas** - As legendas são um elemento muito importante para a compreensão das visualizações por parte do utilizador, visto que a sua apresentação inadequada pode

gerar confusão durante a análise da visualização, sobrepor informação relevante e/ou ocupar espaço desnecessário. Por norma, as legendas deverão situar-se acima ou sobrepor a visualização à qual está associada, sendo que no caso da sobreposição deverá garantir-se que, independentemente dos dados que poderão vir a apresentados, a legenda nunca sobrepõe o espaço ocupado pelos dados. Apenas se deve omitir o título da legenda quando esta por si só é auto-explicativa, o que normalmente apenas acontece com variáveis categóricas e booleanas (e.g. macieira/pereira na figura 3.7). Finalmente, as legendas deverão ser geradas tendo em consideração os fundamentos básicos para variáveis visuais, descritos na secção 2.2.1.

- **Eixos** - A estrutura base de um eixo numa visualização é trivial, no entanto, quando esse eixo é partilhado ou replicado para várias visualizações existem alguns fatores a ter em consideração. Em primeiro lugar, quando o objetivo das visualizações é a comparação da evolução dos dados ao longo de uma determinada variável, deve-se garantir que os eixos são uniformes entre si e ao longo de todo o *dataset* se necessário. Por exemplo, na figura 5.1a apresenta-se duas séries temporais que aparentam ser visualmente semelhantes, no entanto a escala de cima varia entre 0 e 14 e a de baixo de 0 e 5. Se uniformizarmos as escalas para ambas variarem até 14 (como na figura 5.1b), torna-se claro visualmente que a série temporal de baixo tem valores muito inferiores à de cima. Complementarmente, a utilização de etiquetas com o valor de cada ponto poderá facilitar a comparação de dados. Por outro lado, quando existem muitas visualizações a partilhar o mesmo eixo ou estas estão muito distantes entre si e/ou do eixo, este deverá ser replicado no topo ou na posição central das mesmas. Finalmente, os títulos nos eixos não temporais devem conter o nome da variável ou do conjunto observado e, quando necessário, a unidade associada.

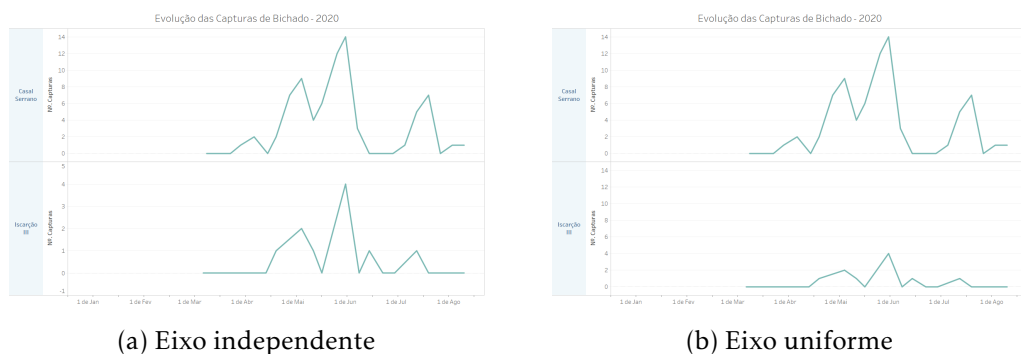


Figura 5.2: Comparação visual pode induzir em erro dependendo do tipo de eixo utilizado na visualização.

- **Navegação e interação vs. Deslocação vertical** - Por vezes a informação que se pretende apresentar é muito extensa, resultando numa falta de espaçamento vertical e consequentemente na necessidade de deslocar a informação verticalmente (i.e. fazer *scroll*). Esse deslocamento poderá resultar no fenómeno perceptual *change blindness*,

explicado na secção 2.2.1, comum na visualização de dados. Por esse motivo, deverá-se optar, sempre que possível, por sumarizar/agregar a informação de modo a que ocupe todo o espaço disponível, sem necessidade de deslocamentos. No entanto, isto poderá resultar na omissão de informação detalhada acerca dos dados, o que poderá ser contornado através de *tooltips* que apresentem esses detalhes ou da navegação para outras visualizações detalhadas. Assim, é possível manter a base de comparação e análise através da visualização agregada e/ou sumarizada e, de igual modo, aceder aos dados detalhados.

5.2 Processo Iterativo de Avaliação

O processo de desenho e implementação de visualizações interativas de dados deverá começar com o levantamento das questões e requisitos analíticos, seguindo-se a avaliação do tipo os dados necessários e, finalmente, a modelação e implementação dessas visualizações. Após concluído esse processo, a primeira versão da visualização deverá ser avaliada em reunião interna, onde podem ser propostas melhorias e alterações que resultarão em novas iterações da visualização. Aquando da definição de uma versão polida da visualização, esta deverá ser apresentada aos *stakeholders* a quem a mesma se destina. Estes poderão apresentar dificuldades e melhorias que deverão ser analisadas, resultando, ou não, em novas iterações e apresentações. Finalmente, a versão aceite pelos *stakeholders* deverá ser apresentada ao gestor do projeto, garantindo assim que a visualização integrase no modelo de produção e é de fácil compreensão não só pelo seu público-alvo, mas por todos os membros intervenientes em projetos de monitorização fitossanitária. Por sua vez, o gestor poderá também sugerir alterações ou melhorias resultando em novas iterações. Na figura 5.3 é possível observar fluxograma que representa todo este processo de iterativo de avaliação.

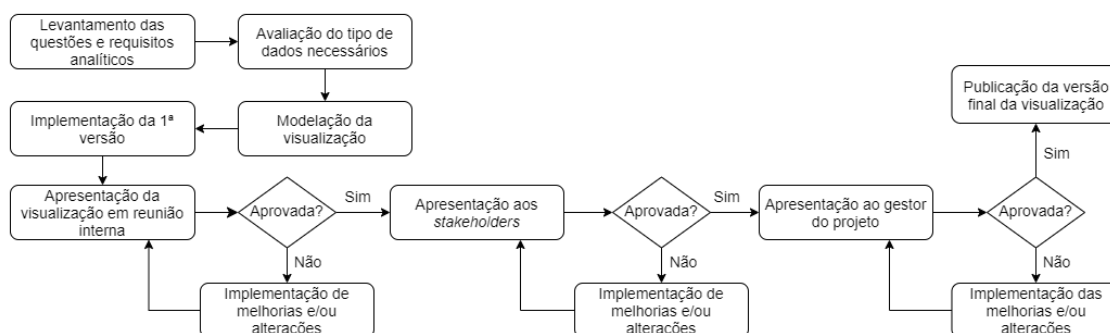


Figura 5.3: Fluxograma do processo iterativo de avaliação que deverá ser seguido para visualização de dados.

5.3 Modelação da Base de Dados Analítica

O sistema de informação do projeto FitoAgro foi desenvolvido *à priori* por A. Malafaia (2019), sendo que a modelação da sua camada de armazenamento já integrava uma base de dados analítica contendo um *data warehouse* [3], sob a forma de vista materializada. Esta vista foi criada com o objetivo de servir de depósito de dados para uma futura componente analítica, no entanto, a sua modelação genérica tornou-se insuficiente para corresponder às expetativas analíticas impostas pelos requisitos descritos no capítulo anterior. Deste modo, decidiu-se em reunião interna estender a base de dados analítica existente com a criação de outro depósito sob a forma de vista materializada, ao qual se denominou *publish*. O objetivo deste depósito é integrar dados apenas para fins analíticos e no formato requisitado para a construção de visualizações interativas de dados. Na figura 5.4 é possível observar a nova estrutura da camada de armazenamento, cuja base de dados analítica passa a ter dois depósitos de dados: o *data warehouse* e o *publish*.

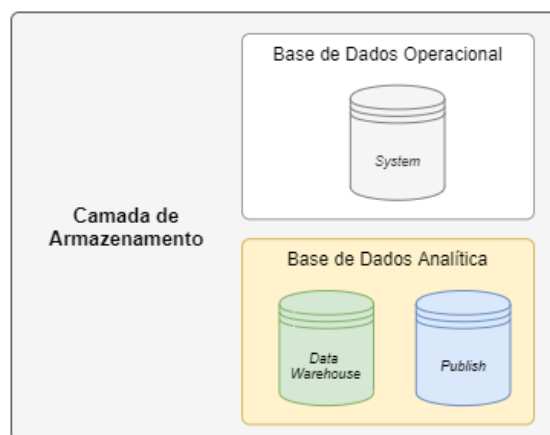


Figura 5.4: Camada de armazenamento do sistema de informação, cuja base de dados analítica contém o *data warehouse* e o *publish*.

Com esta estrutura, o *data warehouse* continua a servir de depósito de dados para a componente analítica sempre que não é necessário processar ou transformar os dados do sistema de informação, como por exemplo para monitorização e controlo de dados. Por outro lado, o *publish* fornece à componente analítica os dados noutros formatos que facilitem a sua análise ou que sejam mais apropriados para a construção de visualizações interativas de dados. É importante realçar que a modelação e manutenção destas vistas e do respetivo sistema de informação não foi realizada no âmbito desta dissertação, sendo que apenas se propôs e discutiu conjuntamente a necessidade de outro depósito de dados.

Na figura 5.5 apresenta-se o modelo de dados utilizado pela componente de visualização, onde se representa as tabelas pertencentes ao *data warehouse* a verde, as do *publish* a azul e as duplicadas para ambas as vistas a amarelo.

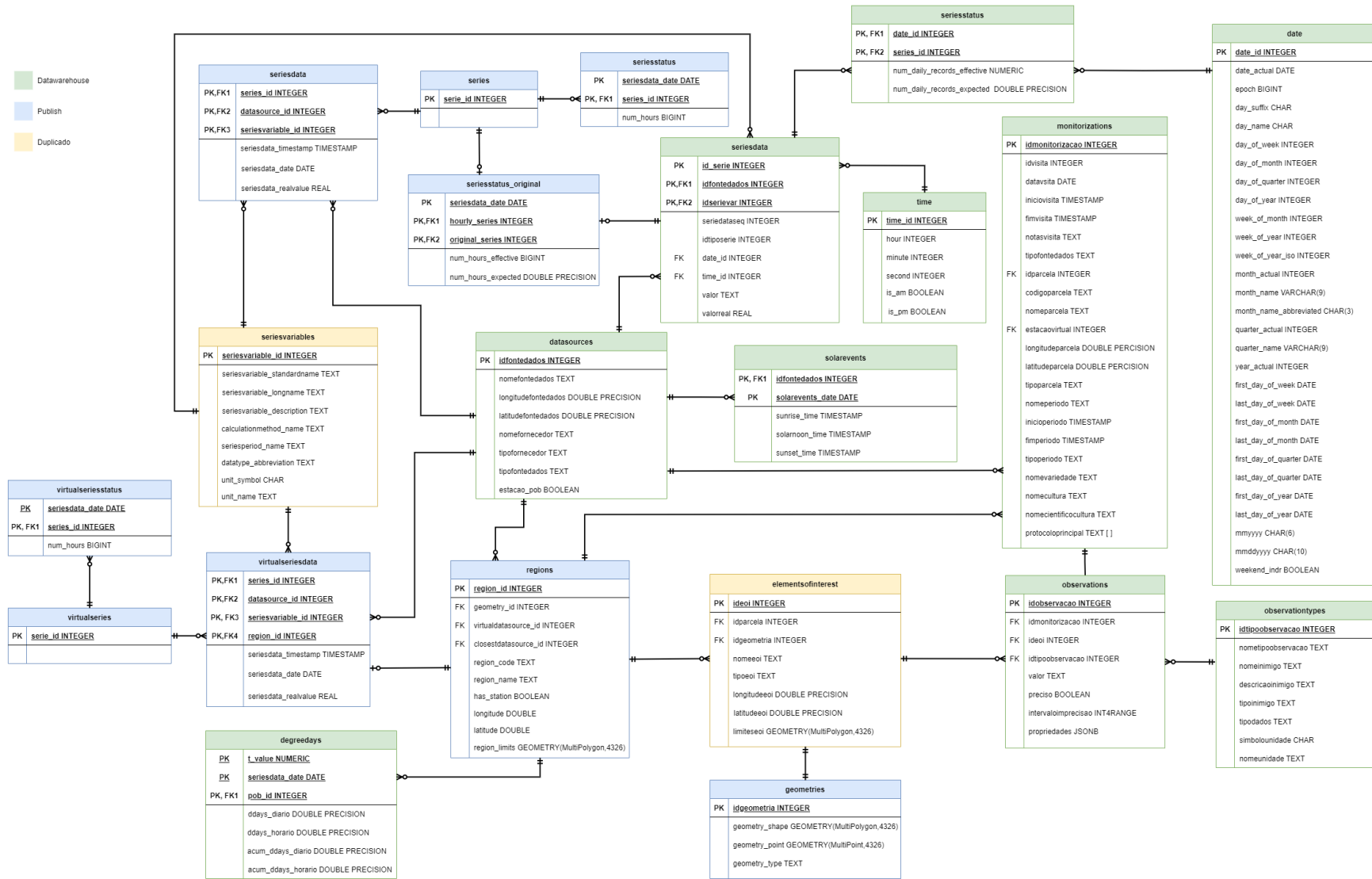


Figura 5.5: Modelo de dados da base de dados analítica, com indicação das tabelas que pertencem a cada vista materializada (*data warehouse* ou *publish*) e as que estão em ambas (duplicadas).

Do *data warehouse* destacam-se as tabelas *monitorizations* e *observations* que contêm todos os dados fitossanitários recolhidos durante as visitas de campo. Para além disso, as tabelas *datasources*, *seriesdata* e *seriesvariables* contêm toda a informação relativa aos dados meteorológicos. Por outro lado, no *publish* também se destaca também a tabela *seriesdata*, mas note-se que esta contém os dados meteorológicos processados e uniformizados horariamente, enquanto que a do *data warehouse* contém os dados originais. Para além disso, o *publish* inclui também a tabela *virtualseriesdata* que contém os dados meteorológicos das estações virtuais associadas aos *POBs* da tabela *regions*. Qualquer uma das tabelas de séries temporais (originais, processadas ou virtuais) associa-se a uma tabela *seriesstatus* que indica o estado e qualidade da estação, isto é, o número de registos recebidos ao longo de cada dia e, para a série original, o número de registos esperado. Finalmente, a tabela *degreedays* permite obter os graus-dia que ocorreram num *POB*, numa determinada data, calculados com diferentes parâmetros de t .

A criação deste segundo depósito permite assim, para além da visualização dos dados de campo e meteorológicos originais, a visualização de dados processados e uniformizados, bem como a associação ou computação *on the fly* de dados que respondam aos requisitos analíticos elicitados. Exemplos dessa aplicação são a associação de estações virtuais aos *POBs* e a computação de graus-dia (*degree days*) para diferentes parametrizações, cuja importância será detalhada no próximo capítulo.

Por fim, o volume de dados meteorológicos acumulados ao longo do projeto FitoAgro é elevado, sendo que a visualização interativa dos dados originais sofre de alguns atrasos e de alguma falta de fluidez. Contudo, a visualização interativa de dados do *publish*, isto é, de dados processados e uniformizados horariamente bem como de dados de estações virtuais, atenuou esse problema, reafirmando a importância da criação deste tipo de depósitos de dados para a construção deste tipo de componentes analíticas.

METEOROLOGIA - DESENHO E IMPLEMENTAÇÃO

Neste capítulo é apresentado o desenho e implementação de visualizações como resposta aos requisitos analíticos levantados no capítulo 4 para a meteorologia. Deste modo, serão apresentadas visualizações para a monitorização e controlo de qualidade das estações e sensores meteorológicos (secção 6.1) e para a analítica dos dados meteorológicos obtidos através desses sensores (secção 6.2).

6.1 Monitorização do Estado e Qualidade das Estações

Através da monitorização da ocorrência/ausência de registos nas diferentes estações é possível identificar quais as que apresentam problemas ou estão atrasadas. Deste modo propôs-se a criação de indicadores de estado com base no último registo e mapeados para um esquema de cores, apresentado de seguida. Note-se que os dados são carregados para o sistema apenas ao final do dia, pelo que o dia de referência é sempre o dia anterior ao corrente. Todos os indicadores apresentados nesta secção foram definidos pela equipa de *computer science* do projeto FitoAgro juntamente com o gestor do mesmo.

- **Online/Ativa** – O último registo ocorreu no dia anterior.
- **Atrasada** – O último registo ocorreu há 2 ou 3 dias.
- **Com Problemas** – O último registo ocorreu entre 4 e 10 dias atrás.
- **Offline/Inativa** – O último registo ocorreu há mais de 10 dias.

A visualização da figura 6.1 apresenta a representação geográfica das estações meteorológicas do projeto que estiveram ativas no último ano, cujos pontos têm a cor mapeada para as dos indicadores de estado apresentados anteriormente. Além disso, a forma dos pontos varia com a indicação de a estação estar ou não localizada num **POB**. Finalmente,

através das caixas de pesquisa no lado esquerdo da visualização é possível procurar e localizar uma estação específica ou destacar todas as estações de um determinado fornecedor.

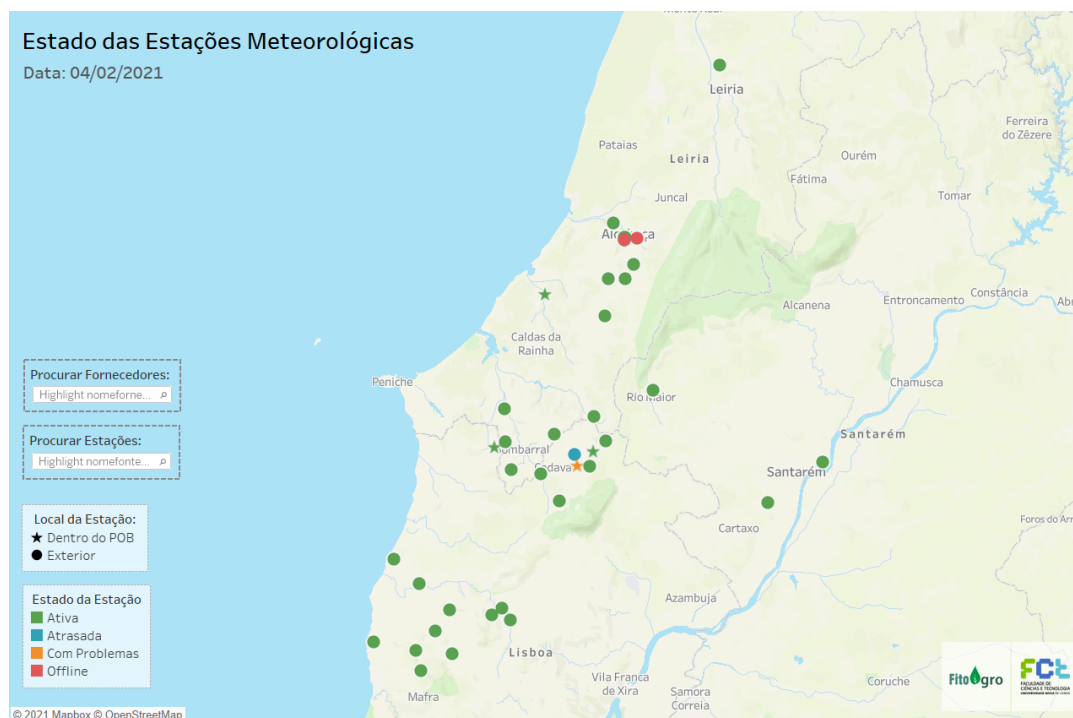


Figura 6.1: Visualização que apresenta o estado e tipo das estações meteorológicas para 4 de fevereiro de 2021.

Cada estação meteorológica tem um conjunto de sensores, dos quais a temperatura do ar, a humidade relativa, a precipitação acumulada, o ponto de orvalho e a velocidade do vento são fundamentais. Para cada um desses sensores, definiu-se o seu estado como o número de valores registados pelo sensor sobre o total de valores esperados pelo mesmo ao longo de um dia. Note-se que o número de valores esperados pelos sensores varia entre estações, uma vez que depende da frequência de amostragem original de cada estação, sem o processamento de dados referido em 3.2.2. Propôs-se, assim, o cálculo do estado diário de cada sensor com base nos seguintes indicadores:

- **Nominal** – O nº diário de valores registados pelo sensor é igual ao nº esperado.
- **Com Falhas** – O nº diário de valores registados pelo sensor é inferior ao nº esperado.
- **Avariado** – Não foram registados valores pelo sensor naquele dia.

Para visualizar o estado diário dos sensores de uma estação nos últimos 30 dias, adicionou-se uma visualização à *tooltip* dos pontos da figura anterior. Na figura 6.2 apresenta-se um exemplo desta *tooltip*, que contém alguma informação adicional relativamente à estação selecionada, bem como uma visualização do estado diário dos sensores dessa estação nos últimos 30 dias, calculados com base nos indicadores apresentados anteriormente. Note-se ainda que essa *tooltip* indica a qualidade média diária da estação ao

no último ano, cujo cálculo diário é explicado mais à frente. Assim, com este conjunto de visualizações (mapa e *tooltips*) é possível responder ao requisito analítico MCQ R1, uma vez que permite a deteção de avarias e falhas nas estações e nos seus respetivos sensores.

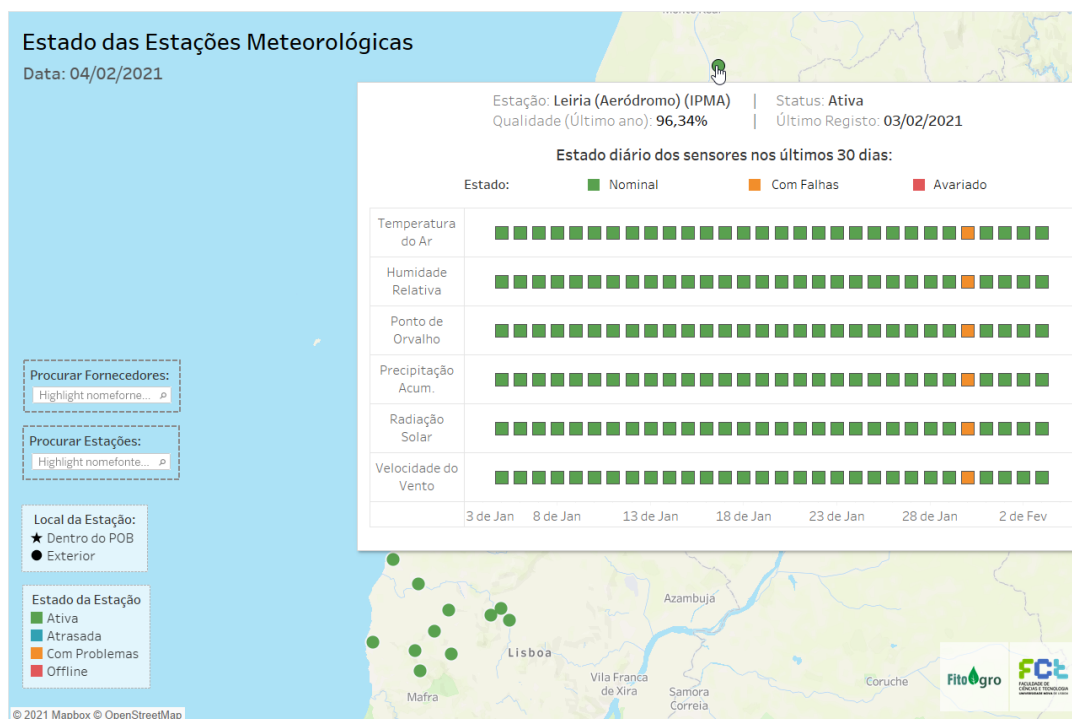


Figura 6.2: *Tooltip* da visualização na figura 6.1 com informação acerca da estação Leiria (Aeródromo) do IPMA e do estado diário dos seus sensores nos últimos 30 dias, referentes a 4 de fevereiro de 2021.

Para demonstrar a aplicabilidade da linguagem de modelação IVML, descrita no capítulo 5, apresenta-se na figura 6.3 o diagrama de modelação do *dashboard* acima. Note-se que se representou a modelação da *tooltip* no mesmo diagrama, dada a sua simplicidade.

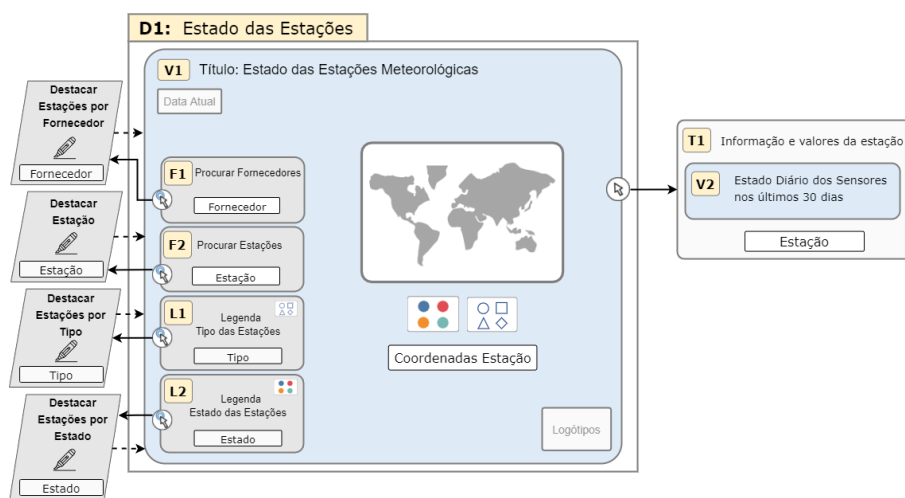


Figura 6.3: Diagrama de modelação IVML que representa o *dashboard* da figura 6.1

Para além do estado atual das estações e sensores necessitou-se de monitorizar a sua qualidade para identificar estações e sensores fiáveis *versus* propícios a falhas, como imposto pelo requisito analítico MCQ R2. Para calcular esse indicador de qualidade, agregou-se a informação ao nível da estação e assumiu-se como qualidade diária o rácio entre o número total diário de valores registados e o número total diário de valores esperados, independentemente do sensor. Com base nesse cálculo propôs-se os seguintes indicadores para representar a qualidade diária de uma estação:

- **Alta** – A qualidade diária dos dados é superior ou igual a 90%.
- **Boa** – A qualidade diária dos dados está entre 60% e 90%.
- **Média** – A qualidade diária dos dados está entre 30% e 60%.
- **Baixa** – A qualidade diária dos dados é inferior a 30%.

Na visualização da figura 6.4 é apresentado, com base nesses indicadores, a qualidade diária nos últimos 30 dias (referente a 8 de fevereiro de 2021) para cada estação do fornecedor selecionado. Ademais, a *tooltip* desta visualização apresenta a estação e o dia selecionado, bem como a qualidade diária da estação e de cada um dos seus sensores fundamentais. Com esta visualização, é possível, assim, responder ao requisito analítico MCQ R2, permitindo a monitorização da qualidade das estações e respetivos sensores.

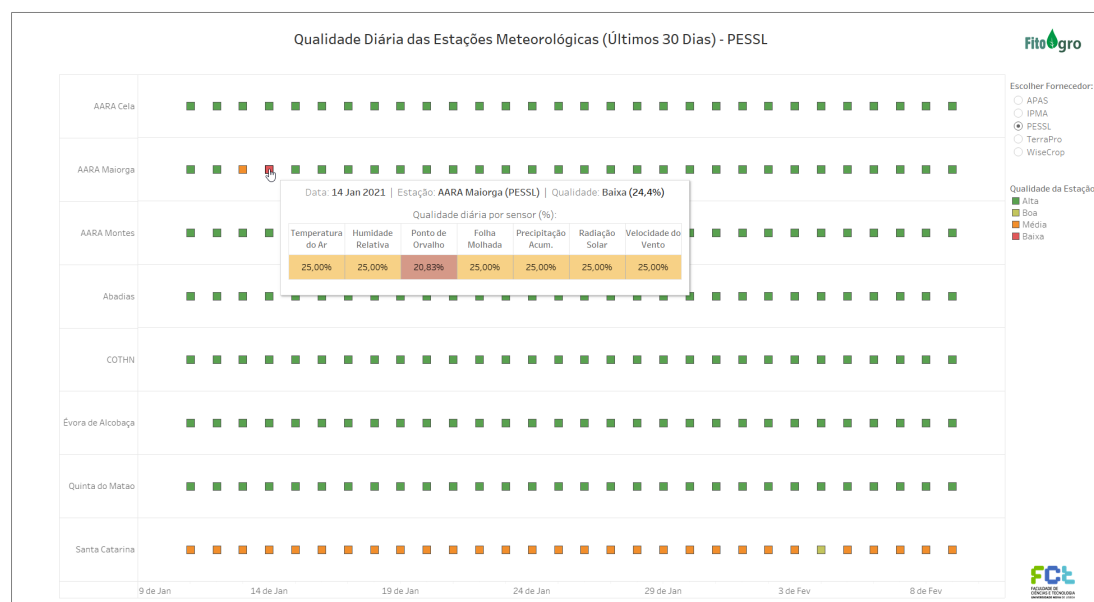


Figura 6.4: Visualização que apresenta a qualidade diária das estações nos últimos 30 dias (para 8 de fevereiro de 2021), com uma *tooltip* visível.

O conjunto das visualizações apresentadas, juntamente com as suas *tooltips*, foi desenhado e implementado através do fluxo iterativo de avaliação apresentado no capítulo 5, tendo no final sido aprovado por satisfazer as necessidades analíticas dos seus *stakeholders*. A utilização das *tooltips* para apresentar visualizações relativamente ao estado

e qualidade dos sensores de cada estação permite, simultaneamente, responder a esses requisitos analíticos e manter, respetivamente através *dashboards* das figuras 6.2 e 6.4, uma percepção espacial e temporal dessas estações.

6.2 Analítica dos Dados Meteorológicos

Para além de monitorizar e controlar a qualidade das estações e dos seus sensores, é importante no âmbito fitossanitário analisar a evolução dos dados meteorológicos. Estes dados podem ser obtidos diretamente dos valores registados pelos sensores (secção 6.2.1), ou através de cálculo de variáveis derivadas (secção 6.2.2).

6.2.1 Sensores Meteorológicos

Uma vez que as estações meteorológicas têm muitos sensores, selecionou-se para visualização apenas aqueles com maior relevância para projetos de monitorização fitossanitária, como referido e enumerado em 3.2.2. Estes foram mapeados para diferentes cores para serem facilmente distinguíveis entre si: temperatura do ar, humidade relativa, precipitação acumulada, ponto de orvalho, folha molhada, velocidade do vento e radiação solar. Ademais, considerou-se relevante calcular e adicionar a depressão do ponto de orvalho, embora não seja fornecido pelas estações.

A seleção destas cores obedeceu a dois critérios, sendo o primeiro a distinção visual clara entre as cores e o segundo a utilização, quando possível, de cores que estejam em consenso com a intuição humana relativamente a cada sensor. Por exemplo, para os sensores folha molhada e radiação solar, respetivamente, associa-se intuitivamente as cores verde e amarelo, enquanto que para a depressão do ponto de orvalho pode não se associar o castanho, mas é uma cor que garante a distinção das restantes. A utilização destas cores foi testada e aprovada pelos *stakeholders* através do fluxo iterativo de avaliação descrito no capítulo 5.

Como já foi referido anteriormente, a frequência de amostragem dos dados originais dos sensores varia consoante a sua estação ou fornecedor. Por um lado, a visualização desses dados permite monitorizar e controlar a qualidade dos dados em bruto, mas, por outro lado, não garantem a consistência necessária para fins analíticos. Por esse motivo, realizou-se, fora do âmbito deste trabalho, um processamento dos dados para os uniformizar numa frequência de amostragem horária [50].

Propôs-se um conjunto de 5 *dashboards* como solução para os requisitos analíticos de monitorização e análíticas de dados provenientes dos sensores meteorológicos, sendo um para monitorização dos dados originais e quatro para fins analíticos. Na figura 6.5 apresenta-se um diagrama destes *dashboards*, sendo que o identificado como D1 é aquele que utiliza os dados em bruto para apresentar os valores originais dos sensores ao longo de um dia. Por outro lado, o D2 utiliza os dados uniformizados para apresentar valores agregados (diariamente, semanalmente ou mensalmente) ao longo de um intervalo de

tempo (1 ano, 6 meses, 3 meses, 1 mês ou 7 dias), ambos selecionáveis. A *tooltip* do D2 apresenta os valores horários do sensor e ao longo do dia selecionado, sendo que através do menu da *tooltip* é possível navegar para o *dashboard* D3. Por sua vez, o D3 é estruturalmente semelhante ao D1, mas utiliza os dados processados e, por isso, apresenta os valores horariamente ao longo do dia selecionado. Tanto no D1 como D3 é possível navegar para os dias seguinte e anterior, sendo que no D3 também é possível navegar de volta para o D2. Note-se que inicialmente construiu-se um *dashboard* semelhante ao D2 utilizando os dados originais, mas posteriormente considerou-se que o D1 seria suficiente para satisfazer os requisitos de monitorização elicitados. Finalmente, os *dashboards* D4 e D5 foram replicados respetivamente do D2 e D3, mas nestes últimos em vez de selecionar uma estação seleciona-se um **POB**. Os dados meteorológicos associados a cada **POB** provêm da estação fiável mais próxima do mesmo, como será explicado posteriormente nesta secção.

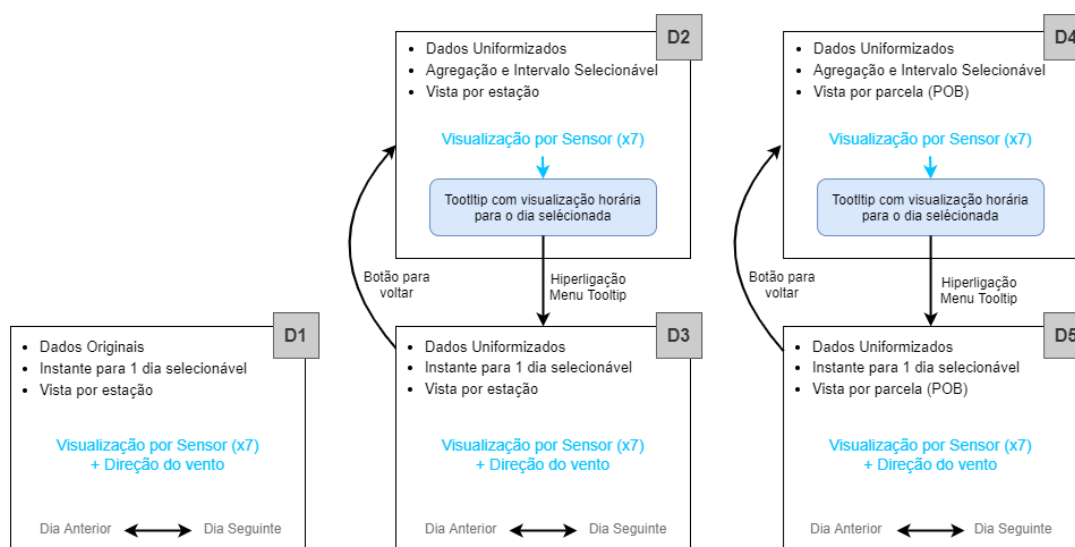


Figura 6.5: Diagrama que apresenta um resumo de todos os *dashboards* propostos para a monitorização e analítica dos dados provenientes de sensores meteorológicos, bem como formas de navegação que existam entre os mesmos.

O *dashboard* da figura 6.6, correspondente ao D1 do diagrama anterior, apresenta os valores originais medidos em cada instante pelos sensores ao longo de um determinado dia, de forma a obter uma visão holística da evolução dos mesmos. Para isso, foi necessário alinhar os seu eixos temporais verticalmente, sendo apenas representado uma vez, em baixo, por uma questão de otimização de espaço e minimização de redundância. Note-se que a estação apresentada (Apas05Do), tem uma frequência de amostragem de 15 minutos, mas que esta pode diferir entre estação, como foi explicado anteriormente. Uma vez que este *dashboard*, utilizando os dados em bruto, permite a monitorização e análise dos valores originais registados por cada sensor, para a estação e dia selecionados, responde ao requisito analítico **MCQ R3**.

6.2. ANALÍTICA DOS DADOS METEOROLÓGICOS

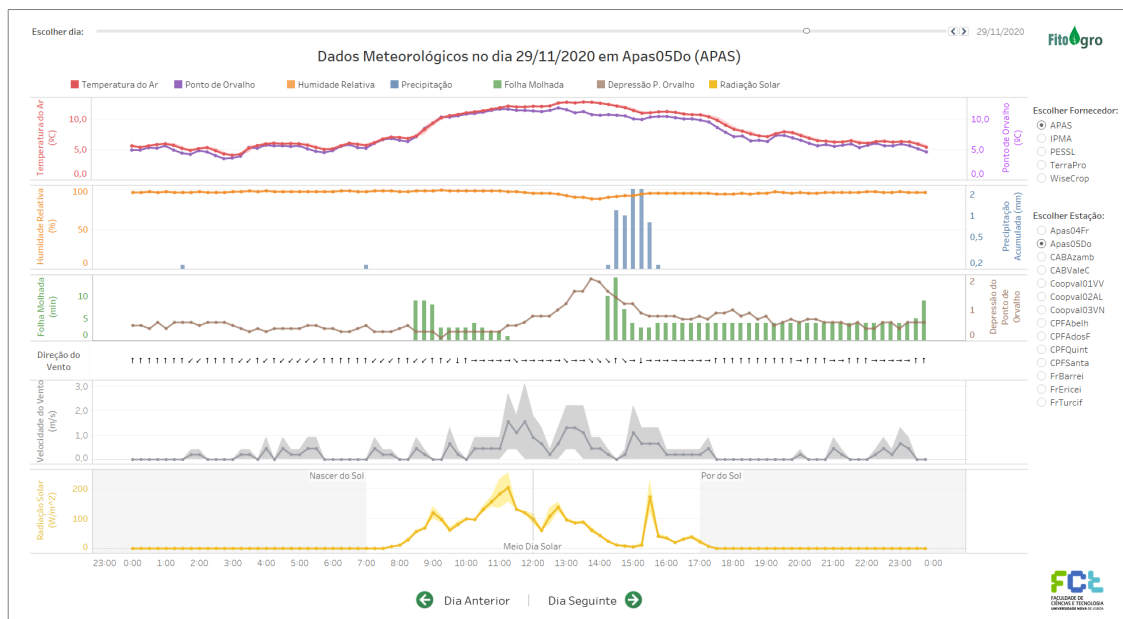


Figura 6.6: *Dashboard* que apresenta os dados meteorológicos em bruto provenientes da estação Apas05Do, ao longo do dia 30 de novembro de 2020.

Ao observar o *dashboard* acima é perceptível que foram utilizadas diferentes abordagens visuais para os diferentes sensores, dependendo do tipo de medição recolhida pelos mesmos. Estas medições podem ser classificadas em 3 grupos, nomeadamente medidas num instante, com mínimo e máximo associado e quantidades acumuladas num período. Cada um destes tipos é representado visualmente de forma diferente, nomeadamente por linhas, áreas e barras, respetivamente. Note-se que, como já foi referido anteriormente, a ausência de dados é facilmente detetada pela quebra de sinal e ausência de *tooltip*. De seguida descreve-se mais detalhadamente a abordagem utilizada para cada sensor, sendo que para melhor exemplificar, agregou-se os valores dos sensores diariamente para um período de 3 meses.

Para o sensor da **temperatura do ar**, representou-se a temperatura média através de um gráfico de linhas e construiu-se uma banda para a variação entre a temperatura máxima e mínima, caso exista. Uma vez que o **ponto de orvalho** é associável à temperatura, adicionou-se este sensor ao mesmo gráfico (figura 6.7) e com a mesma escala, sob a forma de linha.

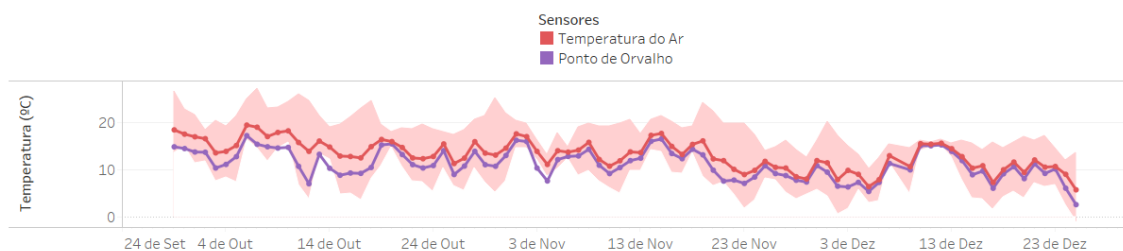


Figura 6.7: Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a temperatura do ar e ponto de orvalho na estação Apas05Do.

Similarmente, representou-se o sensor da **humidade relativa**, através de um gráfico de linhas para a média e de uma banda para a variação entre o máximo e mínimo, caso exista (figura 6.8).

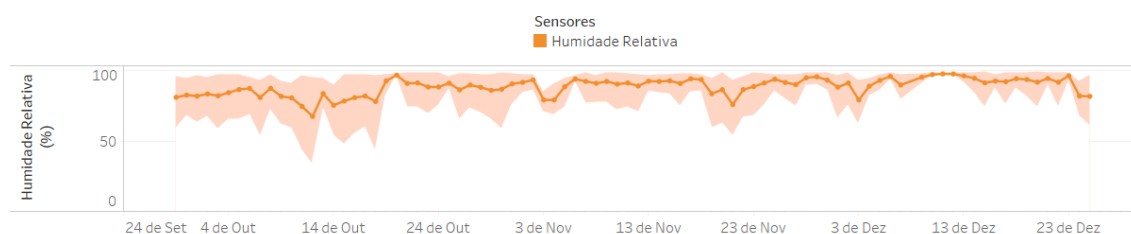


Figura 6.8: Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a humidade relativa na estação Apas05Do.

Para o sensor de **folha molhada**, representou-se a soma de minutos de folha molhada através de um gráfico de barras. Para além disso, calculou-se a **depressão do ponto de orvalho**, através dos valores de temperatura média e de ponto de orvalho, e representou-se no mesmo gráfico da folha molhada (figura 6.9), sob a forma de linha. Como estes sensores têm diferentes unidades utilizou-se duas escalas diferentes, uma para cada sensor.

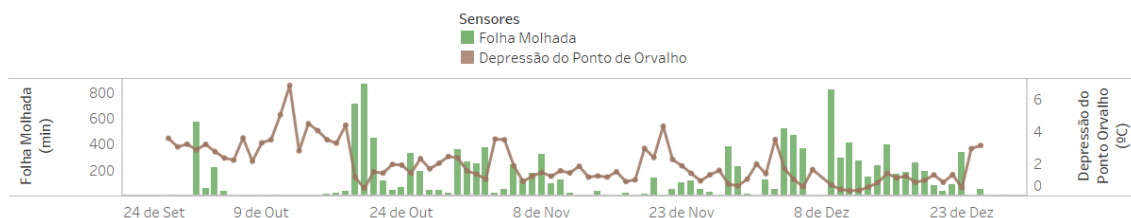
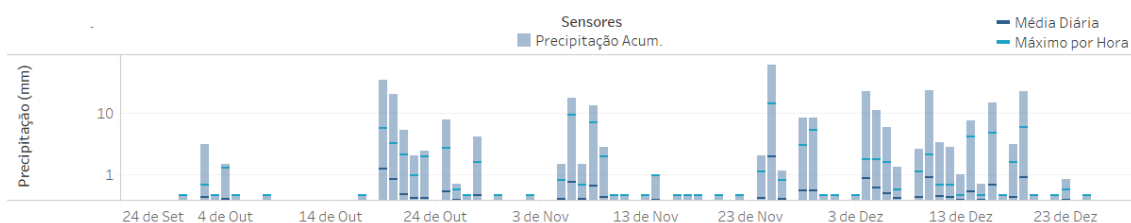


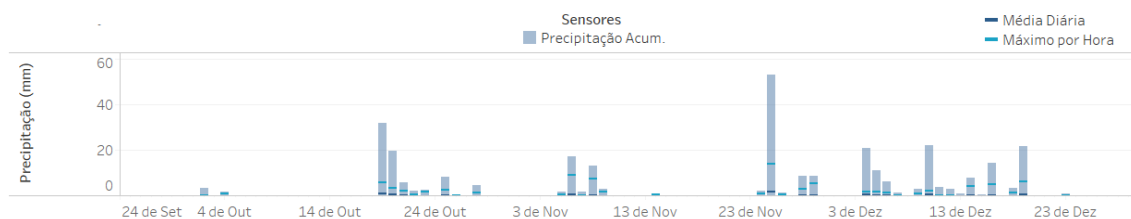
Figura 6.9: Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a folha molhada, na estação Apas05Do, e o cálculo da depressão do ponto de orvalho.

A representação dos valores de **precipitação acumulada** foi efetuada através de um gráfico de barras, como é habitual em visualizações meteorológicas. No entanto, a acumulação diária por si só não permite compreender a distribuição da precipitação ao longo do dia, pelo que se calculou a precipitação horária máxima e a média da precipitação horária. Estes valores permitem assim analisar se a precipitação acumulada registada ocorreu ao longo do dia (máximo e média próximas) ou se ocorreu apenas em pequenos períodos do dia (máximo e média distantes). A representação desses valores na visualização foi realizada através de marcas retangulares sobre a barra de precipitação acumulada, com cores distintas. Por fim, considerou-se que a utilização de uma escala logarítmica (figura 6.10a) se adequava melhor aos valores de precipitação acumulada do que uma escala linear (figura 6.10b).

6.2. ANALÍTICA DOS DADOS METEOROLÓGICOS



(a) Escala logarítmica



(b) Escala linear

Figura 6.10: Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a precipitação acumulada, na estação Apas05Do, e o cálculo do máximo por hora e da média do dia, numa escala logarítmica *versus* linear.

Finalmente, tanto para o gráfico do sensor da velocidade do vento (figura 6.11) como da radiação solar (figura 6.12), utilizou-se a mesma representação da temperatura do ar e humidade relativa, isto é, um gráfico de linhas para a média e uma banda para a variação entre o máximo e mínimo, caso exista.

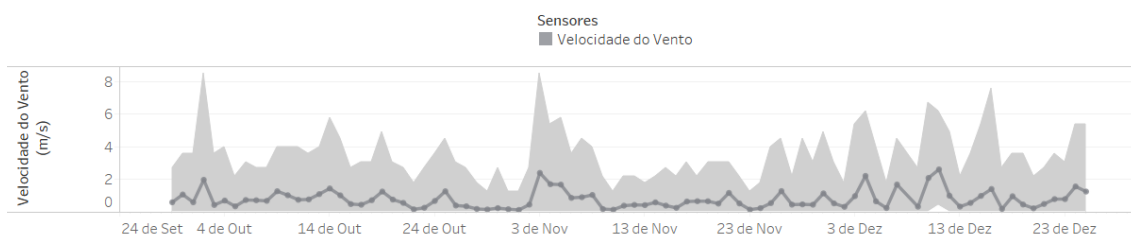


Figura 6.11: Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a velocidade do vento na estação Apas05Do.

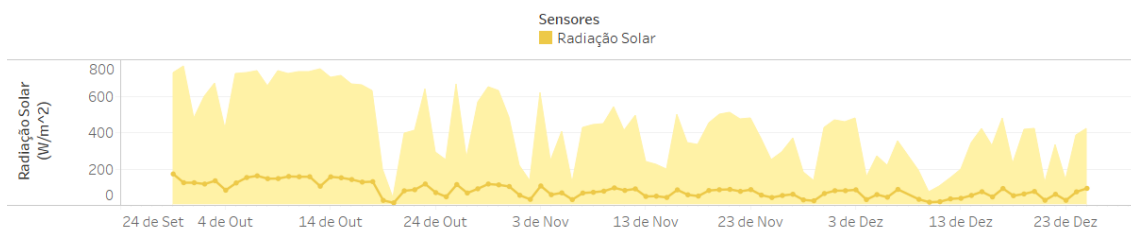


Figura 6.12: Visualização que apresenta os valores diários registados, entre setembro e dezembro de 2020, para a radiação solar na estação Apas05Do.

Utilizando estas abordagens e os dados já uniformizados, construiu-se um segundo *dashboard* (figura 6.13) que apresenta a evolução da meteorologia, de forma dinâmica e

interativa, para uma estação, intervalo de tempo (7 dias, 1 mês, 3 meses, 6 meses ou 1 ano) e granularidade (diária, semanal ou mensal) selecionáveis pelo utilizador. Note-se que a agregação para diferentes resoluções temporais é feita de acordo com a medição de cada sensor, por exemplo a agregação diária da temperatura corresponde à média da medição média ao longo do dia e aos máximos e mínimos dessas respetivas medições, enquanto que para a precipitação apenas corresponde ao acumulado do dia.

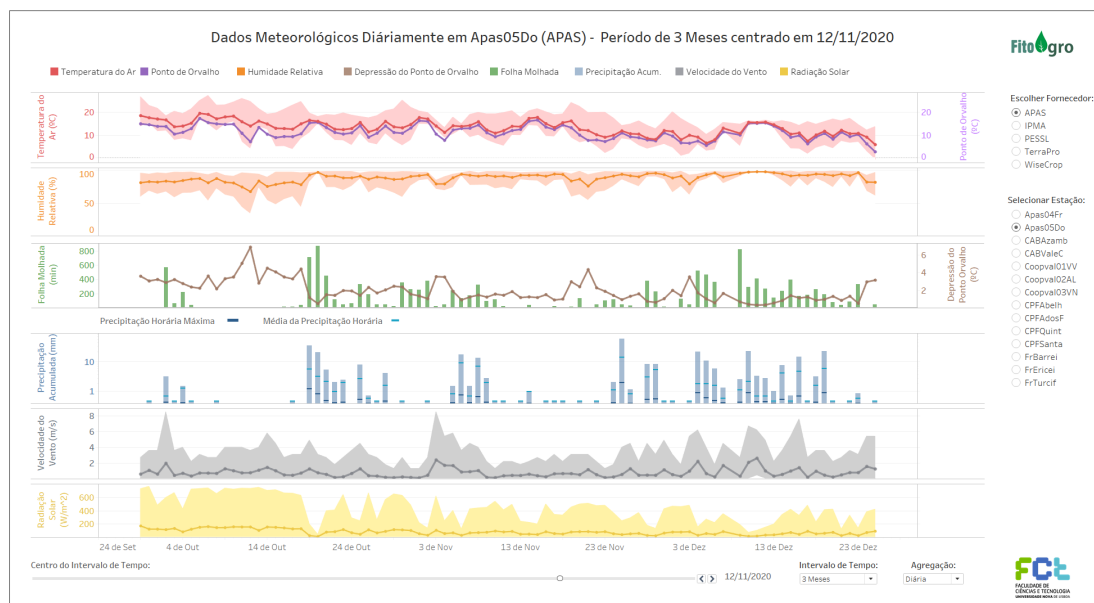


Figura 6.13: *Dashboard* que apresenta os dados meteorológicos da estação Apas05Do, ao longo de um período de 3 meses centrado em 12 de novembro de 2020.

Cada uma das visualizações no *dashboard* acima contém *tooltips* que apresentam informação relativamente aos valores agregados para o(s) sensor(es) selecionado(s) e uma visualização com a evolução horária desse(s) sensor(es) ao longo do dia selecionado (exemplo na figura 6.14). Caso a *tooltip* seja ativada através de um clique (em vez de passar o cursor), mostra-se, na parte inferior da mesma, uma hiperligação para outro *dashboard*, que permite ver os dados detalhados ao longo do dia selecionado.

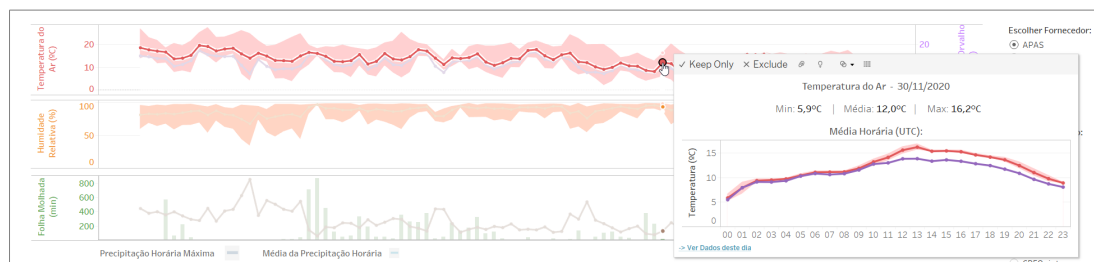


Figura 6.14: Uma *tooltip* da visualização da figura 6.13, onde se apresenta os valores agregados para o dia selecionado (29 de novembro de 2020) e uma visualização da evolução desses sensores ao longo desse dia.

6.2. ANALÍTICA DOS DADOS METEOROLÓGICOS

Por sua vez, o *dashboard* da hiperligação é semelhante ao anterior, mas apresenta a evolução horária dos sensores ao longo do dia selecionado previamente. A sua estrutura também difere ligeiramente do anterior, uma vez que se uniu os sensores de humidade relativa e precipitação acumulada (removendo o cálculo de precipitação máxima e média) numa única visualização, adicionou-se uma visualização para o sensor de direção do vento e anotou-se os pontos críticos da radiação solar (i.e. nascer do sol, meio dia solar e pôr do sol). A nível de navegação, no canto superior esquerdo deste *dashboard* existe uma seta para voltar para o *dashboard* anterior (da figura 6.13) e na parte inferior encontram-se botões para navegar para o dia anterior e seguinte.

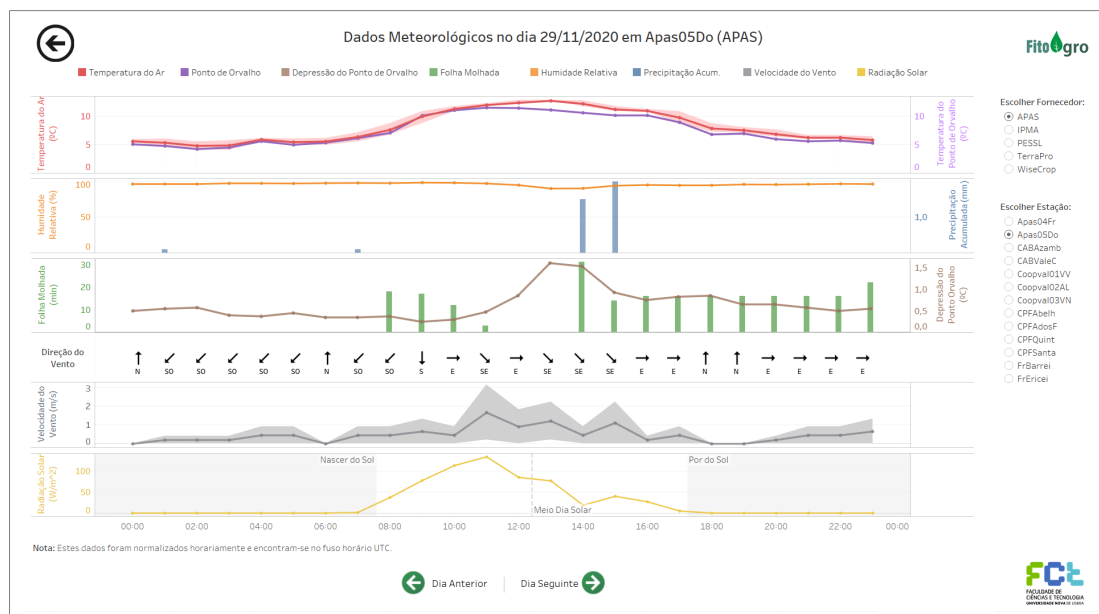


Figura 6.15: *Dashboard* que apresenta os dados meteorológicos da estação Apas05Do, horariamente ao longo do dia 29 de novembro de 2020.

O conjunto destes dois últimos *dashboards* responde ao requisito analítico **AD R1**, uma vez que permite observar a evolução dos dados uniformizados para fins analíticos. Ademais, note-se que os dados apresentados no *dashboard* da figura anterior (6.15) são os mesmos da figura 6.6, mas uniformizados para uma frequência de amostragem horária.

No contexto de projetos de monitorização fitossanitária, e especialmente na **região do Oeste**, a análise de dados meteorológicos toma maior relevância quando os eventos meteorológicos ocorrem no mesmo local dos pontos de observação biológica em estudo. No caso do projeto FitoAgro, apenas algumas das estações meteorológicas estão localizadas dentro dos **POBs** estabelecidos, como foi possível observar anteriormente na figura 6.1. Deste modo, para obter dados meteorológicos fidedignos para os restantes locais, foi necessário associar uma estação virtual a cada um deles. Os dados das estações virtuais são obtidos através das restantes estações utilizando o método IDW (*inverse distance weighting*) que tem em consideração a distância das estações na proximidade ao **POB**.

Este trabalho foi realizado no contexto da dissertação de L. Correia (*unpublished*) [50] e foi o que permitiu obter dados meteorológicos para o local (ou na proximidade) dos POBs. Utilizando estas estações pretende-se promover a descoberta de correlações entre as condições meteorológicas e as observações fitossanitárias registadas nesses locais.

Para validar e analisar os dados das estações virtuais, replicou-se os *dashboards* apresentados nas figuras 6.13 e 6.15, onde em vez da listagem de todas as estações apenas se listam os POBs. Ao selecionar um dos POBs, aparecem os dados meteorológicos da estação virtual associada ao mesmo, cumprindo, assim, o requisito analítico AD R2. Na figura 6.16 exemplifica-se essa adaptação relativamente ao *dashboard* da figura 6.13.

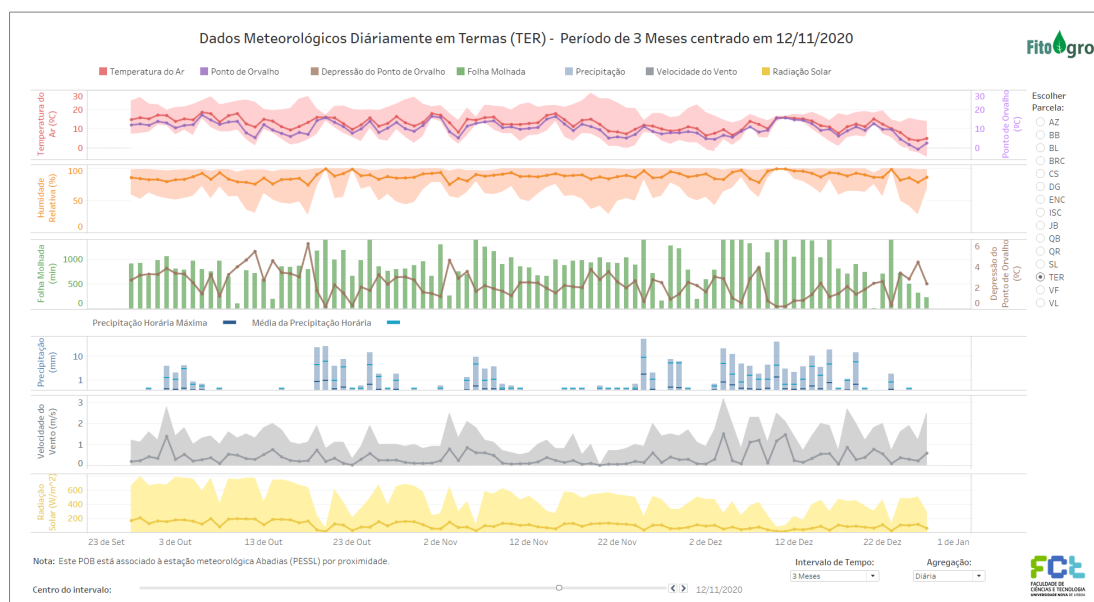


Figura 6.16: Replicação do *dashboard* da figura 6.13 que apresenta os dados meteorológicos da estação virtual do POB Termas, ao longo de um período de 3 meses centrado em 12 de novembro de 2020.

6.2.2 Variáveis Derivadas

Com a criação de estações virtuais, tornou-se possível obter dados meteorológicos relevantes para o desenvolvimento das *pragas* e culturas e calcular variáveis derivadas dos mesmos. O cálculo e análise destas variáveis permite estimar os momentos críticos no desenvolvimento da vegetação e dos insetos. No contexto do projeto, as duas variáveis em foco são as horas de frio e os graus-dia, apresentadas de seguida e cujas visualizações visam responder ao requisito analítico AD R3.

6.2.2.1 Horas de Frio

O frio tem uma função importante no crescimento da vegetação, sendo vital durante o repouso vegetativo e na quebra de dormência de várias espécies. Considera-se quebra de dormência quando as necessidades em horas de frio foram satisfeitas, sendo que o resultado traduz-se numa floração de boa qualidade, favorecendo os calibres e a qualidade final dos frutos e produções [52]. No projeto FitoAgro, as horas de frio são definidas como as horas acumuladas por dia com temperatura abaixo de 7.2°C. Utilizando esse limite para indicar a ocorrência de horas de frio, a visualização da figura 6.17 apresenta a evolução da temperatura média do ar por hora, entre outubro de 2019 e março de 2020, na estação associada ao pomar Quinta do Rato.

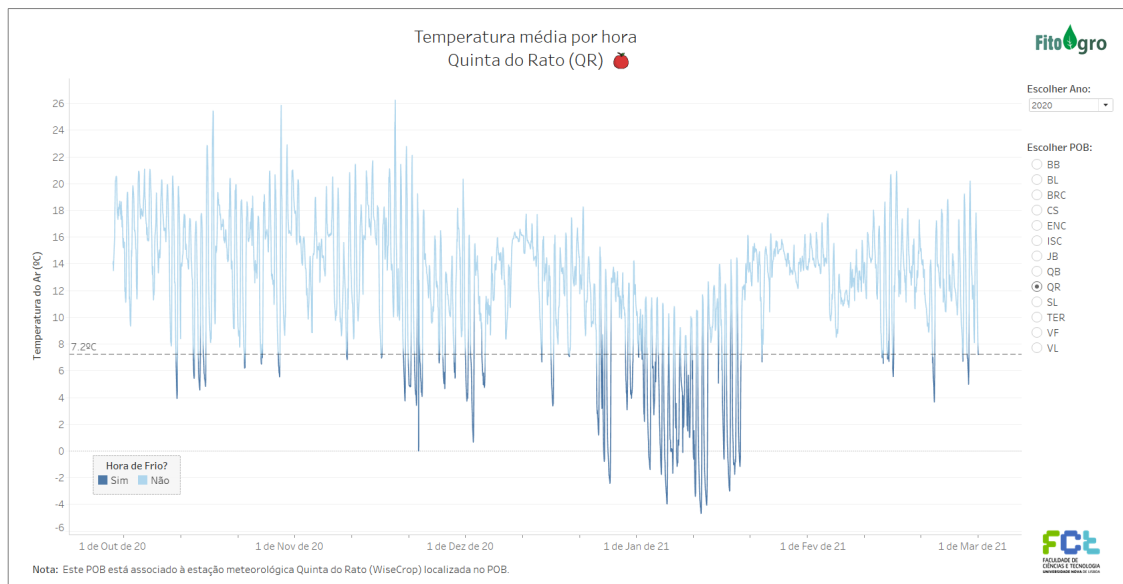


Figura 6.17: Evolução da temperatura média por hora, entre outubro de 2019 e março de 2020, no POB Quinta do Rato, com indicação de horas de frio.

O cálculo das horas de frio é expresso através da equação 6.1, onde $Temp$ corresponde à temperatura média horária do ar.

$$HF(dia) = \sum_{h=0}^{24h} Temp_h \leq 7.2 \quad (6.1)$$

No caso das pomóideas, considera-se que, para atingir uma floração de boa qualidade, deverão ser acumuladas, desde 1 de outubro até 15 de fevereiro de cada ano, 500 e 550 horas de frio para as macieiras e pereiras, respetivamente. Na figura 6.18 é possível observar as horas de frio acumuladas, em 2019/20, no pomar Quinta do Rato e com indicação do limite correspondente à cultura desse pomar (macieira).

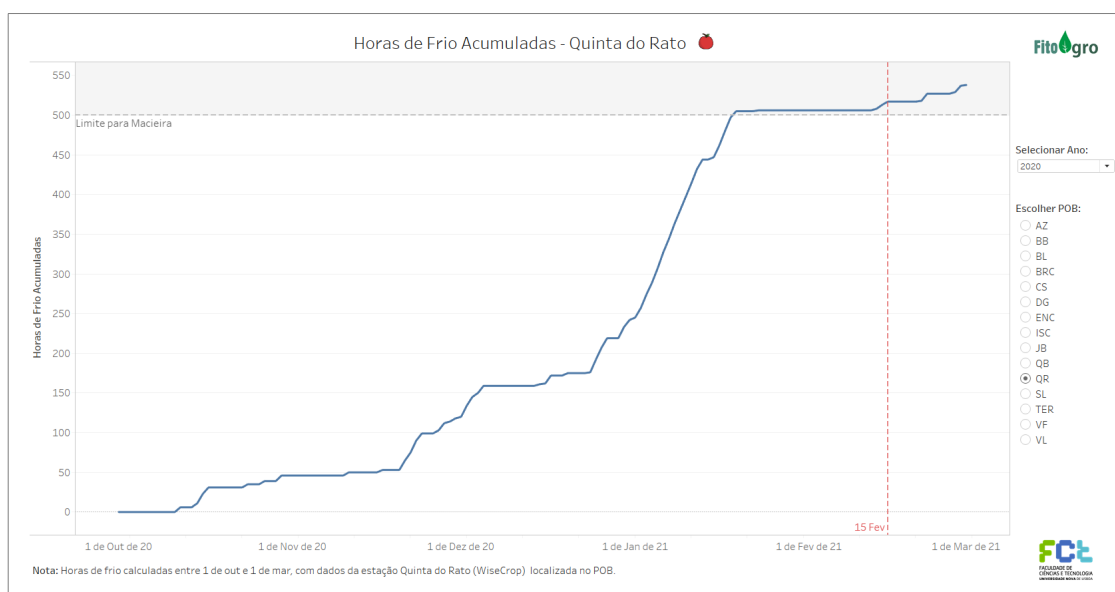


Figura 6.18: Visualização das horas de frio acumuladas, entre outubro de 2019 e março de 2020, no POB Quinta do Rato.

6.2.2.2 Graus-dia

Cada inseto tem um intervalo de temperaturas na qual se desenvolve e fora do qual não o consegue realizar. Esse intervalo é utilizado para calcular o número de graus-dia num dia específico. Um grau-dia acontece quando a temperatura média num determinado dia (24 horas) está um grau acima do limite inferior ótimo para o desenvolvimento do inseto (ou abaixo do limite superior, caso exista). A acumulação deste valor ao longo de vários dias corresponde ao graus-dia acumulados utilizados por muitos dos modelos preditivos que estimam o ciclo de vida dos insetos [53], como no caso do projeto FitoAgro.

Para exemplificar o cálculo dos graus-dia selecionou-se o bichado (*Cydia pomonella*), visto que é um inseto que já foi bastante estudado em laboratório e que, por isso, tem uma parametrização de desenvolvimento bem definida na literatura. Para esta praga, o zero de desenvolvimento (limite inferior ótimo) está estabelecido nos 10°C e os seus graus-dia calculam-se segundo a equação 6.2, onde *Temp* corresponde a uma aproximação da temperatura média diária do ar e 10 ao zero de desenvolvimento.

$$GD(dia) = \begin{cases} Temp - 10 & , \text{ se } Temp \geq 10 \\ 0 & , \text{ caso contrário} \end{cases} \quad (6.2)$$

Na figura 6.19 é possível observar o número de graus-dia que ocorreram entre janeiro e outubro de 2020 na estação do POB Quinta do Brejo, calculados através da fórmula anterior.

6.2. ANALÍTICA DOS DADOS METEOROLÓGICOS

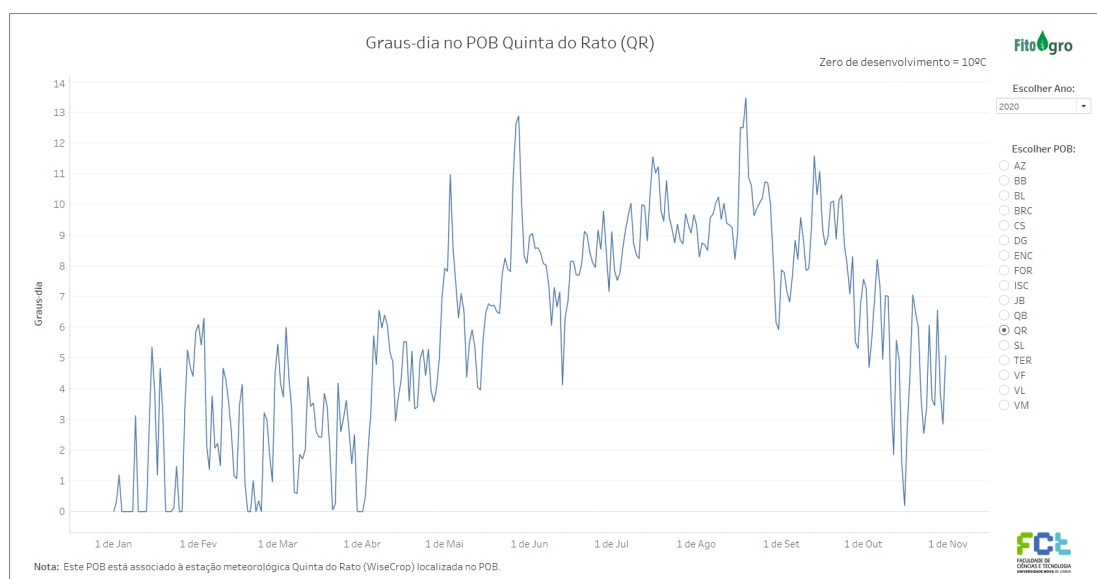


Figura 6.19: Visualização dos graus-dia, entre janeiro e outubro de 2020, no POB Quinta do Rato.

Com base no somatório das temperaturas médias diárias superiores ao zero de desenvolvimento da praga (10°C), i.e. nos graus-dia acumulados, é possível determinar a ocorrência de diferentes fases do desenvolvimento desta. Na tabela 6.1 apresenta-se a associação entre graus-dia acumulados e a ocorrência das diferentes fases de desenvolvimento do bichado (*Cydia pomonella*), na região da Beira Interior de Portugal.

Fase de desenvolvimento da praga	Valores validados para a Beira interior
Início de voo	90°C
Início das penetrações	200°C
Pico de voo da 1ª geração	340°C
Primeiras larvas a saírem dos frutos	470°C
Fim da 1ª geração, início da 2ª	700°C

Tabela 6.1: Valores do somatório de temperaturas médias diárias determinadas para as várias fases de desenvolvimento do bichado (*Cydia pomonella*), na região da Beira Interior de Portugal [54].

Através desta informação e consoante o número de graus-dia acumulados em cada local e em cada ano, é possível determinar em que momentos ocorrem as diferentes fases de desenvolvimento da praga. Por exemplo, na figura 6.20 apresenta-se os graus-dia acumulados (utilizando a parametrização do bichado) entre janeiro e outubro de 2020 no POB da Quinta do Brejo. Ao anotar as diferentes fases de desenvolvimento do bichado na visualização consegue-se estimar o momento em que estas ocorreram.

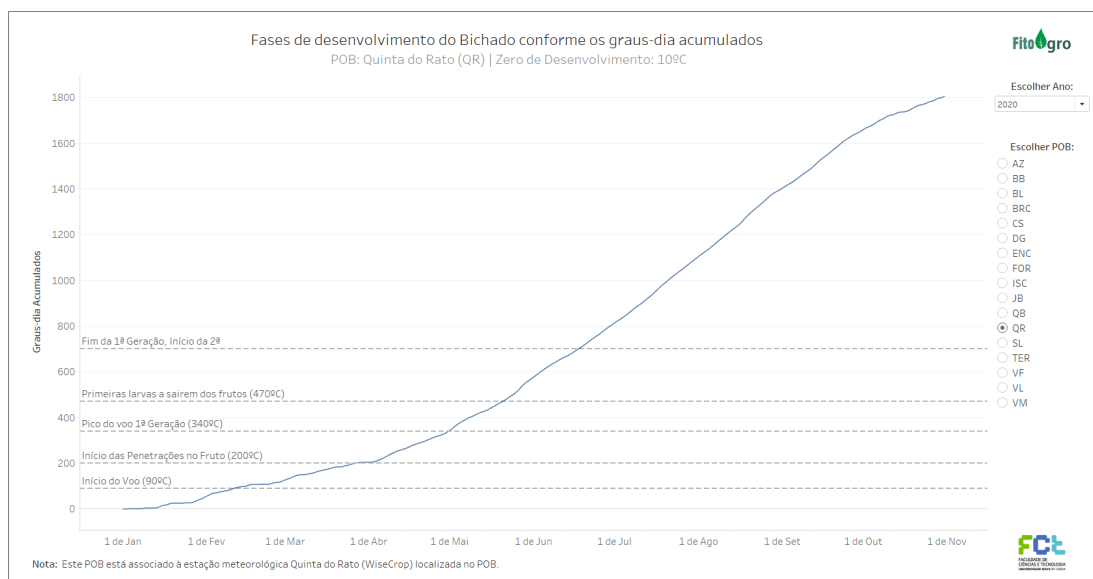


Figura 6.20: Visualização das diferentes fases de desenvolvimento do bichado conforme os graus-dia acumulados, entre janeiro e outubro de 2020, no POB Quinta do Rato.

Uma vez que o cálculo dos graus-dia depende diretamente do intervalo de desenvolvimento ótimo estabelecido para cada praga, este apenas é possível quando o intervalo é conhecido, o que não acontece para todos os inimigos. Deste modo, o ideal seria ter modelos para o cálculo de graus-dia que se ajustem às observações fitossanitárias registadas para cada praga, estimando esse(s) limite(s). Este tópico será discutido posteriormente em detalhe na secção 7.3.2 do próximo capítulo.

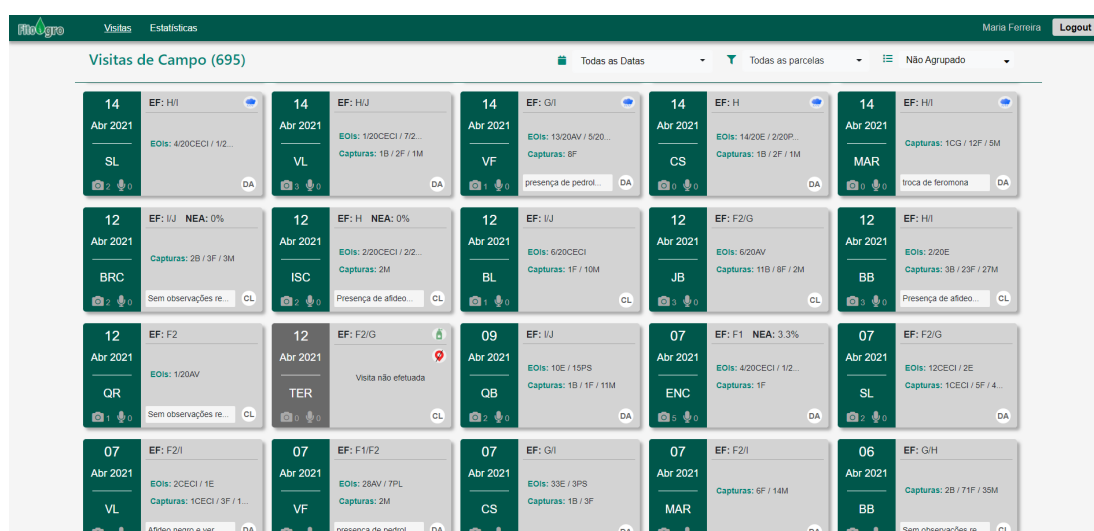
FITOSSANIDADE - DESENHO E IMPLEMENTAÇÃO

Neste capítulo é apresentado o desenho e implementação de visualizações como resposta aos requisitos analíticos levantados no capítulo 4. Deste modo, serão apresentadas visualizações para a monitorização e controlo de qualidade da recolha de dados de campo (secção 7.1), para a analítica dos dados fitossanitários (secção 7.2) e para o estudo do impacto e correlações da meteorologia na fitossanidade (secção 7.3).

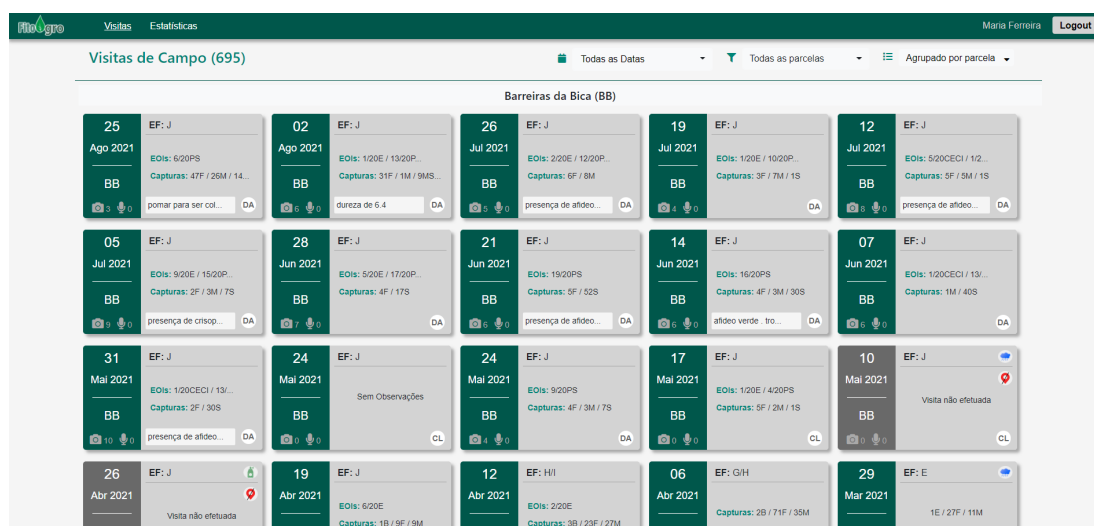
7.1 Monitorização e Controlo das Visitas de Campo

A recolha de dados fitossanitários pelos técnicos de campo, como já havia sido referido, deixou de ser efetuada em papel e em *Google sheets* e passou a realizar-se através de uma aplicação móvel desenvolvida por M. Correia (2020) para esse fim [44]. Com esta mudança, os dados deixaram de poder ser visualizados na *Google Drive* pelo que surgiu a necessidade de fornecer outra forma de acesso aos mesmos, como referido pelo requisito analítico **MCQ R4**. Para isso, desenvolveu-se um *website* complementar à aplicação de recolha que permite visualizar todos os dados de campo observados durante as visitas, cuja recolha tenha sido efetuada através da aplicação. A implementação e desenvolvimento deste *website* é descrito em pormenor por M. Correia (2020) [44], que apresenta a abordagem utilizada para a sua construção e para a leitura dos dados registados através da aplicação móvel. De seguida descreve-se brevemente as principais páginas deste *website*, indicando aquelas que contêm embutidas visualizações interativas de dados.

O acesso ao *website* é efetuado através das mesmas credenciais da *Google Drive* utilizadas para aceder às folhas de registo antigas, ou seja, apenas os utilizadores com acesso a essa pasta privada na *Google Drive* do projeto FitoAgro é que têm acesso a este *website*. Após fazer *login*, a página inicial apresenta uma listagem de todas as visitas realizadas através da aplicação, como na figura 7.1a. Esta listagem está ordenada cronologicamente, da mais recente para a mais antiga, sendo, no entanto, possível filtrar as visitas por datas e/ou parcelas (POBs), bem como agrupar as mesmas por parcela, o permite perceber a evolução entre visitas ao longo de um determinado intervalo de tempo ou dentro de cada POB (figura 7.1b), respetivamente.



(a) Lista de todas as visitas de campo, ordenadas da mais recente para a mais antiga.



(b) Lista de visitas, ordenadas da mais recente para a mais antiga, agrupadas por POB.

Figura 7.2: Página inicial (após *login*) do *website* complementar à aplicação de recolha de dados fitossanitários para o projeto FitoAgro, com e sem agrupar por parcela.

Nessa listagem, resumiu-se a informação de cada visita através de uma etiqueta, cuja cor do lado esquerdo varia entre verde e cinzento consoante a visita tenha sido realizada com sucesso ou não, respetivamente. Desse mesmo lado da etiqueta apresenta-se a data, o código do POB e o número de fotos e anotações vocais associadas à visita. Do lado direito apresenta-se, por sua vez, o estado fenológico (EF), o NEA da filoxera caso exista, um ícone para a ocorrência de eventos (chuva ou tratamentos), um ícone no caso de ausência de traçado GPS, um resumo do número de EOIs (i.e. árvores, cintas, entre outros) atacados e do número de capturas por inimigo e, no inferior da etiqueta, notas gerais da visita seguidas de um ícone com as iniciais do técnico que realizou a visita (ou um indicador de visita cooperativa).

Em vez de meramente listar as visitas, a utilização deste conjunto de elementos visuais permite desde já perceber os aspetos chave de cada visita e comparar com os das restantes. Por exemplo, ao observar a figura 7.1a é perceptível que durante a visita do dia 12 de abril foram realizados tratamentos à parcela Termas (TER), pelo que não foram realizadas observações nem existe traço GPS, apenas se registou o estado fenológico F2/G. Para além disso, é possível observar que choveu no dia 14 de abril durante as visitas a três POBs, no entanto não foi um impedimento para a sua realização. Por outro lado, facilmente se percebe a evolução entre visitas locais ao agrupar as visitas por parcela, como por exemplo na figura 7.1b observa-se um aumento de 15 para 19 EOIs atacados por Psila (PS) no POB Barrerías da Bica, entre 14 e 21 de junho de 2021.

Ao selecionar uma visita, somos redirecionados para a página da figura 7.4, onde se pode ver em detalhe toda a informação da visita, ficheiros multimédia (imagens e anotações vocais) e os dados de recolha observados nos EOIs e armadilhas, em formato de visualização ou tabular. Para além disso, no lado direito da página existe um mapa com o traçado GPS do percurso efetuado pelo(s) técnico(s) (caso exista), sendo que quando este(s) tira(m) fotografias ou faz(em) anotações vocais estas aparecem georreferenciadas sob o traçado através de um marcador com o ícone do tipo de multimédia associado. Note-se que, através dessa mesma página, é possível navegar para a visita anterior e seguinte realizada ao POB, promovendo a análise do histórico dos dados registados nesse local.

Para além do formato tabular dos dados (no separador "Dados em Bruto") com uma estrutura similar às *Google Sheets* antigas, introduziu-se uma visualização interativa de dados, como a da figura 7.3c, para a observação e análise dos dados de uma visita. Esta visualização aparece no separador "Visualização" e a forma como foi construída requer que seja observada exclusivamente no contexto do *website*. A visualização apresenta, para cada inimigo, as observações fitossanitárias registadas em cada EOI e o número de capturas em armadilhas. Para além de fornecer a mesma informação do formato tabular, esta visualização adiciona elementos analíticos como valores médios de observação e a tendência de capturas relativamente às visitas anteriores, cuja *tooltip* apresenta em detalhe esse histórico. Esta visualização tem como objetivo facilitar a análise e monitorização das observações e capturas registadas e responder, assim, ao requisito analítico MCQ R4, descrito no capítulo 4.

CAPÍTULO 7. FITOSSANIDADE - DESENHO E IMPLEMENTAÇÃO

Visitas Estatísticas Maria Ferreira Logout

31 Mar São Lourenço (SL) - 07 Abr 14 Abr

Início: 07 Abr às 12:04 Técnico: Dep. Técnico APAS

EF: F2/G Calibre: Não registado

Chuva: Não Tratamento: Não

Notas: Sem notas.

Multimédia: 2

Visualização Dados em bruto

Notas Sem notas

EOIs

(a) Informação geral acerca da visita, multimédia e mapa com o percurso do técnico.

Visualização Dados em bruto Maria Ferreira Logout

Notas Sem notas

EOIs

Protocolo	Observação	Árvore nº.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Cecidíoma	Presença de lagarta	Branca	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		Amarela	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Laranja	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Rebentos atacados	---	0	0	0	0	1	0	0	0	0	1	0	0	0	2	1	2	0	3	1	1	
Esterifilose	Frutos contíguos atacados	---	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		Folhas atacadas (1 ramo)	---	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0

Capturas

Bichado	Cecidíoma Caixa	Cecidíoma Placa	Cochonilha Algodão	Cryptoblabes gnidiella	Funebrana	Hoplocampa	Molesta	Mosca Alimentar	Mosca Decis trap	Sésia
0	0	1	0	0	5	0	4	0	0	0

(b) Separador "Dados em bruto" com os dados de EOIs e armadilhas, em formato tabular.

Visualização Dados em bruto Maria Ferreira Logout

Protocolo Principal Generalista

Valor Médio Observado

Protocolo	Observação	Intervalo de Valores	Árvores (EOIs)																				Valor Médio Observado	
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Em EOIs Atacados	Global
Cecidíoma	8/20	5 rebentos / árvore	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,50	0,60
Esterifilose	2/20	5 folhas / Ramo (5 folhas terminais e expandidas)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,00	0,10

Capturas:

Bichado	Cecidíoma		Cochonilha A.	Cryptoblabes gnidiella	Funebrana	Hoplocampa	Molesta	Mosca Mediterrânica		Sésia
Armadilha	Caixa	Placa	Armadilha	Armadilha	Armadilha	Armadilha	Armadilha	Alimentar + Tridemelure*	Decis Trap	Armadilha
0	0	1	0	0	5	0	4	0	0	0

(c) Separador "Visualização" com uma visualização dos dados de EOIs e armadilhas.

Figura 7.4: Página da visita ao POB de São Lourenço (SL) no dia 7 de abril de 2021.

Finalmente, se clicarmos em “Estatísticas” na barra do topo da página, navegamos para a página da figura 7.5, que contém um conjunto de visualizações com estatísticas acerca da recolha de dados. A primeira visualização apresenta o número de visitas e observações realizadas por mês desde o início do projeto, enquanto que a segunda mostra o número de visitas por POB onde se executou cada protocolo, ao longo da campanha selecionada. Estas visualizações têm como objetivo apresentar o esforço significativo do trabalho de recolha realizado pelos técnicos de campo. Ambas foram construídas no âmbito da dissertação de A. Malafaia [3] e apenas foram melhoradas para serem apresentadas no *website*.

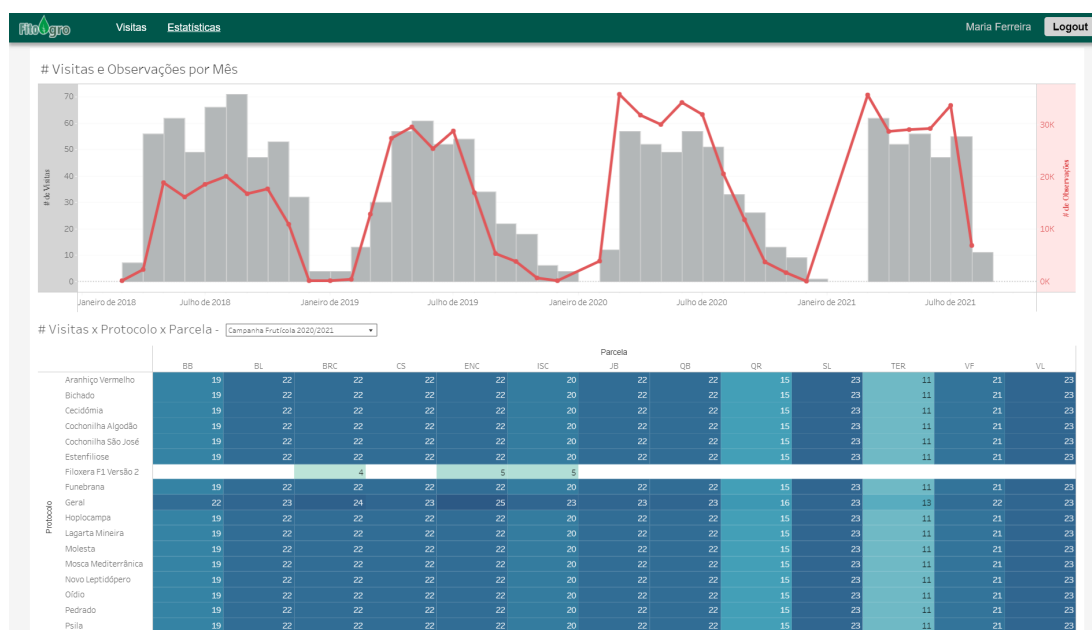


Figura 7.5: Página *web* “Estatísticas” que contém um conjunto de visualizações com estatísticas básicas sobre a recolha de dados.

7.2 Análítica dos Dados Fitossanitários

Nesta secção é apresentada a análise dos dados fitossanitários recolhidos durante as visitas dos técnicos aos POBs (integrados *à posteriori* no sistema de informação do projeto), cujo objetivo é responder aos requisitos analíticos levantados no capítulo 4. Deste modo, serão apresentadas soluções para visualizar e correlacionar os dados fitossanitários recolhidos (em 7.2.1) e para comparar esses dados entre diferentes anos e POBs (em 7.2.2).

7.2.1 Evolução das Observações Fitossanitárias

Dado o foco do projeto em quatro inimigos principais, cada um com um protocolo de recolha específico, foi considerado que o mais apropriado seria construir visualizações específicas para cada inimigo principal e uma generalista para as pragas secundárias. Estas visualizações têm como objetivo não só observar o desenvolvimento de cada inimigo

e a extensão do seu ataque no **POB**, mas também promover a descoberta de correlações entre esse desenvolvimento e outros fatores, como eventos meteorológicos e a aplicação de **produtos fitofarmacêuticos**. Deste modo, estas visualizações surgem como resposta ao requisito analítico **AD R4** e **AD R6**, uma vez que correlacionam as observações com outros fatores relevantes e, para além disso, calculam e apresentam percentagens de ataque e **NEAs** quando estes foram definidos para a praga.

Note-se que a informação recolhida para os diferentes inimigos foi variando ao longo do projeto, uma vez que os protocolos foram sofrendo alterações para tentar alcançar o maior nível de especificidade possível. Por uma questão de brevidade, este documento apenas irá apresentar e descrever as versões mais recentes dos protocolos, mas todas as visualizações estão preparadas para apresentar versões anteriores.

7.2.1.1 Cecidómia (*Dasineura pyri* Bouché)

O protocolo geral da cecidómia consiste na observação de 20 **EOIs** (i.e. árvores), onde em cada um se regista o número de rebentos atacados de um total de 5 observados. Para além disso, requer também a contagem de indivíduos capturados nas armadilhas previamente instaladas com as feromonas adequadas à **praga**. Por outro lado, durante a execução do protocolo específico deve-se ainda observar a presença/ausência de larvas nos rebentos atacados, registando o seu estágio através da sua coloração, que poderá ser branca, amarela ou laranja. Ademais, quando um rebento aparenta estar atacado mas não há larvas visíveis a olho nu, assume-se que contém larvas translúcidas.

De modo a permitir a análise dos dados recolhidos, construiu-se a visualização da figura 7.6, onde se utiliza um gráfico de barras para denotar o número de rebentos atacados por visita (i.e. somatório dos **EOIs**) e um gráfico de linhas para o número de indivíduos capturados nas armadilhas. As barras empilhadas são utilizadas para as observações porque são capazes de representar as partes de um todo e permitir ao mesmo tempo que se observe o todo, nomeadamente para o caso das observações de cecidómia isso corresponde aos rebentos atacados dos quais contém presença de larvas de diferente coloração. Por outro lado, as capturas representam-se por linhas uma vez que mesmo que existam várias espécies ou tipos de armadilhas, estas são medições independentes entre si.

No primeiro gráfico, quando o protocolo específico foi executado, empilha-se as barras pelo tipo de observação, i.e. pela coloração da larva, utilizando cores sequenciais de azul para enfatizar a progressão dos estágios da larva ao longo do tempo e promover a deteção das diferentes gerações no seu ciclo de vida. Repare-se que, para auxiliar a análise dos dados, anotou-se a data da visita e a correspondente percentagem de ataque (a vermelho) acima de cada conjunto de barras empilhadas, bem como o **NEA** da praga através de uma linha tracejada horizontal. Como descrito em 3.2, a percentagem de ataque para a cecidómia corresponde ao número de rebentos atacados em 100 rebentos observados e o **NEA** está estabelecido nos 50% dessa percentagem. Repare-se na utilização de um título dinâmico e explicativo da visualização e de eixos fixados entre 0 e 100, como

7.2. ANALÍTICA DOS DADOS FITOSSANITÁRIOS

definido pelas recomendações de boas prática descritas em 5.1. A fixação em 100 para as observações deve-se ao número máximo de rebentos observados, enquanto que para as capturas, embora não tenham máximo, acordou-se que seria um valor satisfatório para permitir a sua uniformização entre os diferentes POBs e anos.

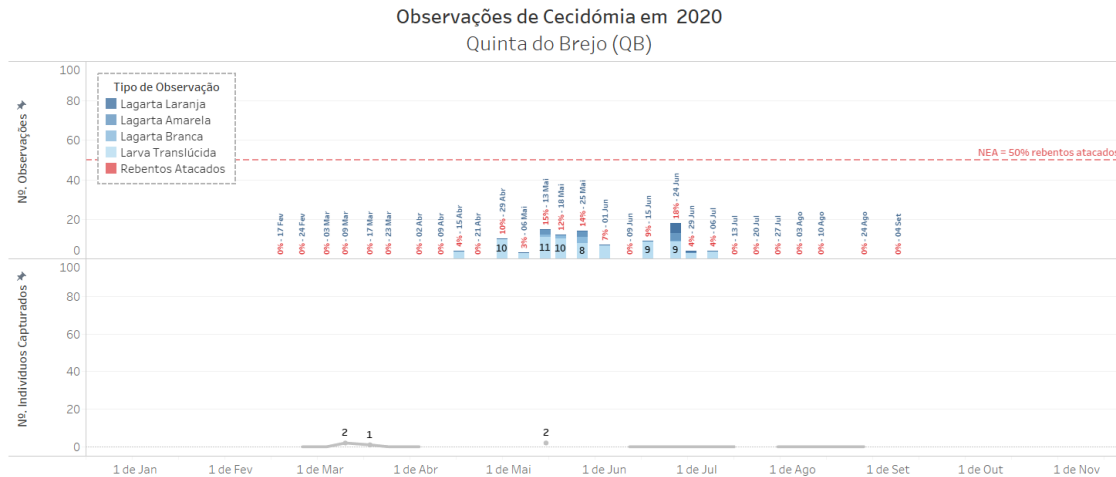


Figura 7.6: Visualização das observações registadas para a cecidómia ao longo de 2020, no POB Quinta do Brejo (QB).

Apesar da visualização anterior permitir a análise do desenvolvimento da praga, não permite correlacionar o mesmo com as restantes observações fitossanitárias e outros fatores que possam ser relevantes. Deste modo, criou-se elementos complementares a essa visualização para promover essa descoberta de correlações, apresentados de seguida.

Em primeiro lugar, construiu-se uma faixa cronológica para as visitas, uma vez que, como referido em 3.2, os técnicos registam informação adicional sobre as mesmas, como a realização ou não da visita, troca de feromonas às armadilhas ou outras notas gerais. Na figura 7.7 é possível observar a faixa construída para esse efeito, onde se indica o tipo de visita através da forma (realização ou não da visita e troca de feromonas) e da cor (existência/ausência de notas). Note-se ainda que ao passar o cursor por uma dessas visitas é apresentada uma **tooltip** com informação sobre a visita e com as notas registadas durante a mesma, caso existam. Com esta faixa é possível ter uma visão holística das visitas efetuadas ao POB, independentemente dos protocolos executados.

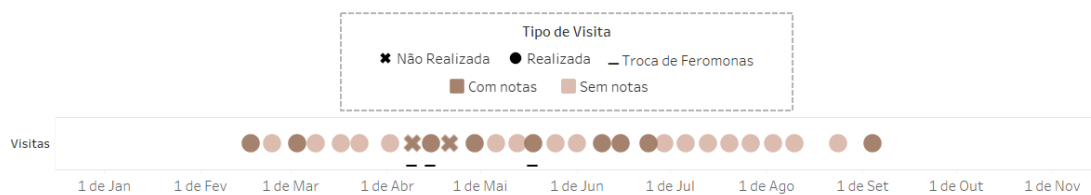


Figura 7.7: Visualização que apresenta uma visão holística das visitas ao longo de 2020, no POB Quinta do Brejo.

Para além disso, os técnicos registam durante a visita o **estado fenológico** dos pomares, visto que o crescimento vegetativo poderá estar relacionado com o desenvolvimento das pragas. Por esse motivo, também se construiu a faixa cronológica da figura 7.8 para representar o **estado fenológico** (E.F.) registado em cada visita, sendo que os diferentes estados foram mapeados para a matiz de cores de forma a enfatizar uma noção de progressão temporal dos mesmos. A utilização dessas cores foi largamente discutida com os parceiros do projeto, tendo sido aceite posteriormente.



Figura 7.8: Visualização que apresenta os estados fenológicos registados ao longo de 2020, no POB Quinta do Brejo.

Adicionalmente, os técnicos registam a ocorrência de eventos durante a visita como estar a chover ou o pomar estar sob tratamento. Note-se que estes eventos são reportados apenas do parecer dos técnicos e das visitas, uma vez que podem impossibilitar a realização de visitas, para além de influenciar o desenvolvimento das pragas. Dada a sua relevância, construiu-se uma faixa cronológica semelhante à das visitas, mas neste caso representa-se a ocorrência de eventos com o seu tipo mapeado para a forma, como é observável na figura seguinte (7.9).

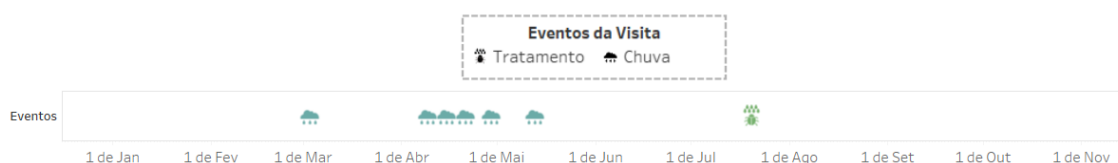


Figura 7.9: Visualização que apresenta os eventos que ocorreram nas visitas de 2020, no POB Quinta do Brejo.

Finalmente, embora a ocorrência de tratamentos durante a visita possa explicar alguns padrões de desenvolvimento da praga, nem sempre os tratamentos ocorrem no mesmo dia ou momento da realização da visita e nem todos os tratamentos têm efeitos semelhantes para todas as pragas. Por esse motivo, alguns *stakeholders* do projeto consideraram que seria de grande interesse a partilha de informação, por parte dos produtores, relativamente da aplicação de **produtos fitofarmacêuticos** nos POBs do projeto. Esses dados foram analisados e divididos em três categorias de acordo com a praga alvo do produto e a possibilidade ou não de efeitos para cada praga em estudo, nomeadamente: praga é o alvo, possíveis efeitos na praga e outras pragas são o alvo. Após esta categorização ser validada pelos técnicos de produção, construiu-se uma faixa cronológica para a realização de tratamentos, apresentada na figura 7.10, e cuja cor mapeia-se para a categorização descrita anteriormente. Note-se que como as datas dos tratamentos, ao contrário dos restantes dados, nem sempre correspondem às datas das visitas, colocou-se abaixo a data do

7.2. ANALÍTICA DOS DADOS FITOSSANITÁRIOS

tratamento. Ademais, visto que por vezes pode haver vários tratamentos num dia, cada ponto poderá ser dividido pelo número de tratamentos efetuados naquele dia. Por exemplo, na figura abaixo é possível observar que apenas a partir de maio é que houveram tratamentos específicos para a praga em estudo, nomeadamente quatro de seguida, enquanto que antes disso apenas o tratamento de 26 de março é que poderá ter tido algum efeito na praga. Por fim, nota-se que no dia 7 de maio realizaram-se dois tratamentos ao **POB**, mas nenhum deles afeta a **praga** em estudo.

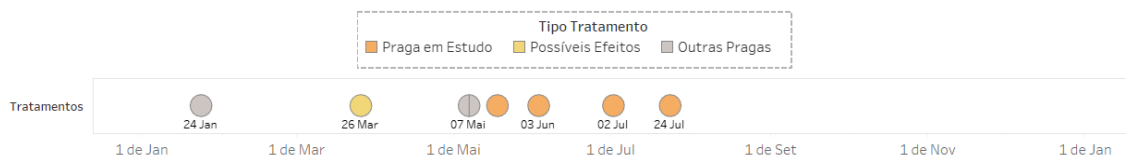


Figura 7.10: Visualização que apresenta os tratamentos efetuados ao longo de 2020, no **POB** Vale Leite.

Ao juntar todas as faixas apresentadas anteriormente à visualização das observações fitossanitárias apresentada inicialmente, obtém-se um *dashboard* que promove a descoberta de correlações entre todos estes elementos. Na figura seguinte (7.11) é apresentado esse *dashboard* com os dados registados no **POB** Quinta do Brejo ao longo do ano de 2020.

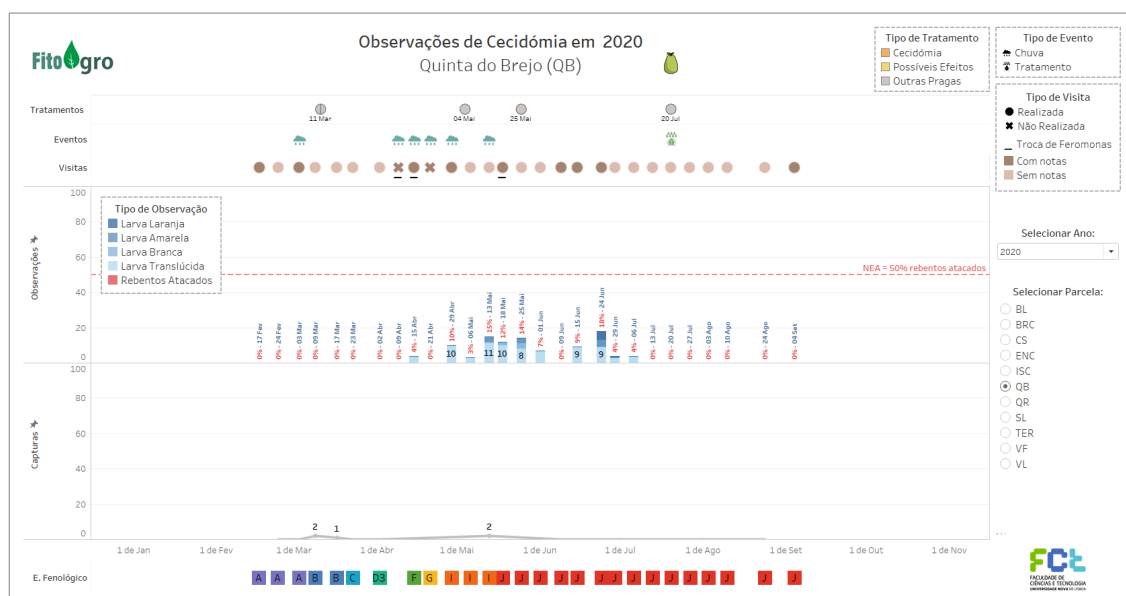


Figura 7.11: *Dashboard* com as observações fitossanitárias de cecidómia e outros fatores relevantes que ocorreram no **POB** Quinta do Brejo, ao longo do ano de 2020.

Do lado direito da visualização pode-se selecionar o ano e escolher o **POB**, cujos dados se pretende visualizar. Repare-se que ao lado do título colocou-se um ícone com a cultura do pomar (i.e. **POB**) selecionado, para que a esta seja facilmente identificável

7.2.1.2 Cochonilha-Algodão (*Pseudococcus viburni*)

Para a cochonilha-algodão o protocolo geral consiste na observação de 25 EOIs (i.e. árvores), onde em cada uma se observa 5 estruturas (i.e. ramos) velhas, 5 estruturas novas e 5 frutos contíguos, registrando-se a quantidade atacada de cada um. Para o protocolo específico, observa-se ainda a quantidade de frutos, estruturas velhas e estruturas novas atacadas por cada estágio da praga, nomeadamente por ovos, ninfas e adultos.

Utilizando a mesma estrutura e os elementos descritos na secção anterior para a cecidómia, construiu-se o *dashboard* da figura 7.12. Este *dashboard* é semelhante ao da cecidómia, no entanto, dada a especificidade de cada protocolo, a forma de representação das observações fitossanitárias requer adaptação conforme os dados recolhidos para o inimigo em foco. Deste modo, para a cochonilha-algodão, utilizou-se um gráfico de barras empilhadas com o número de estruturas ou frutos atacados, sendo que quando o protocolo específico é executado, essas observações são divididas pelos diferentes estágios da praga. Note-se que a estrutura dos restantes elementos (como o estado fenológico, tratamentos, etc.) não sofreu quaisquer adaptações, uma vez que estão associadas à visita e ao pomar e não aos dados protocolares. No entanto, as visitas apresentadas são filtradas de acordo com o protocolo em execução, ou seja, dados de visitas ao POB onde o protocolo não foi executado são omitidas da visualização. O exemplo da figura não apresenta muitos dados, porque no geral não se registaram muitas observações de cochonilha-algodão nos últimos anos.

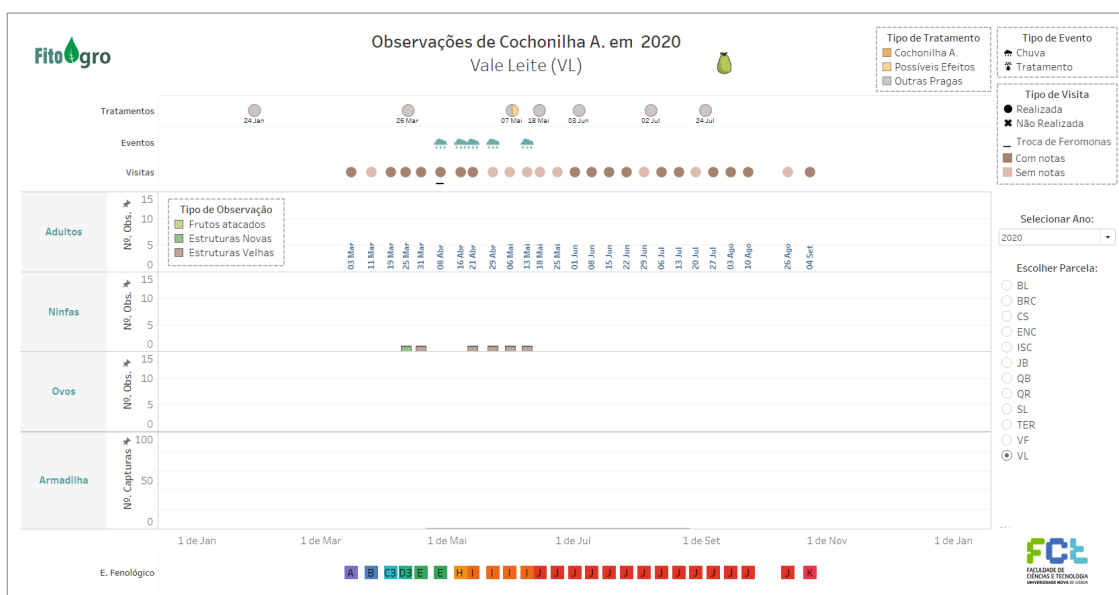


Figura 7.12: *Dashboard* com as observações fitossanitárias de cochonilha-algodão e outros fatores relevantes que ocorreram no POB Vale Leite, ao longo do ano de 2020.

Tal como o da cecidómia, este *dashboard* também utiliza um título dinâmico e explicativo da visualização e eixos fixados entre 0 e 15 para as observações e entre 0 e 100 para as capturas, como definido pelas recomendações de boas prática descritas em 5.1. A

fixação em 15 para as observações deve-se ao número máximo de estruturas observados, enquanto que para as capturas mantém-se a explicação da secção anterior.

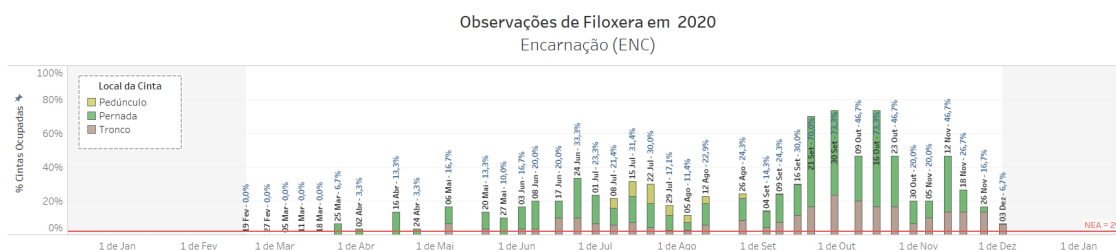
7.2.1.3 Filoxera (*Aphanostigma pyri* Chol.)

O protocolo da filoxera é o mais complexo do projeto, sendo dividido em 4 fases de recolha e cuja primeira fase é dividida em 2 sub-fases, dependendo do momento em que ocorre. As diferentes fases são descritas de seguida.

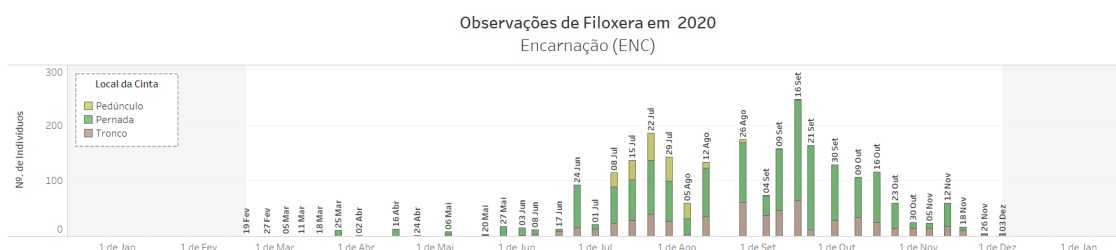
- **Fase I (Pré colheita)** - Inicia-se quando o pomar atinge o estado fenológico B e prolonga-se até ao início da fase II. Durante esta fase são observados 10 EOIs (i.e. árvores), cada um com cintas instaladas previamente no tronco e em duas pernasadas (i.e. ramos grossos da árvore), para um total de três por árvore. As cintas são recolhidas e trocadas nas visitas, sendo posteriormente analisadas em laboratório para registar o número de cintas ocupadas pela praga, bem como o número de indivíduos que as ocupam.
- **Fase II** - Decorre desde que os frutos atingem um calibre entre 25 a 35mm até à colheita do pomar. Nesta fase observa-se o mesmo do que a fase anterior, com a adição de 4 cintas divididas por 4 frutos ou ramos de cada EOI observado. Como na fase anterior é efetuada a recolha e troca de cintas e posteriormente registado em laboratório o número de cintas ocupadas por local (tronco, pernasadas e frutos/pedúnculos), bem como o número de indivíduos que se encontravam em cada uma.
- **Fase III** - Esta fase ocorre em simultâneo com a fase II, iniciando-se no entanto 3 semanas antes da colheita e com duração de 2 visitas/recolhas. Durante a mesma, recolhe-se 3 frutos de árvores marcadas (i.e. EOIs) e 3 frutos de árvores não marcada. Posteriormente em laboratório, os frutos sofrem um corte longitudinal e são analisados através da lupa binocular para registar a presença de sintomas, nomeadamente auréola amarela ou podridão, bem como a presença de filoxera no seu cálice, através da contagem do número de indivíduos presentes para cada estágio da praga (ovos, ninfas e fêmeas). Quando o fruto apresenta ambos os tipos de sintomas, regista-se apenas o que indica um estado mais avançado, neste caso a podridão.
- **Fase IV** - Esta fase decorre durante a colheita (1 visita/recolha) de forma complementar à fase anterior, uma vez que é recolhida uma maior quantidade de frutos, nomeadamente 5 frutos em cada um dos 20 EOIs, para um total de 100 frutos. A observação em laboratório é a mesma da fase III, sendo também registado, para cada um dos 100 frutos, a presença de sintomas e o número de indivíduos presentes de cada estágio.
- **Fase I (Pós colheita)** - A recolha e o registo durante esta fase é exatamente igual à fase de pré colheita, no entanto esta fase decorre durante 4 semanas a partir do momento da colheita.

Dada a complexidade do protocolo, a construção da visualização das observações fitossanitárias teve de ser várias vezes iteradas, de forma a obter uma solução que fosse capaz de representar todas as fases simultaneamente ao longo do mesmo eixo temporal. A estrutura desta visualização manteve, sem alterações, as faixas cronológicas para o estado fenológico, tratamentos, eventos e visitas apresentadas nas secções anteriores.

Para ambas as sub-fases da fase I e para a fase II construiu-se um gráfico de barras empilhadas com duas vertentes analíticas: extensão e intensidade do ataque. Para a primeira vertente, apresentada na figura 7.13a, utilizou-se o número de cintas ocupadas para calcular a percentagem ocupada em cada local da árvore (tronco, pernas e pedúnculos), permitindo assim analisar a extensão do ataque da filoxera nas árvores do pomar. Repare-se que, uma vez que a percentagem de ataque da filoxera corresponde à percentagem total de cintas ocupadas, essa percentagem aparece indicada no topo das barras e o NEA associado (2%) aparece como uma linha horizontal a vermelho. Por outro lado, para analisar a intensidade do ataque da praga, utilizou-se a contagem do número de indivíduos presentes em cada local das cintas, como é possível observar na figura 7.13b. Nos gráficos de ambas as vertentes, o local das cintas foi mapeado para a cor e a data da visita foi indicada no topo de cada barra. Note-se ainda que a ordem utilizada para representar o local das cintas, segue a mesma ordem vertical destas nas árvores, isto é, de baixo para cima temos primeiro as cintas do tronco, depois das pernas (ramos grossos) e, por fim, dos pedúnculos (ramos dos frutos).



(a) Visualização para analisar a extensão do ataque através da percentagem de cintas ocupadas por filoxera, consoante o local da árvore.



(b) Visualização para analisar a intensidade do ataque através do número de indivíduos de filoxera a ocupar cintas, consoante o local da árvore.

Figura 7.13: Visualizações para as fases I e fase II, onde se analisa a extensão e intensidade do ataque de filoxera ao longo de 2020, no POB Encarnação.

7.2. ANALÍTICA DOS DADOS FITOSSANITÁRIOS

Para as fases III e IV construiu-se os dois gráficos de barras empilhadas da figura 7.14, sendo um para analisar a presença de indivíduos nos frutos e outro para a de sintomas nos mesmos. Representou-se o primeiro através da contagem, para cada estágio, do número de indivíduos presentes nos frutos, mapeando a progressão dos estágios através escalas sequenciais de azul. Por outro lado, para a presença de sintomas apresentou-se a percentagem de frutos com auréola amarela e a percentagem com podridão, do total de frutos observados em cada fase, e distinguindo o tipo de sintoma pela cor.

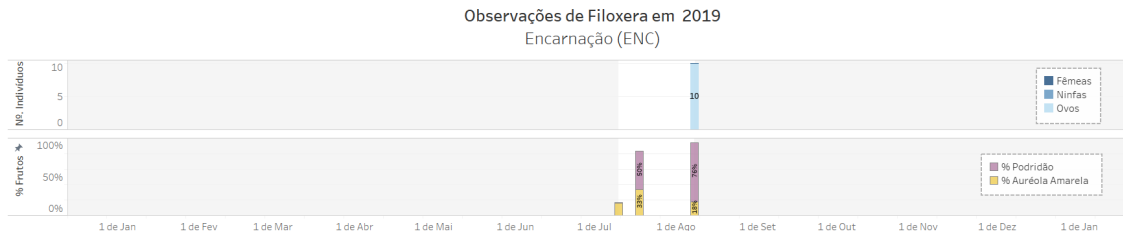


Figura 7.14: Visualização para a fase III e IV, onde se analisa os sintomas dos frutos e o número de indivíduos que contêm de cada estágio, para o POB Encarnação em 2019.

Ao juntar estas visualizações às restantes faixas cronológicas achou-se que seria importante identificar através de gráficos de *gantt*, em cima de cada visualização, o período em que ocorreu cada fase. Na figura 7.15 é apresentada a *dashboard* para o protocolo da filoxera, através do qual é possível observar cronologicamente todas as fases, bem como analisar e correlacionar estas com os outros elementos relevantes, descritos anteriormente. Note-se que na figura apresenta-se a apenas a visualização da extensão do ataque, mas acima desta encontra-se um botão que permite alternar facilmente entre as duas vertentes.

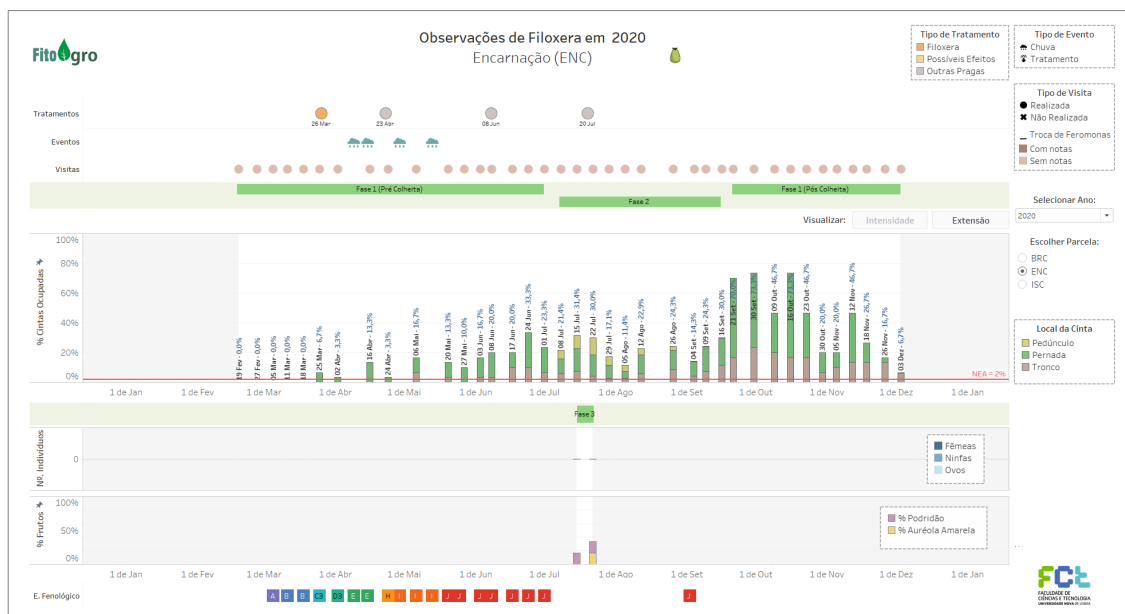


Figura 7.15: *Dashboard* com a extensão do ataque de filoxera e outros fatores relevantes que ocorreram no POB Encarnação, ao longo de 2020.

Uma vez que este *dashboard* tem muitos componentes e que algumas visualizações estão um pouco distantes entre si, o eixo temporal surge duplicado (ao centro e em baixo) como recomendado pelas boas práticas descritas em 5.1.2.

7.2.1.4 Lepidópteros

Uma vez que a espécie do novo lepidóptero não é conhecida, o seu protocolo abrange todas as espécies de lepidópteros, exceto o bichado da macieira (*Cydia pomonella*) pois este é visualmente fácil de distinguir dos restantes. O protocolo para os lepidópteros consiste na visita a 20 EOIs, onde em cada um se observa 20 frutos contíguos, registando quantos foram atacados. Para além disso, regista-se também o número de indivíduos de cada espécie capturado nas armadilhas, cuja análise poderá ter de ser efetuada posteriormente em laboratório para melhor identificar cada espécie.

Deste modo, construiu-se uma visualização para os lepidópteros com a espécie mapeada para a cor, de forma a permitir a comparação entre estas e promover a identificação da espécie do novo lepidóptero. Esta visualização foi dividida em dois gráficos, sendo o primeiro um gráfico de barras com o número de frutos atacados por lepidópteros e o segundo um gráfico de linhas com o número de indivíduos capturados por espécie, traçando assim a chamada curva de voo dos insetos. A união desta visualização com as faixas cronológicas apresentadas anteriormente resultou no *dashboard* da figura 7.16. A visualização deste *dashboard* permite a observação do desenvolvimento das diferentes espécies de lepidópteros e a descoberta de correlações com outros fatores, promovendo também a identificação da espécie do novo lepidóptero. Note-se que não foi possível apresentar um exemplo com frutos atacados, uma vez que, até ao momento da escrita deste documento, não existem registos dessa observação no sistema.

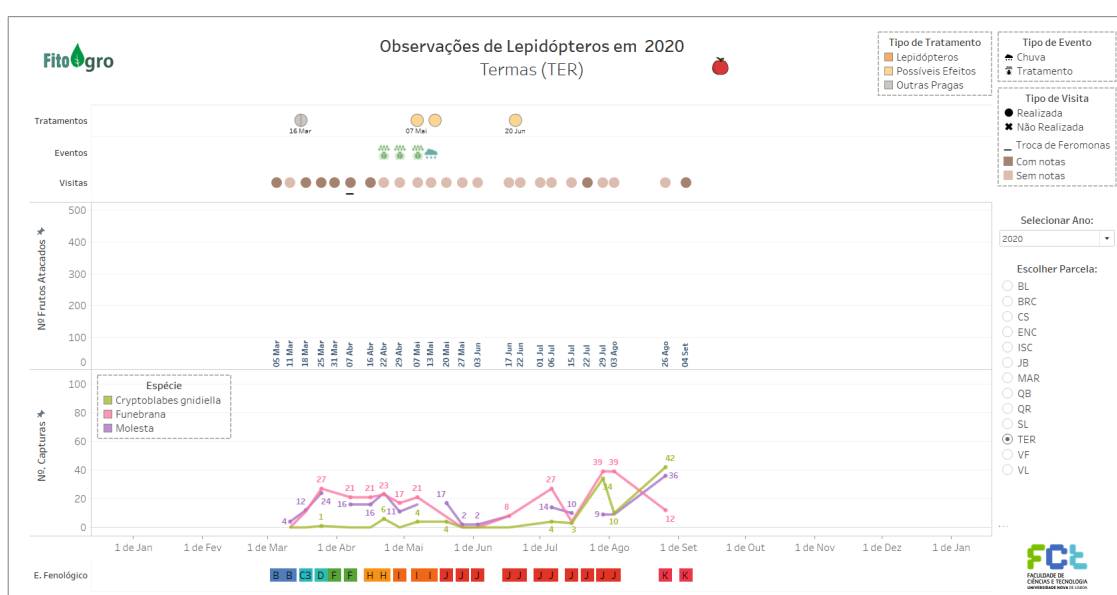


Figura 7.16: *Dashboard* com as observações fitossanitárias das diferentes espécies de lepidópteros, bem como outros fatores relevantes que ocorreram no POB Termas, em 2020.

7.2. ANALÍTICA DOS DADOS FITOSSANITÁRIOS

Por outro lado, o bichado é um dos principais lepidópteros estudados pela comunidade científica, como já foi referido na secção 6.2.2, sendo que, por isso, tem o seu ciclo de vida e as condições ótimas para o seu desenvolvimento bem definidas na literatura. No projeto FitoAgro, o bichado não é um inimigo principal pelo que não tem um protocolo de recolha específico, fazendo apenas parte do protocolo generalista. No entanto, por ser um inimigo bastante estudado, decidiu-se que seria relevante a construção de visualizações específicas para o mesmo, até porque serviria como base para o estudo da parametrização e ajuste dos modelos preditivos, descrito no final deste capítulo.

Por esse motivo, construiu-se para o bichado uma visualização praticamente idêntica à dos restantes lepidópteros, diferindo apenas no protocolo do bichado a observação de 50 frutos contíguos por árvore em vez de 20. A estrutura da visualização das observações mantém-se assim a mesma da anterior, com um gráfico de barras para os frutos atacados e um gráfico de linhas para a curva de voo. Na figura 7.17 é apresentado o *dashboard* construído para o bichado, onde se anotou o seu NEA que corresponde desde 0.5 a 1% de frutos atacados ou a 3 capturas em macieiras e 4 em pereiras. Note-se que novamente não foi possível apresentar um exemplo com frutos atacados.

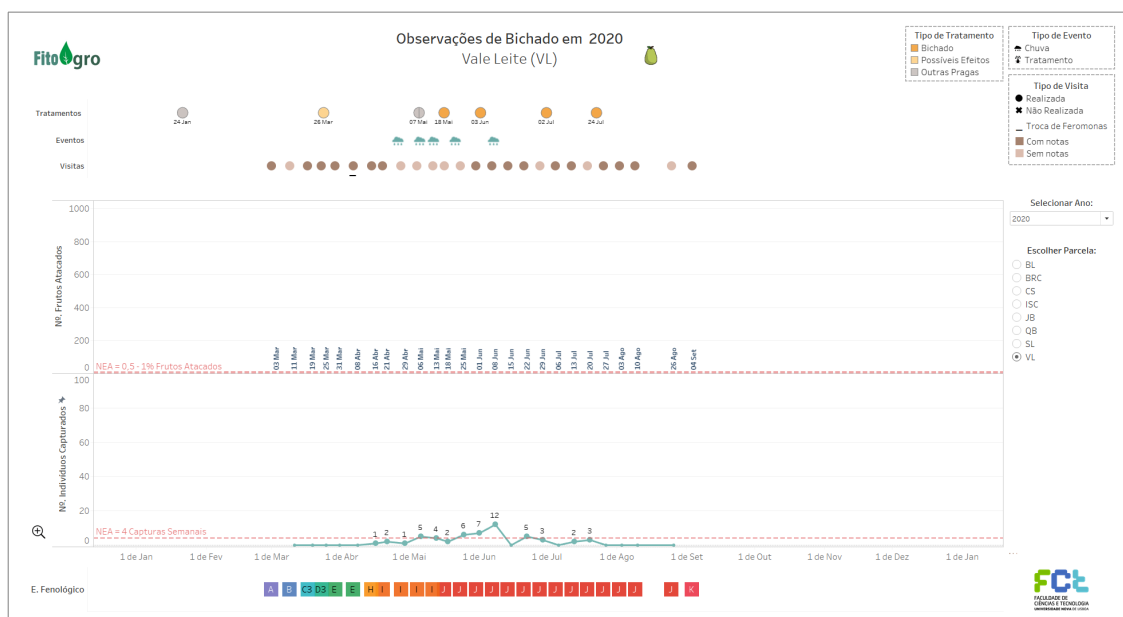


Figura 7.17: *Dashboards* com as observações fitossanitárias das diferentes espécies de lepidópteros, bem como outros fatores relevantes que ocorreram no POB Vale do Leite, ao longo de 2020.

Seguindo novamente as recomendações de boas práticas da secção 5.1.2, o eixo dos frutos atacados foi fixado em 1000, por ser esse o número total de frutos observados, e o das capturas entre 0 e 100 para permitir uma correta comparação dos dados entre os POBs. No entanto, quando se quer analisar os dados em maior detalhe ou quando o número de frutos atacados e/ou de indivíduos capturados é reduzido, poderá ser útil desprender o eixo. Por esse motivo, colocou-se uma botão em forma de lupa ao lado dos

eixos que permite desprender/fixar o eixo conforme seja necessário. A necessidade desta funcionalidade tornou-se evidente para o bichado dado que o número total de observações é superior ao dos restantes protocolos, no entanto esta solução seria facilmente aplicável às restantes pragas.

7.2.1.5 Pragas Secundárias

Para as restantes 12 pragas, como são inimigos secundários no projeto, é apenas executado o protocolo generalista, onde se recolhem também dados diferentes consoante a praga, mas com um menor grau de especificidade. Utilizando a mesma estrutura das visualizações dos inimigos principais apresentadas anteriormente, construiu-se uma visualização genérica que se adapta aos dados fitossanitários recolhidos para qualquer praga secundária. Para isso, dividiu-se os dados de recolha em dois tipos: observações em EOIs e capturas em armadilhas. Para o primeiro tipo, colocou-se um gráfico de barra por tipo de observação, sendo que para alguns casos pode não ser o mais intuitivo mas é a abordagem mais genérica. Por outro lado, para as capturas em armadilhas utilizou-se um gráfico de linhas por armadilha, caso sejam mais que uma. Como já foi referido, as visualizações deste *dashboard* adaptam-se aos dados recolhidos para o inimigo selecionado (do lado direito), sem necessidade de navegar entre *dashboards*. Na figura 7.18 é apresentado o *dashboard* generalista para o inimigo sésia.

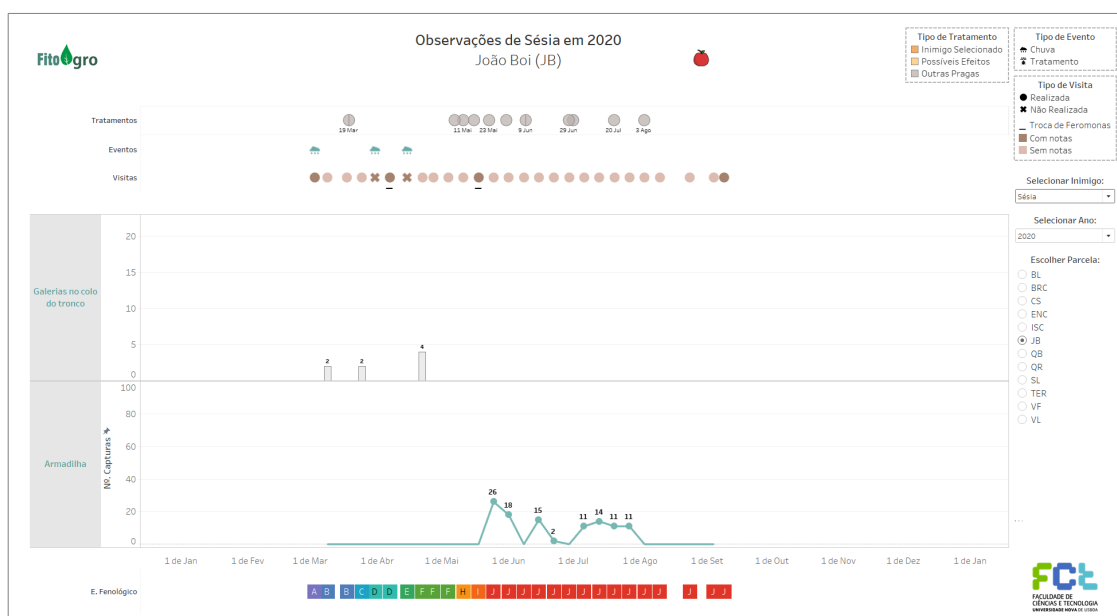


Figura 7.18: *Dashboard* generalista com as observações fitossanitárias da sésia, bem como outros fatores relevantes que ocorreram no POB João Boi, ao longo de 2020.

As variantes deste *dashboard* para os restantes inimigos secundários encontram-se no apêndice C, sendo que algumas pragas não têm dados porque não foram observadas em 2020 nos POBs. Note-se que todos os eixos deste *dashboard* têm o eixo das observações fixo para o número de observações do protocolo, como recomendado pelas boas práticas

da secção 5.1.2. No exemplo da figura acima, regista-se a presença/ausência de sésia nas galerias do colo do tronco de 20 EOIs, logo só podem haver no máximo 20 presenças, sendo por isso o máximo fixado para desse eixo.

Em suma, utilizou-se a mesma estrutura base para todos os inimigos, sendo que apenas a visualização das suas observações fitossanitárias foi customizada para cada inimigo principal e generalizada para os restantes, mas mantendo sempre gráficos de barras para as observações em EOIs e de linhas para as capturas em armadilhas. No caso da filoxera, dado a complexidade do seu protocolo, essa visualização foi dividida por fases, agrupando a fase 1 com a 2 e a 3 com a 4. No diagrama seguinte (7.19) ilustra-se a estrutura genérica dos *dashboards*, indicando a visualização customizável e as adaptações para a filoxera.

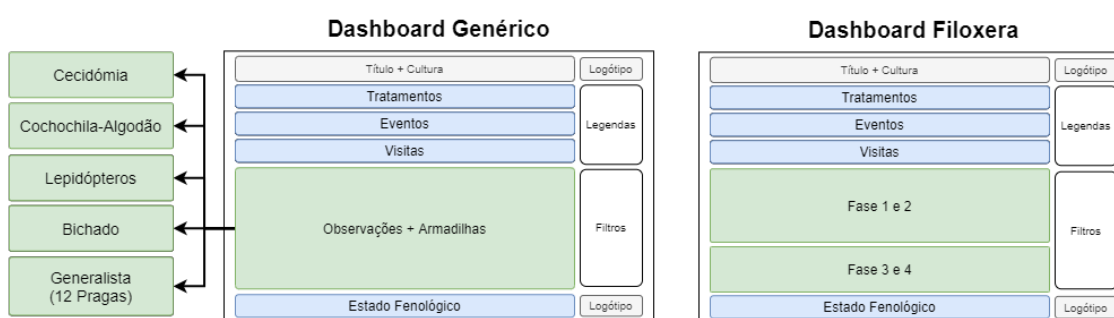


Figura 7.19: Estrutura genérica dos *dashboards* para observação dos dados fitossanitários, com indicação da visualização que varia por inimigo e da variante para a filoxera.

7.2.2 Comparação entre anos e POBs

Para estudar o desenvolvimento dos insetos e a sua correlação com as condições climáticas e outros fatores, é necessário primeiro analisar dados de vários locais e em diferentes anos para encontrar padrões comuns. No entanto, a comparação através da análise iterativa dos dados de cada POB e de cada ano, além de menos eficaz, pode levar à ocorrência do fenómeno *perception blindness*, descrito anteriormente nos fundamentos básicos da visualização interativa de dados (secção 2.2.1).

Nesse sentido surgiu o requisito analítico **AD R6**, que descreve a necessidade de comparar POBs e anos através da análise simultânea de todos os seus dados. Para isso, duplicou-se e reordenou-se a estrutura base dos *dashboards* da secção anterior, de forma a suportar duas dessas estruturas na vertical e depois alinhar verticalmente os eixos temporais. Na figura 7.20 é apresentado um exemplo do *dashboard* que resulta dessa transformação, onde é possível comparar verticalmente toda a informação da cecidómia em dois POBs diferentes, ao longo do mesmo ano.

Por outro lado, na figura 7.21 é apresentado um exemplo de um *dashboard* semelhante, mas onde é possível comparar verticalmente toda a informação da cecidómia no mesmo POB, ao longo de dois anos diferentes. Note-se que ainda não houve partilha dos dados relativamente aos tratamentos efetuados em 2021.

CAPÍTULO 7. FITOSSANIDADE - DESENHO E IMPLEMENTAÇÃO



Figura 7.20: *Dashboard* que permite comparar toda a informação recolhida para a cecidómia nos POBs Quinta do Brejo e Casal Serrano, ao longo do ano de 2020.



Figura 7.21: *Dashboard* que permite comparar toda a informação recolhida para a cecidómia ao longo de 2020 e 2021, no POB Quinta do Brejo.

7.3 Impacto da Meteorologia na Fitossanidade

Nesta secção é discutida a importância de promover a descoberta de correlações entre a meteorologia e a fitossanidade, de forma a compreender a extensão do seu impacto no desenvolvimento dos diferentes inimigos. Para isso, descreve-se a relevância dos graus-dia acumulados para efetuar essa correlação (em 7.3.1), enfatizando a sua importância para os modelos de desenvolvimento das pragas e para o ajuste de modelos preditivos (em 7.3.2).

7.3.1 Importância dos Graus-dia Acumulados

O conceito e aplicação dos graus-dia já foi explicado em 6.2.2, bem como um exemplo do cálculo do seu valor acumulado, os graus-dia acumulados. Como foi referido também nessa secção, o cálculo desta variável está diretamente ligada às condições ótimas definidas para cada inseto. Deste modo, o desenvolvimento da praga varia ao longo do tempo entre POBs e anos, uma vez que este depende das condições climáticas que ocorrem naquele local e momento. Pode-se concluir assim que a utilização do tempo/calendário para efetuar comparações não é a mais adequada, uma vez que o inseto depende diretamente das condições climáticas consideradas ótimas para o seu desenvolvimento, que, por sua vez, variam ao longo do tempo para diferentes locais e anos.

Perante esta conclusão, foi necessário encontrar uma variável temporalmente dependente e diretamente correlacionada com as condições ótimas para o desenvolvimento do inimigo, ou seja, no caso do FitoAgro com as condições climáticas. Naturalmente que a variável selecionada foi os graus-dia, nomeadamente os graus-dia acumulados, até porque é também a variável adotada por maioria dos modelos de desenvolvimento para pragas.

Nesse sentido, propôs-se a replicação dos *dashboards* de comparação apresentados na secção anterior, sendo que o eixo em comum passou a ser os graus-dia acumulados em vez do tempo. Para isso, começou-se por efetuar um ensaio para a cecidómia, cujas condições ótimas associadas ao cálculo de graus-dia não estão definidas na literatura para a região do Oeste. No entanto, no projeto adotou-se o uso de 0°C como o seu zero de desenvolvimento e o modelo *Modified Average* de P. Damos & M. Savopoulou-Soultani (2012) [55] para o cálculo dos graus-dia. Por sua vez, este cálculo é feito através da equação 7.1, onde $maxT$ é a temperatura máxima diária e T_{base} é o zero de desenvolvimento do inseto.

$$DD(dia) = \begin{cases} \frac{T_{base} + maxT}{2} - T_{base} & , \text{ se } Temp \geq T_{base} \\ 0 & , \text{ caso contrário} \end{cases} \quad (7.1)$$

Uma vez que no projeto é utilizado 0°C como zero de desenvolvimento (T_{base}) da cecidómia, essa equação simplifica-se neste caso para a metade da temperatura máxima

diária ($maxT$). A acumulação desses valores diários corresponde então aos graus-dia acumulados. Ao associar a cada data o número de graus-dias acumulados até aquele instante, obtém-se um gráfico como o da figura 7.22. Note-se que neste indica-se o número de graus-dia acumulados até ao início de cada mês, sendo que existem meses onde se acumulam mais e outros menos, como por exemplo o mês de julho acumulou 651 grau-dia enquanto o mês de janeiro apenas acumulou 337.

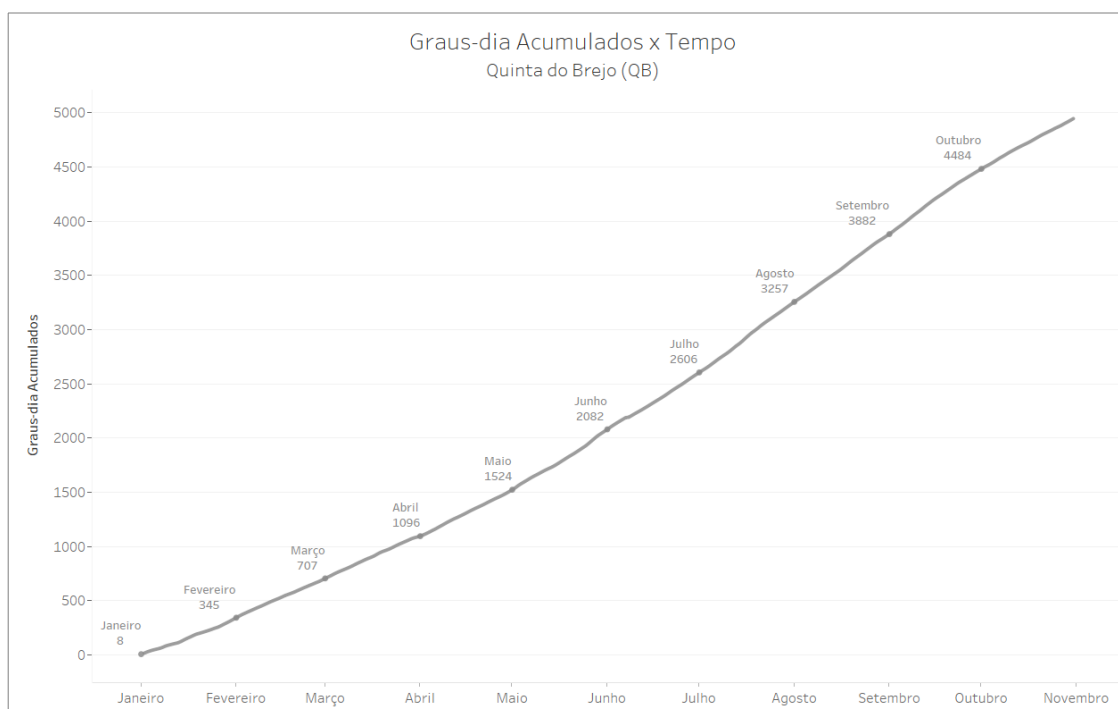


Figura 7.22: Graus-dia acumulados no POB Quinta do Brejo, ao longo de 2020.

Ao remover o eixo temporal mantendo o número de grau-dia acumulados associados às datas, obtém-se faixas de comprimento variado entre meses, que correspondem ao número de grau-dia acumulados ao longo de cada um, tal como exemplificado na figura anterior. Na figura 7.23 ilustra-se isso para o mesmo exemplo da figura anterior, onde é observável que o comprimento da faixa para janeiro é muito inferior a julho.

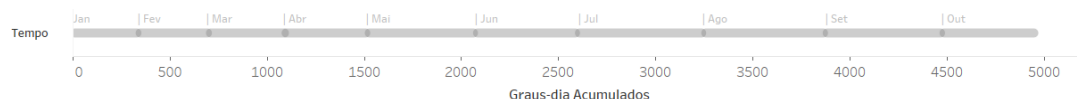


Figura 7.23: Faixa de graus-dia acumulados no POB Quinta do Brejo, ao longo de 2020.

Dado que os valores acumulados definem funções monótonas não-decrescentes, garantem progressão temporal, isto é, a ordem das datas é a mesma dos valores acumulados, visto que não existem dias com valores acumulados menores do que os seus dias anteriores. Assim, ao associar às datas das visitas o valor acumulado de graus-dia até aquele instante, torna-se possível representar toda a informação fitossanitária ao longo de um eixo de graus-dia acumulados.

7.3. IMPACTO DA METEOROLOGIA NA FITOSSANIDADE

Nas figuras 7.24 e 7.25 apresentam-se os *dashboards* para comparação, respetivamente, entre POBs e entre anos ao longo dos graus-dia acumulados, relativamente à cecidómia. Note-se que foi adicionada uma linha cronológica à faixa das visitas que, de forma semelhante à figura anterior, marca o número de graus-dia acumulados visualmente para cada mês, preservando a noção temporal das visitas.

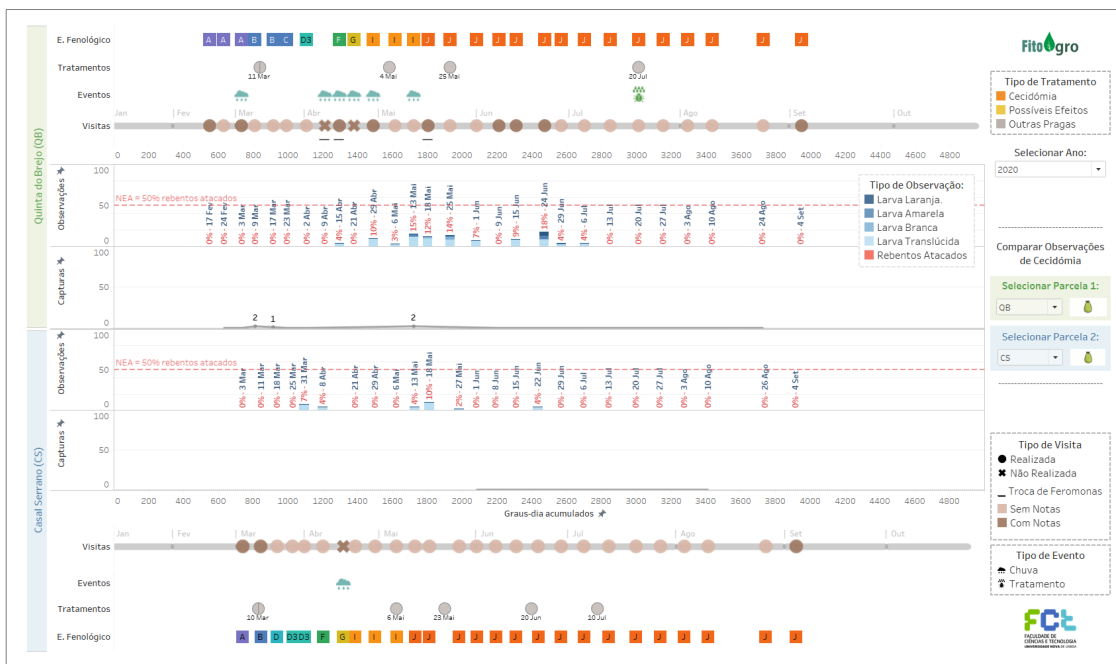


Figura 7.24: *Dashboard* que permite comparar toda a informação de cecidómia recolhida nos POBs Quinta do Brejo e Casal Serrano, ao longo dos graus-dia acumulados em 2020.

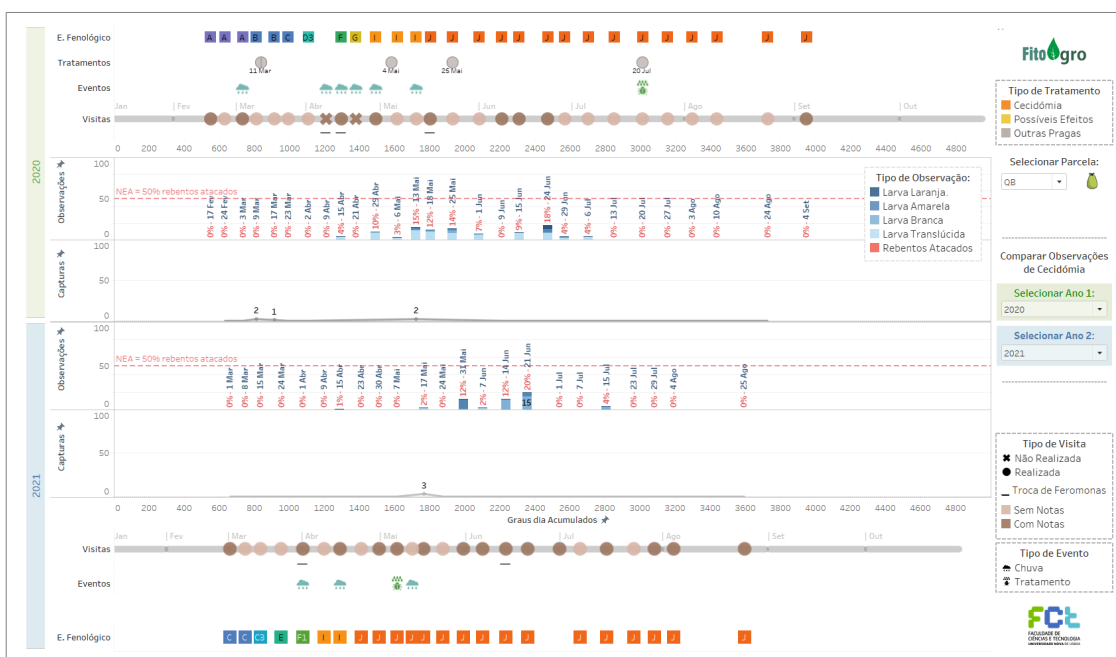


Figura 7.25: *Dashboard* que permite comparar toda a informação de cecidómia recolhida ao longo dos graus-dia acumulados em 2020 e 2021, no POB Quinta do Brejo.

Esta solução original permite comparar o desenvolvimento das pragas através das condições climáticas registadas nos diferentes locais e anos, possibilitando uma análise comparativa mais correta e precisa, bem como a descoberta de outros fatores influentes nesse desenvolvimento, para além da temperatura. Este conjunto de *dashboards*, sendo facilmente replicado para outros inimigos, estende a solução inicial do requisito analítico **AD R5** para responder também ao requisito **AD R7**.

Dada a originalidade e importância desta solução, a mesma foi apresentada em painel no Congresso de Fruticultura da Feira Frutos 2020, tendo sido *à posteriori* escrito, em conjunto com a restante equipa da FCT/UNL e com o suporte da NOVA LINCS, um artigo científico sobre esta temática, intitulado "FitoAgro: Ferramenta para recolha de informação e tomada de decisão visando a sustentabilidade e competitividade das explorações frutícolas", publicado na 34ª edição da coleção de *Actas Portuguesas de Horticultura* (2020) [56]. Para além disso, este artigo foi adaptado, em conjunto com alguns membros do COTHN pertencentes ao projeto FitoAgro, para publicação no livro *Grupos Operacionais de Fruticultura no período 2018-2022* (2021) [57].

7.3.2 Ajuste de Modelos Preditivos

Os modelos preditivos, como foi explicado em 3.3, têm como função ajustar os modelos de desenvolvimento das pragas às condições climáticas e fitossanitárias em cada momento e local. No entanto, muitas pragas não têm modelos de desenvolvimento bem definidos, isto é, não se sabe ao certo quais são os seus limites ótimos de desenvolvimento e/ou como identificar as suas fases de desenvolvimento ao longo das variáveis derivadas calculadas. Por exemplo, o modelo de desenvolvimento utilizado pelos produtores do projeto para a cecidómia não está definido na literatura para a *região do Oeste* e, conseqüentemente, as fases identificadas por esse modelo nem sempre se ajustam às observações fitossanitárias registadas nos POBs. Na figura 7.26 é apresentado um *dashboard* semelhante ao da secção anterior, mas que acrescenta, sob as observações fitossanitárias, as fases identificadas pelo modelo de desenvolvimento adotado, permitindo observar o seu ajuste às observações registadas ao longo dos graus dia acumulados em 2020, para um zero de desenvolvimento igual a 0°C (estabelecido pelo mesmo modelo).

Para ajustar modelos de desenvolvimento é necessário identificar em primeiro lugar os momentos críticos do desenvolvimento dessas pragas, isto é, o(s) instante(s) ou intervalo(s) de tempo em que ocorrem algumas das fases do seu ciclo de vida, nomeadamente aquelas onde é crucial aplicar medidas de combate para evitar perdas económicas na produção da cultura. No caso da cecidómia, o modelo adotado pelo projeto identifica três momentos críticos (1ª, 2ª e 3ª geração) associados à acumulação de graus-dia, no entanto não identifica de que forma estes se manifestam nas observações fitossanitárias registadas. Nesse sentido, realizou-se um levantamento dos momentos críticos para a cecidómia através do registo das datas ou intervalos de tempo em que os técnicos de produção consideram que estes ocorrem.

7.3. IMPACTO DA METEOROLOGIA NA FITOSSANIDADE

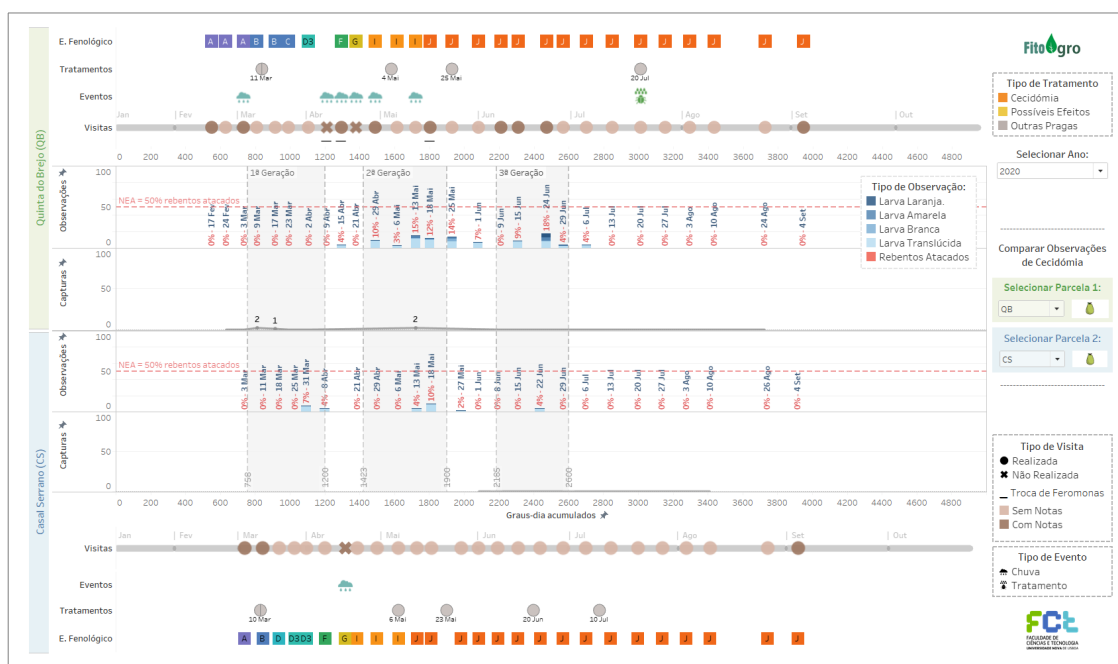


Figura 7.26: *Dashboard* da figura 7.25 com o modelo de desenvolvimento da praga visível sob as observações fitossanitárias registradas.

Neste levantamento apenas se apresentaram visualizações das observações de cecidómia e do estado fenológico da cultura. Não se especificou quais os tipos de momentos a registrar, para permitir extrair uma maior variedade de opiniões e conhecimento dos técnicos. No anexo I apresenta-se parte desse documento (apenas para um POB) preenchido por um dos técnicos de produção. Note-se que o preenchimento deste documento foi solicitado a um técnico de produção de cada entidade, sendo que de 8 entidades obteve-se resposta de 4 técnicos de produção. Posteriormente, digeriu-se toda a informação recolhida e discutiu-se com os técnicos as divergências encontradas para chegar a um consenso na identificação da ocorrência de momentos críticos de cecidómia. Em suma, concluiu-se que o primeiro aparecimento da praga identifica o início da 1ª geração e que o aparecimento de maioritariamente larvas translúcidas após uma visita com presença de larvas laranja indica o início de uma nova geração. No entanto, poderá haver sobreposição de gerações o que dificulta essa percepção.

Após a identificação dos momentos críticos, foi necessário proceder à parametrização de modelos preditivos dos momentos críticos, procurando o melhor ajuste possível destes às observações fitossanitárias registradas. Este trabalho foi realizado no contexto da dissertação de L. Correia (*unpublished*) [50], tendo sido construídas visualizações para auxiliar nesse processo, no entanto não serão apresentadas, uma vez que extravasam o âmbito deste documento. Em suma, nesse processo selecionam-se as parametrizações para o cálculo dos graus-dia acumulados que melhor se ajustam às observações fitossanitárias registradas nos diferentes POBs (através do erro quadrático médio) e utilizando diferentes combinações para os tipos de observação envolvidos na parametrização. Realizou-se esse processo para prever o início de observações (i.e. o início do ataque da 1ª geração)

e obteve-se um conjunto promissor de modelos preditivos que utilizam diferentes parametrizações e tipos de observação. Esse conjunto consiste nos dois modelos por cada tipo (ou conjunto de tipos) de observação que contêm o menor erro quadrático médio, sendo que para um deles considerou-se apenas zeros de desenvolvimento entre 7 e 10°C (intervalo considerado mais adequado pelos cientistas do projeto) e para o outro não se impôs restrições na parametrização.

Finalmente, é necessário observar o ajuste desse conjunto de modelos a todas as observações fitossanitárias registadas, para selecionar o modelo que melhor se ajusta, com o menor intervalo de previsão possível e com uma previsão por antecipação (i.e. prever antes do momento acontecer e nunca depois). Para isso, construiu-se uma visualização que permite, de uma forma simples, observar o ajuste desse conjunto de modelos às observações registadas num POB (figura 7.27) ou de um desses modelos às observações de vários POBs (figura 7.28), ao longo de um ano.

Por exemplo, na figura 7.27 é possível observar o ajuste dos modelos às observações fitossanitárias registadas na POB Quinta do Brejo, tendo-se selecionado aquele que prevê que o início do ataque ocorra entre 9 e 19 de Março de 2020, utilizando um zero de desenvolvimento de 7.2°C e as observações registadas nas parcelas cujo inimigo principal é a cecidómia, juntamente com todas as capturas em armadilhas.

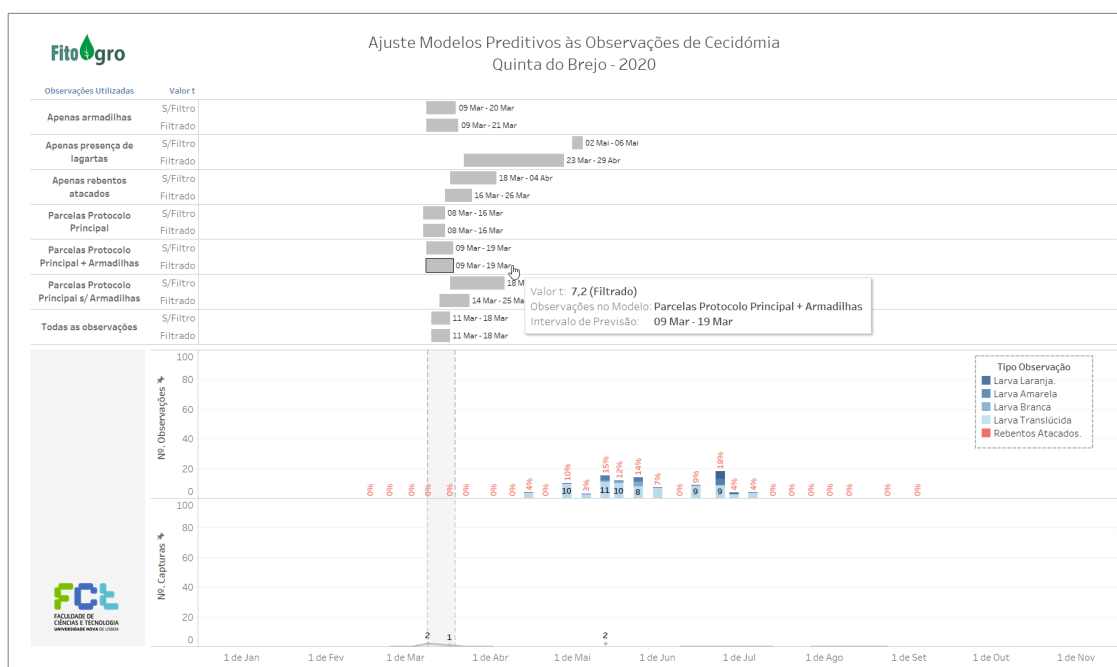


Figura 7.27: Visualização para observar o ajuste dos modelos preditivos de cecidómia às observações fitossanitárias registadas num POB (Quinta do Brejo) em 2020.

Similarmente, na figura 7.28 observa-se o ajuste da previsão desse mesmo modelo às observações registadas ao longo de 2020, em todos os POBs. Note-se que as visualizações das observações de campo são as mesmas das secções anteriores para a cecidómia, mas simplificadas para facilitar a observação do ajuste dos modelos.

7.3. IMPACTO DA METEOROLOGIA NA FITOSSANIDADE

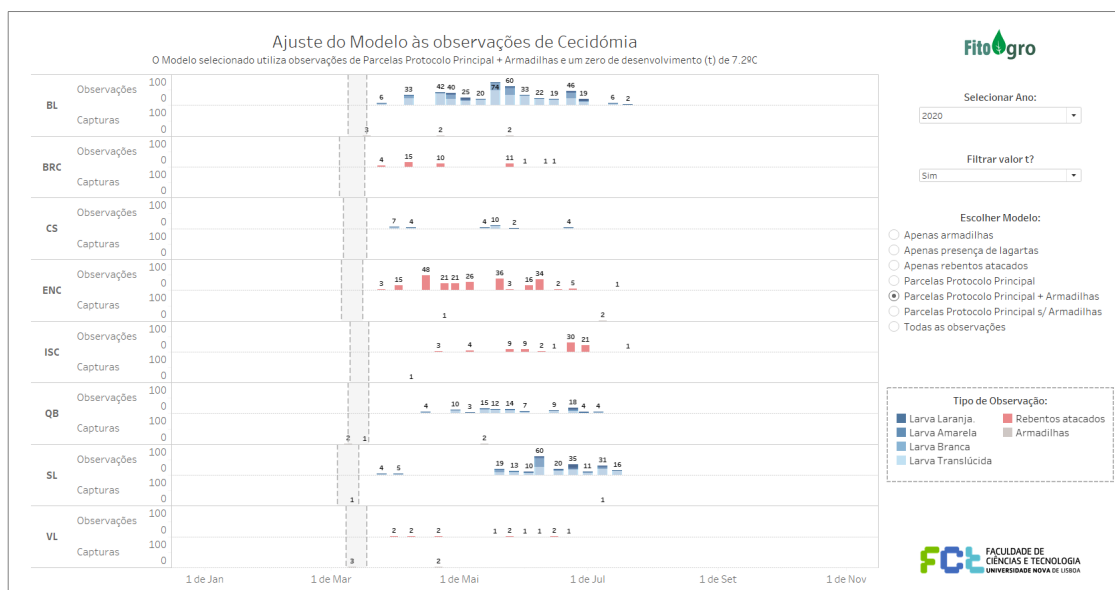


Figura 7.28: Visualização para observar o ajuste de um dos modelos preditivos de cecidómia às observações fitossanitárias registadas em todos os POBs.

Estas visualizações permitem assim validar o ajuste dos modelos preditivos às observações fitossanitárias registadas, como imposto pelo requisito **AD R8** e gerando confiança dos *stakeholders* nesses modelos. Adicionalmente, servem de instrumento de validação do ajuste de modelos preditivos a observações fitossanitárias, uma vez que são facilmente adaptadas para permitir a integração de mais momentos críticos e/ou parametrizações.

No futuro, estes modelos utilizarão dados de previsão meteorológica juntamente com o histórico das observações fitossanitárias registadas em cada local para prever os momentos críticos. Nessa altura, a visualização apresentada inicialmente nesta secção (figura 7.26) é facilmente adaptada para representar a previsão desses modelos, tornando-se um instrumento de suporte à tomada de decisão dos produtores.

AVALIAÇÃO E DISCUSSÃO

Neste capítulo apresenta-se a metodologia de avaliação utilizada (secção 8.1), o questionário realizado para avaliar a acessibilidade e usabilidade da componente (secção 8.2) e adaptações de soluções desenvolvidas a outros contextos (secção 8.3). No final do capítulo são discutidas algumas conclusões da avaliação (secção 8.4).

8.1 Metodologia de Avaliação

O desenvolvimento das soluções que integram a componente seguiu o processo iterativo de avaliação recomendado em 5.2. Este processo iniciou-se com a elicitação de requisitos aos *stakeholders* do projeto FitoAgro, seguida de uma avaliação iterativa onde se apresentou diferentes soluções para serem discutidas e iteradas com cada um dos *stakeholders*, até se obter as versões finais apresentadas neste documento. Esta abordagem permitiu construir solução intuitivas e que atendessem todas as necessidades analíticas elicidadas. Contudo, apenas através deste processo não é possível validar e medir a usabilidade e acessibilidade desta componente, pelo que se criou um questionário com esse propósito para ser preenchido pelos diferentes *stakeholders*, como será apresentado secção seguinte.

8.2 Questionário

A avaliação e validação da usabilidade e acessibilidade da componente analítica desenvolvida foi realizada através de um questionário para ser respondido pelos diferentes *stakeholders*. Este foi dividido em seis secções, onde em cada uma se apresenta exemplos de visualizações, seguidas de questões acerca das mesmas e de algumas afirmações para avaliar, através de escalas de *Likert*, a apreciação pela solução apresentada. O primeiro conjunto de questões pretende avaliar a acessibilidade das soluções, enquanto que as afirmações para apreciação validam a sua usabilidade, nomeadamente a satisfação dos

requisitos analíticos correspondentes. Na tabela 8.1 apresenta-se por secção o número de questões e quais foram os requisitos avaliados.

Secção	Nº Questões	Requisitos
S1 - Monitorização e Controlo dos Dados Meteorológicos	13	MCQ 1 a 3
S2 - Analítica da Meteorologia	10	AD 1 e 2
S3 - Monitorização e Controlo dos Dados de Campo	5	MCQ 4
S4 - Fitossanidade	11	AD 4 a 6
S5 - Variáveis Derivadas da Meteorologia	9	AD 3
S6 - Impacto da Meteorologia na Fitossanidade	2	AD 7 e 8

Tabela 8.1: Nº Questões por secção e enumeração dos requisitos analíticos avaliados.

As questões podem requer uma ação de contagem, leitura ou perceção da visualização apresentada. Por exemplo, uma contagem requer que o utilizador identifique na visualização a informação relevante conte o número de valores ou instâncias correspondentes. Por exemplo, a pergunta número 1 "Quantas estações estão *offline*?" requer que se conte o número de estações meteorológicas na visualização 6.1 que estão a vermelho (indicador de *offline*). Por outro lado, uma leitura requer que o utilizador identifique na visualização o local onde a informação está explicitada e que a reescreva como resposta. Por exemplo, a pergunta número 5 "Quando foi enviado o último registo da estação selecionada?" requer que se faça uma leitura da *tooltip* da visualização anterior, identificando a data do último registo. Por fim, perguntas que requerem perceção são perguntas cuja resposta não é identificada diretamente na visualização através de contagens ou leitura, exigindo uma análise e assimilação da visualização de forma a chegar à resposta pretendida. Por exemplo, a pergunta 11 "Por volta de que hora se deu o pico da temperatura e depressão do ponto de orvalho?" requer que se percecione a visualização 6.6 identificando o alinhamento vertical do eixo temporal entre os sensores e verificando.

Para além disso, o questionário apenas aceita resposta de escolha única ou aberta, consoante a pergunta. A resposta de escolha única caracteriza-se por um conjunto de opções pré estabelecidas, onde apenas uma pode ser selecionada. Por outro lado, a resposta aberta permite qualquer tipo de resposta desde que seja do tipo pretendido, isto é, se a resposta for um número inteiro então só se aceita números inteiros. Note-se que o objetivo das respostas abertas não é obter opiniões ou justificações, mas sim respostas que não sejam induzidas por opções pré-estabelecidas. Este tipo de respostas foi na sua maioria utilizado para perguntas que requerem contagens e leituras, de modo a garantir que as respostas fossem imparciais.

Por fim, as afirmações para apreciação têm como objetivo determinar, através de uma escala de *Likert*, o nível de satisfação global com a componente, bem como dos requisitos analíticos descritos no capítulo 4. As afirmações foram feitas pela positiva, isto é, com a formulação "Através desta visualização sou capaz de...", sendo que o nível 1 (mais baixo) corresponde a "Discordo Totalmente" e o 5 (mais alto) a "Concordo Totalmente".

No apêndice D apresenta-se, para cada secção do questionário, as visualizações apresentadas, seguidas das perguntas e afirmações correspondentes.

Dada a extensão do questionário e uma vez que nem todos os *stakeholders* têm conhecimento e/ou interesse acerca de todas as temáticas, optou-se que, no geral, estes respondessem apenas às secções que incluíssem os requisitos analíticos que lhes foram atribuídos no capítulo 4. De forma excepcional, considerou-se que os participantes de *computer science* deveriam ser capazes de responder a todas as secções.

Para simplificar o processo de participação no questionário, este foi realizado através do *google forms*. No entanto, através deste apenas se apresentou imagens das visualizações, perdendo-se o aspeto de interação da componente. Embora este seja muito importante para a componente e para a sua respetiva avaliação, a maioria dos *stakeholders* demonstrou, ao longo do processo iterativo de desenvolvimento, ser capaz de analisar as visualizações através de imagens. Deste modo, considerou-se que, para a análise de resultados, seria desnecessário o *overhead* introduzido pela interação com filtros, legendas, etc.

8.2.1 Análise de Resultados

O questionário contou com a participação de 13 *stakeholders*, dos quais 1 gestor de projeto (7,7%), 5 técnicos de produção (38,5%), 2 cientistas (15,4%), 3 *computer scientists* (23,1%) e 2 técnicos de recolha de campo (15,4%). Cada participação no questionário foi acompanhada através de uma sessão remota, onde foi dado suporte técnico e realizadas anotações relativamente às interpretações e dificuldades demonstradas. Note-se que, por isso, a análise de resultados utiliza os dados obtidos, mas a sua explicação poderá incluir as interpretações e anotações realizadas durante as sessões.

Relativamente aos resultados obtidos, podemos observar através da figura 8.1 que, à exceção da secção 1, a esmagadora maioria dos *stakeholders* respondeu corretamente às perguntas do questionário. Na secção 1 a amostra de resultados é consideravelmente inferior, sendo que apenas responderam *stakeholders* de *computer science* e gestão do projeto, para um total de 4 participantes. Ainda assim, apenas uma pergunta da secção teve 2 respostas incorretas enquanto que as restantes respostas foram todas corretas.

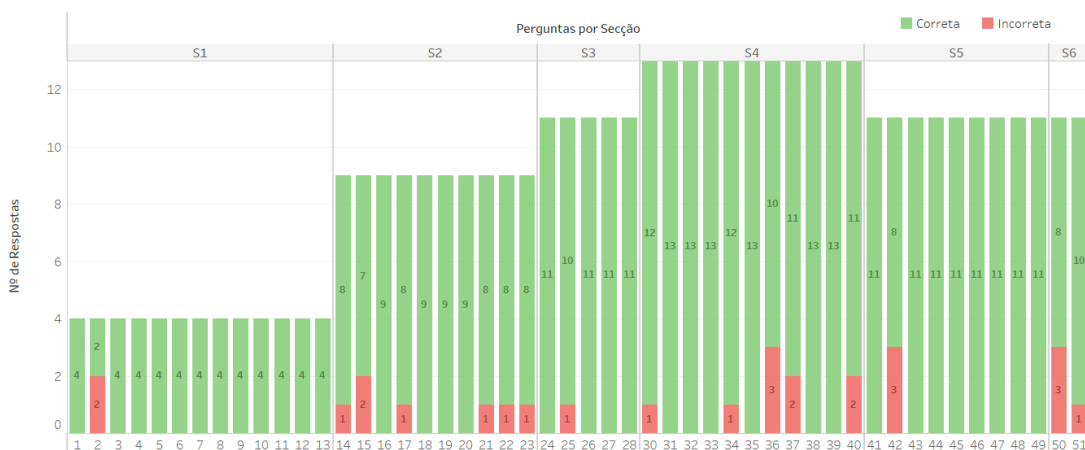


Figura 8.1: Número de respostas corretas/incorretas por pergunta.

CAPÍTULO 8. AVALIAÇÃO E DISCUSSÃO

Se analisarmos os resultados em termos percentuais, a média de respostas corretas por pessoa é de 94.2% e por pergunta é de 94.6%, pelo que os resultados foram muito positivos. No gráfico da figura 8.2, podemos observar a percentagem de respostas corretas/incorrectas por pergunta, onde se nota que existiram algumas perguntas onde surgiu uma maior percentagem de respostas incorretas relativamente às restantes.

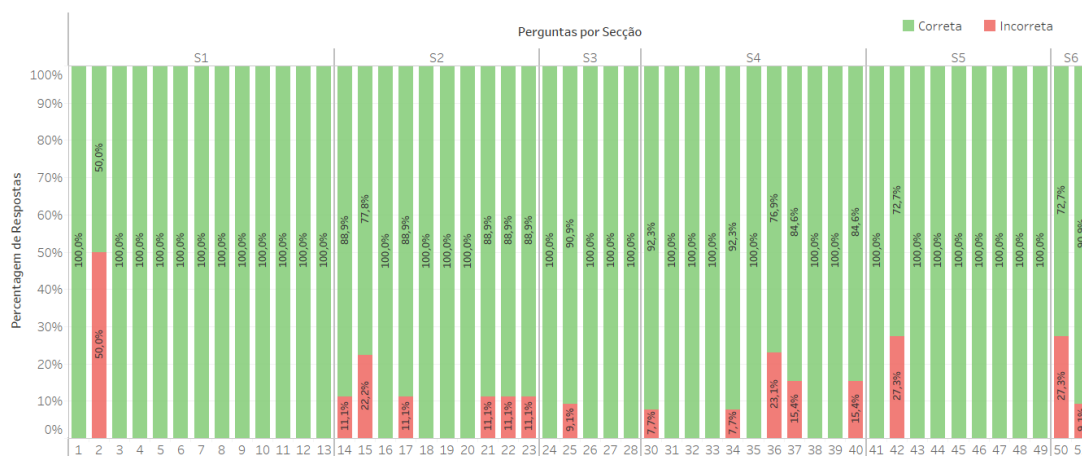


Figura 8.2: Percentagem de respostas corretas/incorrectas por pergunta.

Se considerarmos para análise detalhada as perguntas com mais de 20% de respostas incorretas, destacam-se as perguntas número 2, 15, 36, 42 e 50, apresentadas na figura 8.3 com indicação das diferentes respostas obtidas.

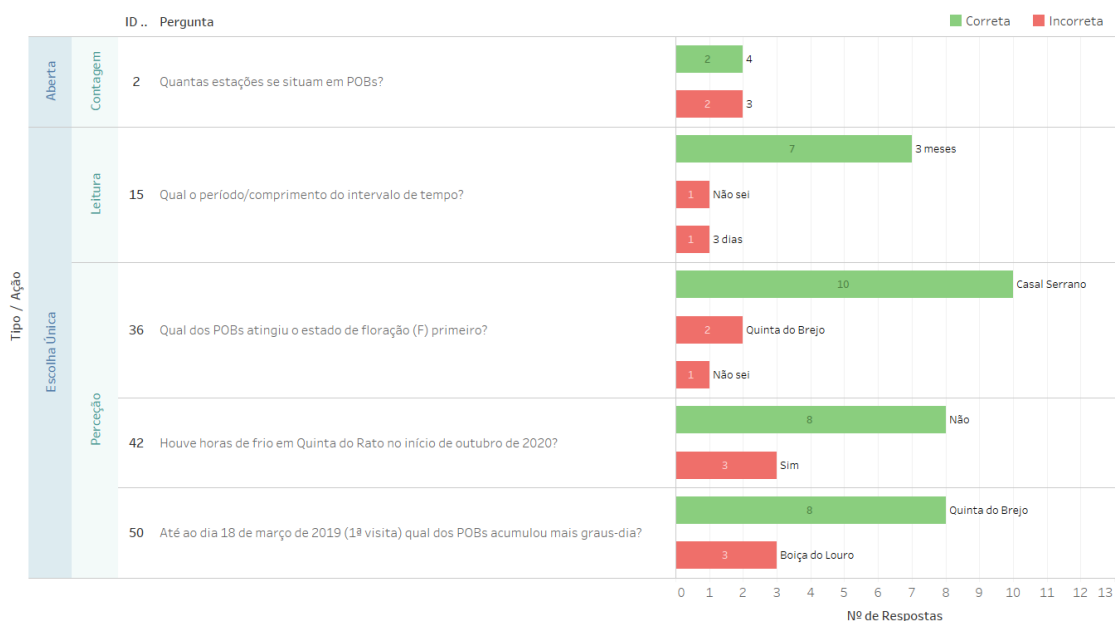


Figura 8.3: Respostas obtidas nas perguntas com mais de 20% de respostas incorretas.

A pergunta número 2 é de resposta aberta e requer a contagem do número de estações meteorológicas que se situam em POBs no mapa da visualização 6.1 apresentada no capítulo 6. Os valores das respostas obtidas variam apenas entre 3 e 4, sendo o último a resposta correta. Os participantes demonstraram perceber a visualização e a associação da cor e do formato das estações aos indicadores de estado e de localização em POBs, respetivamente. Contudo, a contagem do número de estações correspondentes foi por vezes incorreta, ainda que se considere que a discrepância de valores não foi significativa dado o número de estações apresentadas no mapa. Com interação esta problemática teria tido menos impacto, uma vez que ao clicar no(s) indicador(es) que se pretende a visualização automaticamente destaca as estações correspondentes.

Por outro lado, a pergunta número 18 é de escolha única (i.e. escolhe-se uma opção de várias possíveis respostas) e requer que o participante faça uma leitura no título da visualização 6.13, apresentada no capítulo 6. Alternativamente, o utilizador poderá observar a caixa de seleção do intervalo de tempo ou inferir através do eixo temporal. A resposta correta a esta pergunta é um intervalo de 3 meses, no entanto obteve-se uma resposta com 3 dias e outra onde o participante não soube responder. Visto que as opções de resposta são 1 dia, 3 dias, 3 meses e 6 meses, a primeira resposta incorreta poderá resultar de um lapso de atenção relativamente à resposta dada. Quanto à falta de capacidade de resposta, não é possível retirar conclusões, uma vez que representa apenas um sétimo das respostas.

Por sua vez, a pergunta 36 também é de escolha única, mas requer a perceção e compreensão do utilizador relativamente à visualização 7.20 (apresentada no capítulo 7) e aos dados que esta apresenta, particularmente neste caso do estado fenológico. A resposta correta a esta pergunta é Casal Serrano, no entanto 2 das respostas obtidas indicavam Quinta do Brejo e outra em que se foi incapaz de responder. Na generalidade os participantes do inquérito queixaram-se que a falta de interação dificultava a perceção do tempo na faixa do estado fenológico dada a sua distância do eixo temporal. Deste modo, era esperado que os resultados refletissem essas dificuldades, o que aparenta ter sido o caso.

A pergunta 42 é igualmente de escolha única e requer também perceção e compreensão, desta vez da visualização 6.17 apresentada no capítulo 6. A resposta que se considerou correta é a de que a afirmação é verdadeira, contudo surgiram dúvidas relativamente à formulação da pergunta, nomeadamente porque o termo "início de outubro" pode ser interpretado de diferentes formas. A pergunta refere-se aos primeiros dias do mês, mas houve quem assumisse que seria a primeira metade. Assim, considera-se que os resultados desta pergunta não avaliam corretamente a componente.

Finalmente, a pergunta número 50 é de escolha única e requer perceção e compreensão da visualização 7.24, apresentada no capítulo 7. A resposta correta a esta pergunta é Quinta do Brejo, no entanto houve 3 participantes que assinalaram Boiça do Louro como resposta. A utilização dos grau-dia como eixo é uma solução inovadora e de carácter estranho principalmente para aqueles que se encontram no meio agrónomo, uma vez que estes se regem muito pelo tempo. Assim, tendo em conta a natureza inovadora desta solução, considera-se que 3 respostas incorretas em 11 é um resultado satisfatório.

Contudo, espera-se que a interatividade da componente analítica ajude a atenuar a curva de aprendizagem desta solução, uma vez que através desta é possível obter informação detalhada acerca dos graus-dia acumulados até cada visita.

Uma vez que o tipo de ação e de resposta poderá ter influência no desempenho do questionário, construiu-se o gráfico da figura 8.4 para observar o número e respetiva percentagem de respostas corretas/incorrectas por tipo de resposta e ação. Através deste gráfico concluiu-se que apesar das perguntas de perceção representarem a maioria das perguntas do questionário, a sua percentagem de respostas incorretas não é mais elevada. Para os restantes tipos os resultados são positivos para o tamanho de cada amostra. Considera-se assim que o tipo de ação e de resposta não tiveram influência nos resultados.

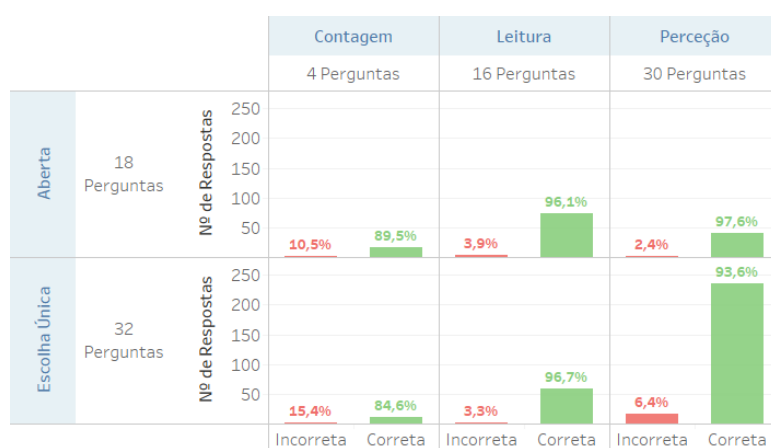


Figura 8.4: Respostas obtidas por tipo e pela ação que requer.

Para além disso, se analisarmos os dados em função do tipo de *stakeholder* envolvido obtemos o gráfico da figura 8.5, onde facilmente percebemos que as respostas incorretas distribuem-se similarmente entre os técnicos de produção e de recolha, bem como pela gestão do projeto. Ademais, através do gráfico é possível observar que todas as respostas obtidas por cientistas foram corretas e que menos de 2% das respostas de *computer science* foram incorretas.

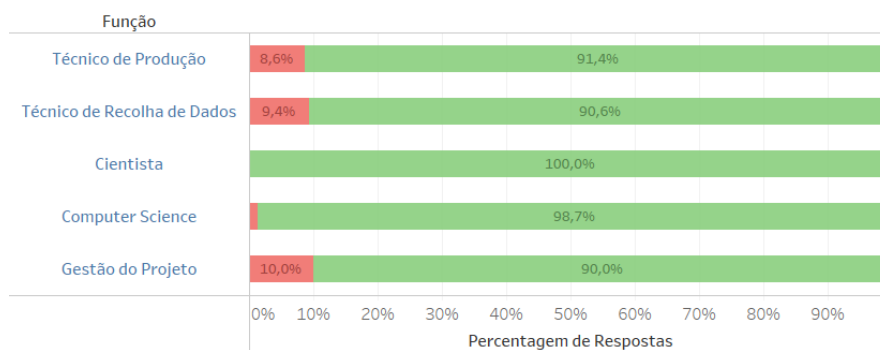


Figura 8.5: Percentagem de respostas obtidas por tipo de *stakeholder*.

Deste modo, podemos concluir que a análise de resultados indica que, no geral, a componente foi considerada acessível por todos os tipos de *stakeholders*.

8.2.2 Satisfação dos Requisitos Analíticos

Para avaliar a usabilidade da componente analítica desenvolvida, colocou-se no questionário afirmações para apreciação, cujas respostas são dadas através de escalas de *Likert*. Como referido anteriormente, o objetivo é determinar o nível de satisfação global com a componente, bem como dos requisitos analíticos descritos no capítulo 4.

Neste sentido, a figura 8.6 apresenta para cada requisito a percentagem de respostas obtidas para cada nível da escala de *Likert*. Através desta facilmente se conclui que a satisfação global dos requisitos analíticos é elevada, uma vez que apenas se obteve pontuações entre 4 e 5 na escala de *Likert*.

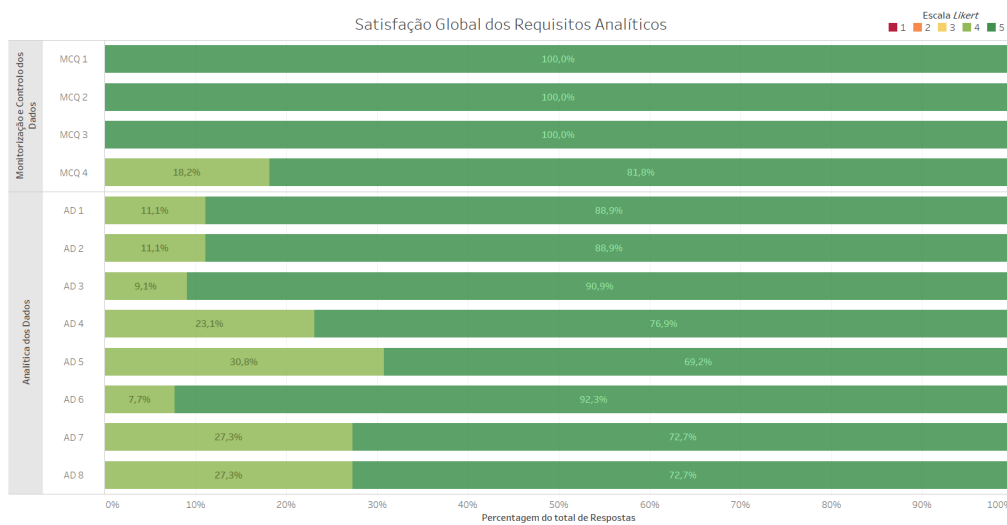


Figura 8.6: Percentagem obtida por requisito analítico, em cada nível da escala de *Likert*.

Ao classificar os requisitos consoante o tipo de informação que envolvem e o tipo de abordagem utilizada pela solução desenvolvida, obtemos requisitos com tipo de informação diversa ou única e soluções com abordagens clássicas, inovadoras ou mistas.

Relativamente ao tipo de informação, este caracteriza a complexidade do requisito no sentido em que quanto maior a diversidade da informação, mais complexa tendencialmente é a solução. Um exemplo de requisitos com tipo de informação única são os de monitorização e controlo de dados, enquanto que, por exemplo, a analítica de dados fitossanitários englobam informação diversa (e.g. observações, visitas, tratamentos, etc.).

Quanto à abordagem da solução, esta pode ser clássica, isto é, já utilizada como norma para o tipo de informação apresentada, pode ser inovadora (inversamente à clássica) ou ainda pode ser uma mistura das duas abordagens, isto é, introduzindo apenas alguns elementos inovadores para uma abordagem clássica. Exemplos de uma abordagem clássica são as soluções de monitorização de dados meteorológicos ao longo de um dia, enquanto que, por exemplo, a introdução da noção de estação virtual nesses dados já a torna numa solução mista. Por outro lado, considera-se como abordagens inovadoras a utilização de técnicas de alinhamento vertical entre diferentes tipos de dados, a substituição do eixo temporal por funções monótonas não-decrescentes, entre outros exemplos.

Se reordenarmos os requisitos da figura anterior de forma ascendente pela percentagem de avaliações de nível 4 e associarmos a cada requisito a classificação do tipo de informação e de abordagem, obtemos o gráfico da figura 8.7. Através deste, é possível observar que existe uma tendência na diminuição da satisfação quanto maior a diversidade do tipo de informação e quanto maior a inovação da abordagem utilizada.

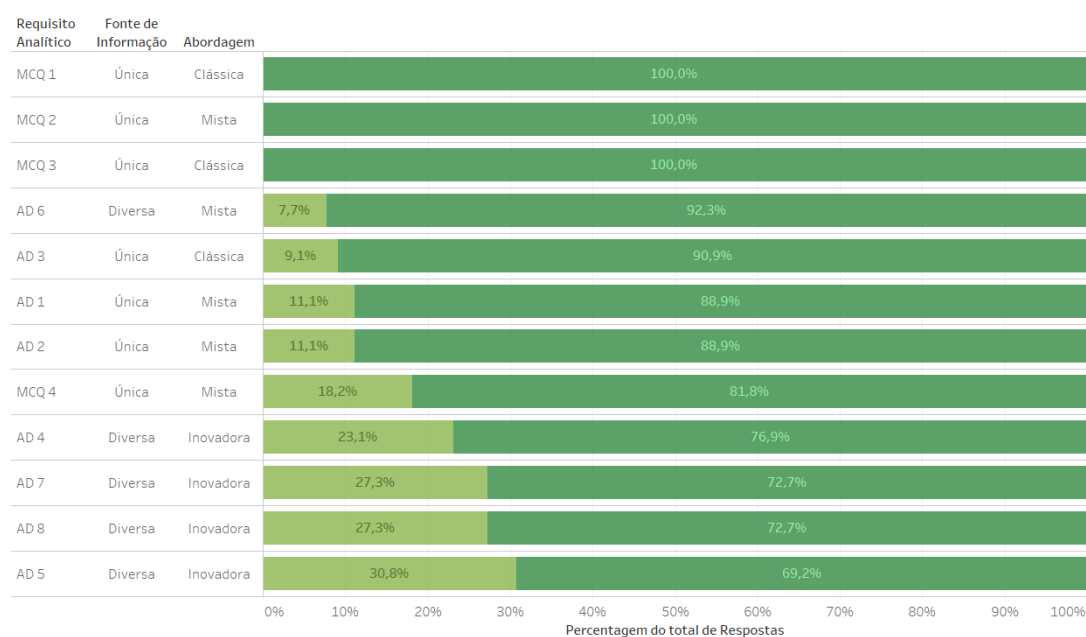


Figura 8.7: Percentagem obtida por requisito analítico, em cada nível da escala de *Likert* e ordenada de forma ascendente pela percentagem de avaliações de nível 4 e categorizado por tipo de informação e abordagem utilizada.

Podemos concluir que a componente sucedeu em termos de usabilidade e satisfação dos requisitos analíticos. Esperava-se que a opinião fosse mais dividida para as soluções com abordagens mais inovadoras e com tipos de informação diversa, mas embora isso se tenha verificado através de uma tendência na diminuição da satisfação, os resultados permaneceram muito positivos.

8.3 Adaptação a Outros Contextos

Uma vez que o foco da componente incidiu no projeto FitoAgro, não foi possível testar a adaptação desta a outros contextos para as várias áreas no âmbito da monitorização fitossanitária, como a recolha *in situ* e o ajuste de modelos preditivos. No entanto, facilmente se adaptou a componente desenvolvida para apresentar dados meteorológicos de outra fonte de dados obtida de forma externa ao projeto FitoAgro. Na figura 8.8 é possível observar a adaptação da componente desenvolvida aos dados meteorológicos do IPMA [47] referentes a 2018 para a área da grande Lisboa. Embora se utilize menos sensores, as regras utilizadas para os apresentar mantiveram-se consistentes, sendo que,

para além disso, foi possível disponibilizar a seleção do intervalo de tempo apresentado e do seu respetivo centro, bem como o nível de granularidade dos dados.

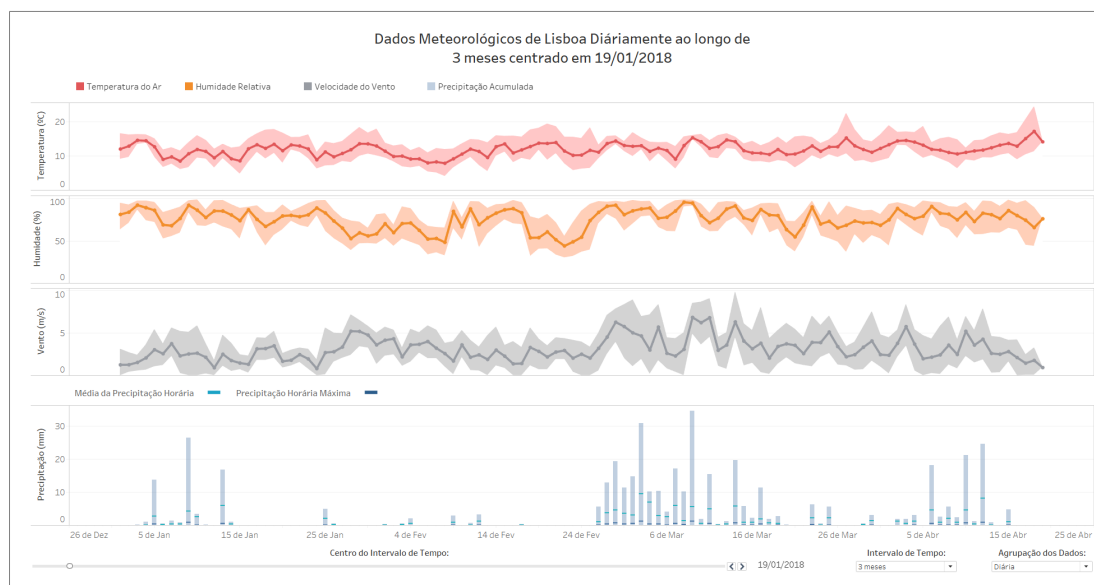


Figura 8.8: Dados Meteorológicos do IPMA referentes a 2018 para Lisboa, apresentados com a componente desenvolvida.

Embora este exemplo seja simples, mostra que a utilização da componente não está limitada aos dados do projeto FitoAgro, sendo que a sua reutilização ou adaptação para outros dados meteorológicos (idem para recolha *in situ* ou ajustes de modelos preditivos) é facilmente aplicável a outros contextos nesse âmbito.

Para além disso, tentou-se aplicar noutros contextos a solução mais inovadora da componente, a substituição do eixo temporal por uma função monótona não-decrescente temporalmente dependente. Para isso, selecionou-se para ensaio os dados do *Our World in Data* [58], relativamente à atual pandemia COVID-19, causada pelo vírus SARS-CoV-2. Considerou-se que esta seria uma temática interessante até porque um dos maiores problemas da análise de dados pandémicos surge devido ao desalinhamento temporal da evolução pandémica entre países.

Atualmente, uma das soluções para esse desalinhamento caracteriza-se por fixar âncoras temporais, como por exemplo utilizar o número de dias a partir do momento em que se alcança um determinado número de casos positivos confirmados. Na figura 8.9 apresenta-se um gráfico do *Our World in Data* com a evolução do número de casos positivos em Portugal (por milhão de habitantes) em relação aos restantes países, alinhados ao longo do número de dias a partir do 1º caso positivo confirmado em cada país.

Como proposta alternativa a esta solução, utilizou-se o conceito desenvolvido na componente analítica FitoAgro relativamente à substituição do eixo temporal por uma função monótona não-decrescente e aplicou-se aos dados da pandemia COVID-19. Para isso, utilizou-se o acumulado de casos positivos (por milhão de habitantes), para observar a evolução do número de pacientes nos cuidados intensivos (por milhão de habitantes).

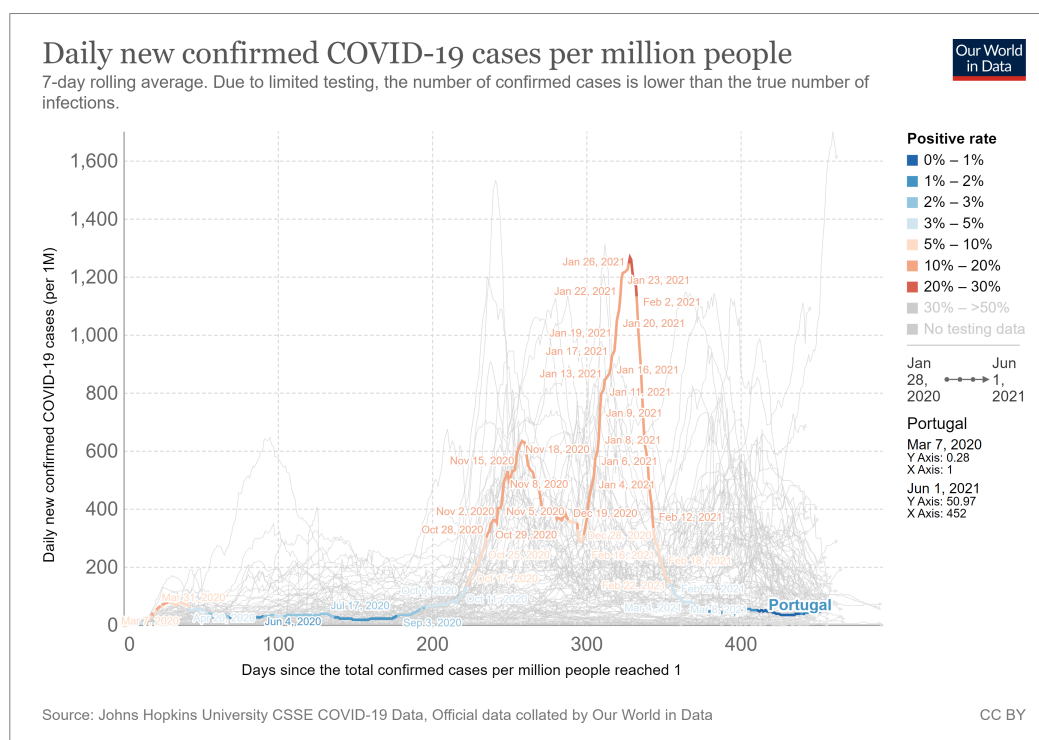


Figura 8.9: Evolução do número de casos positivos em Portugal em relação aos restantes países, ao longo do número de dias a partir do 1º caso positivo confirmado em cada país. Fonte: *Our World in Data* [58].

Em primeiro lugar, na figura 8.10 apresenta-se essa evolução e acumulado ao longo do tempo para Portugal e Itália. Estes países foram atingidos pela pandemia em momentos diferentes, verificando-se assim um desalinhamento temporal entre eles.

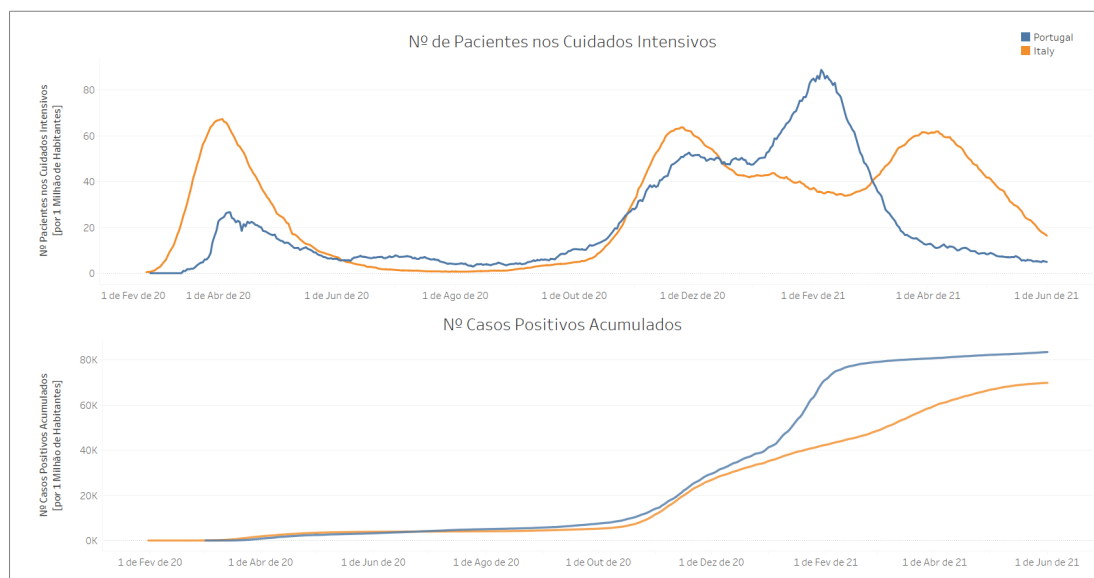


Figura 8.10: Evolução temporal do número de pacientes nos cuidados intensivos e do acumulado de casos positivos em Portugal e em Itália (por 1 milhão de habitantes).

Após esta primeira análise, substituiu-se então o tempo pelo acumulado do número de casos positivos, mantendo-se no eixo vertical a evolução do número de pacientes nos cuidados intensivos, em Portugal e em Itália (por 1 milhão de habitantes). Na figura 8.11 apresenta-se o gráfico que resulta desta transformação, sendo que, tal como no caso do FitoAgro, adicionaram-se ao *dashboard* faixas cronológicas para representar as contrações e dilatações do tempo ao longo do novo eixo.

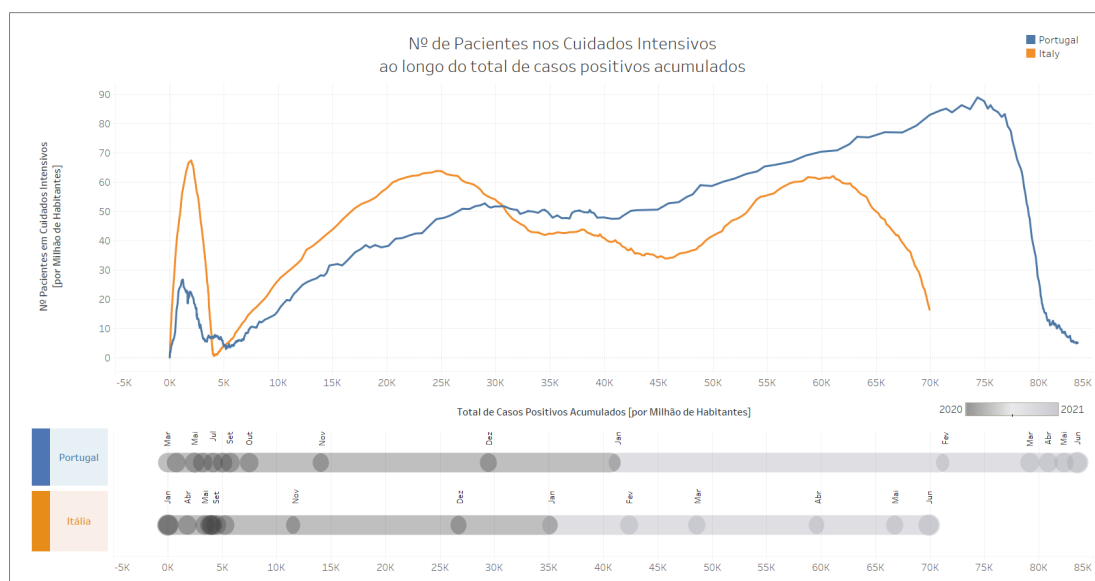


Figura 8.11: Evolução do número de pacientes nos cuidados intensivos ao longo do acumulado de casos positivos em Portugal e em Itália (por 1 milhão de habitantes).

Com este ensaio demonstrou-se que algumas soluções da componente desenvolvida são adaptáveis a contextos fora do âmbito da monitorização fitossanitária. Cada vez mais a problemática do desfasamento temporal se aplica a várias áreas na atualidade, tornando principalmente cada vez mais evidente na natureza e na meteorologia devido às alterações climáticas. Deste modo, a componente cumpre o seu objetivo ao utilizar soluções originais para promover a descoberta de correlações nos dados observados, independentemente do calendário, e adaptando-se assim facilmente outros contextos.

8.4 Conclusão

Os resultados dos questionários e a satisfação demonstrada verbalmente pelos *stakeholders* do grupo operacional FitoAgro permitem concluir que a componente analítica desenvolvida cumpriu os requisitos e as expectativas propostas. Ademais, estes resultados demonstram que o processo iterativo de avaliação é uma mais valia para o desenvolvimento de componentes de visualização interativa de dados, uma vez que este permite um desenvolvimento ágil ao longo da hierarquia dos *stakeholders*, promovendo a satisfação e a confiança destes nas soluções desenvolvidas. Contudo, através dos questionários foi possível observar a importância da interação para a visualização de dados, que, embora

não estivesse presente na avaliação feita por questionário, é um elemento omnipresente em todas as visualizações apresentadas.

Deste modo, é possível afirmar que a componente desenvolvida é considerada acessível e útil pelos *stakeholders* do grupo operacional, sendo que estes inclusivamente demonstraram um grande interesse e disponibilidade durante todo o processo, bem como posteriormente durante a avaliação e utilização da componente. Note-se que já existem *stakeholders* que têm acesso à componente e a toda a interatividade que esta oferece, nomeadamente a gestão de projeto, técnicos de campo e alguns técnicos de produção. Num futuro próximo pretende-se disponibilizar a componente aos restantes elementos do grupo operacional, para que esta possa continuar a servir no futuro como ferramenta de análise e de apoio à tomada de decisão.

Não foram realizados testes de desempenho dado que na plataforma utilizada (*Tableau*) o desempenho está diretamente ligado ao tamanho e estruturação dos dados. No entanto, podemos afirmar que no geral a componente apresenta os dados de forma quase instantânea, exceto os dados meteorológicos, devido ao seu volume elevado. O sistema de informação inclui dados meteorológicos de diversas estações (com vários sensores) ao longo de diversos anos, daí que o volume de dados seja mais elevado e que as visualizações (principalmente as que utilizam os dados originais) sofram de pequenos atrasos na apresentação dos resultados. Apesar disso, a visualização dos dados já processados é mais fluida, sendo que a que utiliza estações virtuais é instantânea, o que reafirma que o desempenho deve-se ao volume e estrutura dos dados e não à componente desenvolvida. Do ponto de vista do analítico isto é uma observação importante, sendo que enfatiza a importância de soluções de armazenamento de dados como o *publish* (apresentado em 5.3) para o desenvolvimento de componentes de visualização interativa de dados.

Por fim, a componente tinha como objetivo manter algum carácter genérico para que pudesse ser adaptada no futuro por outros projetos que integrassem dados do âmbito da monitorização fitossanitária, como dados meteorológicos, de recolha *in situ* ou ajustes de modelos preditivos. A componente analítica desenvolvida cumpre esse objetivo, sendo que estabelece regras e linhas de referência para o desenvolvimento de visualizações interativas de dados adaptáveis noutros contextos do âmbito. Ademais, demonstrou-se que algumas soluções originais desenvolvidas podem ser adaptadas para se reaplicarem noutros cenários, fora do âmbito da monitorização fitossanitária.

CONCLUSÕES E TRABALHO FUTURO

Este capítulo conclui o trabalho realizado ao longo desta dissertação (secção 9.1) e, para além disso, refere alguns aspetos que poderiam ser abordados futuramente para a modelação, desenho e construção de visualizações interativas de dados (secção 9.2).

9.1 Conclusão

Este trabalho teve dois grandes objetivos, sendo que o primeiro consistiu-se na delimitação de recomendações e regras gerais para a modelação, desenho e implementação de visualizações interativas de dados para contextos gerais e da monitorização fitossanitária. Para isso, criou-se uma linguagem de modelação específica para a visualização interativa de dados, o *IVML*, e, para além disso, estabeleceu-se um conjunto de regras gerais para o desenho de visualizações no âmbito da monitorização fitossanitária e áreas envolventes, bem como um processo de construção e avaliação iterativa destas.

Por outro lado, o segundo objetivo incidiu na elaboração de uma componente de visualização interativa de dados para o projeto FitoAgro, que fosse de fácil compreensão e utilização pelos seus *stakeholders*, que incorporasse soluções originais para a descoberta de correlações nos dados e cuja estrutura tivesse um carácter genérico no âmbito da monitorização fitossanitária e das áreas nesse contexto, nomeadamente da meteorologia, da recolha de dados *in situ* e de modelos de desenvolvimento e previsão. Para este objetivo, foi necessário, em primeiro lugar, fazer um levantamento dos requisitos analíticos para os diferentes *stakeholders* do projeto, para os quais se construiu visualizações interativas apropriadas. Por sua vez, incorporou-se nessas visualizações soluções originais para promover a correlação de dados, nomeadamente a construção de elementos que permitem a observação temporal simultânea de vários fatores relevantes, a substituição do eixo temporal por variáveis temporais derivadas da meteorologia, como os graus-dia acumulados,

entre outros. Finalmente, determinou-se estruturas base para a visualização de diferentes tipos de dados meteorológicos, de recolha *in situ* e de modelos de desenvolvimento e previsão, permitindo facilmente a sua adaptação para incorporar outros dados dentro dessas áreas.

Em suma, considero que os objetivos estabelecidos foram cumpridos com sucesso. Deixa-se assim contribuições para o desenho e modelação de visualizações interativas de dados, bem como um conjunto de normas e estruturas gerais que se espera que sejam utilizadas e adaptadas no futuro para outras componentes visuais no âmbito da monitorização fitossanitária e áreas envolventes.

9.2 Trabalho Futuro

Embora se tenham cumprido todos os objetivos propostos e ainda contribuído para a modelação de visualizações interativas de dados para além do contexto fitossanitário, alguns aspetos interessantes ficaram por explorar. Nesta secção são apresentadas algumas temáticas e soluções que poderiam ser aplicadas ou estudadas no futuro, de forma complementar ao trabalho desenvolvido ao longo desta dissertação.

9.2.1 Melhoria e extensão da linguagem IVML

A linguagem de modelação IVML, descrita em apêndice (B), ainda não se encontra aprimorada nem completa ao ponto de definir a modelação para todos os possíveis casos de uso da visualização interativa de dados. Para além disso, a ferramenta utilizada pela linguagem, a *diagrams.net* [59], é *open-source* e facilita a construção dos diagramas, no entanto necessita que os componentes da linguagem sejam importados como uma biblioteca externa escrita em *XML*, sempre que se inicie um novo diagrama. Finalmente, o foco da linguagem prevalece na modelação e documentação, sendo que a sua utilização para fins de desenho e apresentação ao utilizador final acaba por se tornar demasiado complexa.

Aspira-se que no futuro a linguagem continue a ser desenvolvida de modo a que esta possa satisfazer todas as necessidades específicas da visualização interativa de dados, tornando-se a linguagem *standard* para a modelação e desenho dessa área.

9.2.2 Ferramenta para a construção de visualizações

A componente desenvolvida estabelece regras e estruturas base para o desenho de visualizações no contexto fitossanitário. No entanto, a construção, replicação e/ou adaptação dessas visualizações requer um conhecimento *à priori* da plataforma *Tableau* ou outra tecnologia de visualização igualmente capaz.

Uma solução seria ter uma ferramenta que permitisse conceber, explorar e implementar visualizações com base nas regras e estruturas pré estabelecidas, sem necessitar de conhecimento técnico acerca da plataforma. Isto permitiria que um utilizador ou *stakeholder* fosse capaz de, durante o processo analítico, modificar ou adaptar essas visualizações ou até mesmo conceber outras.

Por exemplo, utilizaram-se diferentes tipos de abordagens para a análise dos dados meteorológicos de acordo com o tipo de dados fornecido pelos diferentes sensores. Se quiséssemos introduzir um outro sensor, a ferramenta poderia fornecer automaticamente uma visualização adequada ao tipo de dados do mesmo, nomeadamente se fossem valores acumulados seria um gráfico de barras e se fossem valores de instante seria um gráfico de linhas, sendo que se para cada instante existissem intervalos de valores (e.g. máximo e mínimo) adicionaria-se ainda uma banda de variação. Por outro lado, se, por exemplo, se alterasse ou adicionasse um protocolo ao projeto FitoAgro, através dessa ferramenta qualquer utilizador facilmente concebia visualizações para esse inimigo que relacionassem todas as informações fitossanitárias recolhidas e idem para os modelos de desenvolvimentos e de previsão.

A criação de uma ferramenta deste tipo poderia ser feita através de extensões ou *plugins* programados na própria plataforma do *Tableau* (Extensions API [37]), sendo que esta, por si só, já é uma plataforma com uma curva de aprendizagem pouco acentuada.

9.2.3 Visualizar a aplicação de modelos preditivos

A componente de visualização interativa de dados construída para o projeto FitoAgro sucedeu no cumprimento de todos os requisitos analíticos elicitados pelos seus *stakeholders*, nomeadamente permite a visualização do ajuste de modelos preditivos às observações fitossanitárias registadas. No entanto, o projeto tinha alguns objetivos adicionais que pretendia alcançar, dos quais o desenvolvimento de modelos preditivos com base em dados de previsão meteorológica e no histórico de dados observados em cada local. Para esse requisito, seria apenas necessário adaptar as visualizações da componente existente para incorporar os novos dados e modelos, como foi explicado em 7.3.2. Adicionalmente, outro objetivo do projeto seria a disponibilização de mapas ou boletins de risco para cada *praga*. Nesse caso seria necessário estender a componente com novas visualizações que permitissem observar a interpolação do risco na *região do Oeste*.

9.2.4 Reaplicabilidade do ajuste de componentes temporais ao longo de funções monótonas não-decrescentes temporalmente dependentes

A substituição do eixo temporal por valores acumulados de graus-dia (variável derivada da meteorologia) foi a solução mais inovadora desenvolvida para esta componente. Este princípio funciona no contexto fitossanitário, uma vez que a variável em causa é temporalmente dependente e reflete as condições climáticas em cada local, o que, por sua vez, está diretamente correlacionado com o desenvolvimento das *pragas*.

No entanto, seria interessante reaplicar este princípio para outros cenários, onde se iria validar o ajuste da componente temporal ao longo de uma variável temporalmente dependente, que se comporte como uma função monótona não-decrescente como é o caso dos valores acumulados. Por exemplo, na secção 8.3 do capítulo anterior, apresentou-se um ensaio onde se reaplica essa solução para observar o ajuste da evolução do número de pacientes nos cuidados intensivos devido à COVID-19 ao longo do número de casos positivos confirmados, em Portugal e em Itália (por milhão de habitantes). Esta reaplicação carece de validação científica, mas serve como ponto de partida para trabalho futuro nessa temática.

BIBLIOGRAFIA

- [1] M. Centeno. *FitoAgro*. URL: inovacao.rederural.gov.pt/grupos-operacionais/13-projectos-grupos-operacionais/87-fitoagro. Visitado a 27-12-2019.
- [2] *Lei 26/2013*. Diário da República. URL: dre.pt/pesquisa/-/search/260454/details/maximized. Visitado a 27-12-2019.
- [3] A. Malafaia. “Sistema de Informação para Projetos de Monitoração Fitossanitária”. Universidade Nova de Lisboa, 2018.
- [4] *Cropio*. URL: about.cropio.com/. Visitado a 29-01-2020.
- [5] *FiedClimate*. URL: metos.at/pb/fieldclimate/. Visitado a 29-01-2020.
- [6] *Protomate*. URL: agromais.pt/artigos.aspx?ID=311. Visitado a 29-01-2020.
- [7] F. Ferreira. “Cartas de Risco para Pragas e Doenças das Plantas”. Universidade Nova de Lisboa, 2016.
- [8] F. Pinto. “Sistema para acompanhamento da cultura do brócolo”. Universidade Nova de Lisboa, 2017.
- [9] A. Maia. *SAFEBRÓCOLO - Manual de Utilizador*.
- [10] *WiseCrop*. URL: wisecrop.com/pt/. Visitado a 27-01-2020.
- [11] M. Ward, G. G. Grinstein e D. A. Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition*. ISBN: 9781482257373. Visitado a 31-01-2020.
- [12] *John C. Hart*. URL: <https://ieeexplore.ieee.org/author/37301161800>. Visitado a 31-01-2020.
- [13] *Tableau*. URL: tableau.com/. Visitado a 31-01-2020.
- [14] S. Carpendale. *Visual Variables*. URL: pages.cpsc.ucalgary.ca/~saul/hci_topics/pdf_files/visual-variables.pdf. Visitado a 31-01-2020.
- [15] *Gapminder Tools*. URL: [gapminder.org/tools/?from=world#\\\$chart-type=bubbles](http://gapminder.org/tools/?from=world#\$chart-type=bubbles). Visitado a 31-01-2020.
- [16] T. Munzner. *Visualization Analysis & Design*. 2015. ISBN: 9781466508910.
- [17] *The Data Visualisation Catalogue*. URL: datavizcatalogue.com/index.html. Visitado a 02-02-2020.

BIBLIOGRAFIA

- [18] *Polar Curves*. URL: brilliant.org/wiki/polar-curves/. Visitado a 02-02-2020.
- [19] *Circular barplot*. URL: r-graph-gallery.com/circular-barplot. Visitado a 02-02-2020.
- [20] W. Aigner, S. Miksch, H. Schumann e C. Tominski. *Visualization of Time-Oriented Data*. 2011. ISBN: 978-0-85729-078-6.
- [21] *The TimeViz Browser*. URL: timeviz.net/. Visitado a 06-02-2020.
- [22] *Apache ECharts (incubating)*. URL: echartsjs.com/en/index.html. Visitado a 02-02-2020.
- [23] *Dynamic Population Mapping Using Mobile Phone Data*. URL: web.flowminder.org/case-studies/dynamic-population-mapping-using-mobile-phone-data. Visitado a 02-02-2020.
- [24] *Building the Global Heatmap*. URL: medium.com/strava-engineering/the-global-heatmap-now-6x-hotter-23fc01d301de. Visitado a 02-02-2020.
- [25] *How Climate Change May Affect the Plants in Your Yard*. URL: nytimes.com/interactive/2019/05/23/climate/plant-hardiness-zones-shifting.html. Visitado a 02-02-2020.
- [26] M. O. Ward, G. Grinstein e D. Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. 2nd Edition. 2015. ISBN: 9781482257373.
- [27] *Advanced Data Analytics Platform*. URL: looker.com/platform/overview. Visitado a 20-11-2021.
- [28] *Uploading Spreadsheet (Excel, Google Sheets) Data into Looker*. URL: help.looker.com/hc/en-us/articles/360023724114-Uploading-Spreadsheet-Excel-Google-Sheets-Data-into-Looker. Visitado a 20-11-2021.
- [29] *Business Intelligence (BI) Software That Goes Beyond | Sisense*. URL: sisense.com/product/business-teams/. Visitado a 20-11-2021.
- [30] *Data Discovery | Sisense for Cloud Data Teams*. URL: dtdocs.sisense.com/article/data-discovery. Visitado a 20-11-2021.
- [31] *Qlik Sense | Modern Cloud Analytics*. URL: qlik.com/us/products/qlik-sense. Visitado a 21-11-2021.
- [32] *Visualização de Dados | Microsoft Power BI*. URL: powerbi.microsoft.com/pt-pt/. Visitado a 21-11-2021.
- [33] *Power BI Custom Visuals - Data Visualization Tools | Power BI*. URL: powerbi.microsoft.com/en-us/developers/custom-visualization/. Visitado a 21-11-2021.
- [34] *Accelerate your business with analytics | Domo*. URL: <https://www.domo.com/business-intelligence>. Visitado a 21-11-2021.
- [35] *Custom Charts - Domo*. URL: domohelp.domo.com/hc/en-us/articles/360042924454-Custom-Charts. Visitado a 21-11-2021.

- [36] *Tableau Prep Builder & Prep Conductor: A self-service data preparation solution*. URL: tableau.com/products/prep. Visitado a 31-01-2020.
- [37] *Extensions API | Tableau*. URL: tableau.com/developer/extensions/. Visitado a 31-01-2020.
- [38] *Cecidómia*. URL: infoagro.cothn.pt/portal/index.php?id=1307. Visitado a 29-12-2019.
- [39] *CECIDÓMIA, Dasineura pyri Bouché*. URL: sapecagro.pt/internet/webteca/artigo.asp?id=170&url_txt=&link=. Visitado a 29-12-2019.
- [40] *Dasineura pyri – Plant Parasites of Europe*. URL: bladmineerders.nl/parasites/animalia/arthropoda/insecta/diptera/nematocera/cecidomyiidae/cecidomyiinae/lasiopteridi/dasineurini/dasineura/dasineura-pyri/. Visitado a 29-12-2019.
- [41] A. Martins. “As cochonilhas-algodão da vinha (*Hemiptera:Pseudococcidae*) na região do Oeste”. Instituto Superior de Agronomia, 2009. URL: repository.utl.pt/handle/10400.5/2053. Visitado a 14-01-2020.
- [42] *Filoxera*. URL: infoagro.cothn.pt/portal/index.php?id=1320. Visitado a 28-12-2019.
- [43] *Traça Oriental*. URL: infoagro.cothn.pt/portal/index.php?id=1879. Visitado a 25-01-2020.
- [44] M. Correia. “Registo de dados baseado em protocolos de recolha de informação”. Universidade Nova de Lisboa, 2020.
- [45] *Sistema Integrado de Metainformação - conceitos*. URL: smi.ine.pt/Conceito/Detalhes/2123. Visitado a 27-07-2021.
- [46] *APAS*. URL: apas.com.pt/. Visitado a 08-02-2021.
- [47] *Instituto Português do Mar e da Atmosfera*. URL: ipma.pt/. Visitado a 08-02-2021.
- [48] *Pessl Instruments*. URL: metos.at/. Visitado a 08-02-2021.
- [49] *TerraPro - Technologies*. URL: terra-pro.net/. Visitado a 08-02-2021.
- [50] L. Correia. “Sistema de Informação FitoAgro e Modelos Adaptativos para Pragas”. Universidade Nova de Lisboa, unpublished.
- [51] *IFML: The Interaction Flow Modeling Language | The OMG standard for front-end design*. URL: ifml.org/. Visitado a 18-06-2021.
- [52] *Horas de Frio (7,2°C)*. URL: infoagro.cothn.pt/portal/index.php?id=2792. Visitado a 09-02-2021.
- [53] D. Johnson, R. Bessin e L. Townsend. *Predicting insect development using degreedays*. URL: entomology.ca.uky.edu/files/efpdf1/ef123.pdf. Visitado a 29-01-2020.

BIBLIOGRAFIA

- [54] *Bichado - Graus dia*. URL: infoagro.cothn.pt/portal/index.php?id=3362. Visitado a 07-07-2021.
- [55] P. Damos e M. Savopoulou-Soultani. "Temperature-driven models for insect development and vital thermal requirements". Em: *Psyche: A Journal of Entomology* (2012). DOI: [10.1155/2012/123405](https://doi.org/10.1155/2012/123405). Visitado a 28-03-2021.
- [56] A. P. Nunes, C. Lucas e A. C. Ramos, eds. *Congresso Frutos 2020: Inovação e Estratégia para a Fruticultura Nacional*. Actas Portuguesas de Horticultura 34. 2020. ISBN: 978-972-8936-39-6.
- [57] C. Martins, C. Lucas e A. P. Simões, eds. *Livro Grupos Operacionais de Fruticultura*. COHTN-CC. 2021. Por publicar.
- [58] H. Ritchie, E. Mathieu, L. Rodés-Guirao, C. Appel, C. Giattino, E. Ortiz-Ospina, J. Hasell, B. Macdonald, D. Beltekian e M. Roser. "Coronavirus Pandemic (COVID-19)". Em: *Our World in Data* (2020). URL: <https://ourworldindata.org/coronavirus>. Visitado a 06-11-2021.
- [59] *Diagram Software and Flowchart maker Maker*. URL: diagrams.net/. Visitado a 15-09-2021.

A P Ê N D I C E



RESUMO DOS REQUISITOS ANALÍTICOS

Classificação	Requisitos Analíticos
Monitorização e Controlo de Qualidade (MCQ)	MCQ R1. Detetar estações avariadas e falhas de sensores para que se possa sinalizar a sua reparação.
	MCQ R2. Monitorizar a qualidade das estações e respetivos sensores ao longo do tempo, de forma a identificar estações propícias a falhas <i>versus</i> estações fiáveis.
	MCQ R3. Observar, ao longo do tempo, os dados em bruto dos diferentes sensores meteorológicos, para diferentes estações, identificando falhas através da ausência de dados ou inconsistências nos mesmos.
	MCQ R4. Verificar a correção e consistência os dados recolhidos em cada visita de campo.
Analítica dos Dados (AD)	AD R1. Observar a evolução dos valores registados pelos diferentes sensores meteorológicos para diferentes estações, granularidades e intervalos de tempo.
	AD R2. Observar a evolução dos valores registados pelos diferentes sensores meteorológicos nos locais de cada POB , para diferentes granularidades e intervalos de tempo.
	AD R3. Calcular e observar a evolução de variáveis derivadas da meteorologia relevantes para o desenvolvimento das pragas e culturas em estudo.
	AD R4. Observar, para cada protocolo ou inimigo, a evolução das visitas, observações fitossanitárias, estado fenológico , tratamentos e ocorrências meteorológicas registadas, promovendo a descoberta de correlações nos mesmos.
	AD R5. Comparar, para cada protocolo ou inimigo, a evolução de toda a informação fitossanitária em diferentes locais e/ou anos.
	AD R6. Calcular e apresentar, para cada inimigo, a respetiva percentagem de ataque e nível económico de ataque (NEA).
	AD R7. Correlacionar ou promover a descoberta de correlações entre a meteorologia (ou variáveis derivadas da mesma) e a informação fitossanitária registada, para diferentes locais e/ou anos.
	AD R8. Observar, para cada inimigo, o ajuste dos modelos preditivos existentes às observações fitossanitárias registadas.

Tabela A.1: Resumo dos requisitos analíticos levantados, classificados por tipo.

DESCRIÇÃO TÉCNICA DO IVML

O IVML - *Interaction Visualization Modeling Language* é uma linguagem de modelação gráfica para descrever *dashboards*, nomeadamente, os seus componentes, interações e ações de dados, em diagramas de fácil compreensão. Pode servir assim para desenhar e documentar visualizações interativas de dados numa linguagem específica desse âmbito.

B.1 Descrição da Linguagem

No geral, o IVML é composto por dois tipos de diagramas, com diferentes propósitos. O diagrama de interações, como o nome indica, pretende modelar a interação dos utilizadores com os elementos que compõem *odashboard* ou interações que possam existir entre os mesmos. Por outro lado, o diagrama de *tooltips* tem como objetivo representar apenas as *tooltips* associadas às visualizações. Note-se que qualquer interatividade ou navegação que exista com as *tooltips* também deverá ser modelada neste diagrama específico. A razão para a separação das *tooltips* dos restantes elementos deve-se à complexidade que este pode alcançar por si só, isto é, uma visualização ou *dashboard* pode conter diferentes tipos de *tooltips* que, por sua vez, podem conter visualizações e permitir navegar para outras ou visualizações ou páginas *web*. Quando o grau de complexidade não justifica a utilização dos dois diagramas, pode-se utilizar o de interações para representar também as *tooltips*.

No diagrama de interações é onde são representados todos os componentes do *dashboard*, formando assim a sua estrutura, isto é, cada componente deverá ser representado proporcionalmente na página e no local que iria ocupar na visualização final. Para além disso, este diagrama também deverá representar os diferentes tipos de interação disponíveis e, para cada interação, quais os componentes afetados e as ações de dados associadas. Na figura seguinte (B.1) apresenta-se um exemplo do esqueleto de um diagrama de interações, contendo uma interação.

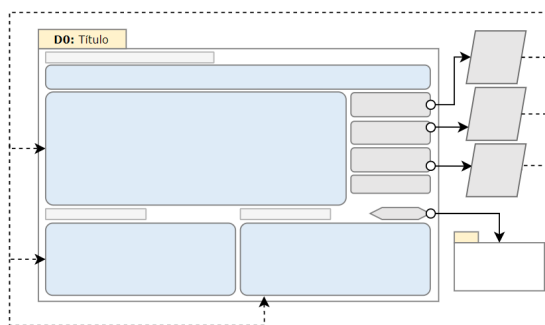


Figura B.1: Exemplo do esqueleto de um diagrama de interações, com três interações.

O diagrama de *tooltips* deverá manter como esqueleto a estrutura de componentes para associar as *tooltips* às visualizações correspondentes e, deste modo, manter uma noção espacial da visualização ou *dashboard* a ser modelado. Por defeito, as *tooltips* são automaticamente visíveis quando o utilizador sobrepõe o cursor na visualização, mesmo que não clique na mesma, no entanto, no caso de clicar, a tooltip apresentada poderá sofrer alterações, pelo que o tipo de interação associado à mesma deverá ser sempre especificado. Na figura B.2 apresenta-se o esqueleto do diagrama de *tooltips* para o exemplo apresentado na figura anterior.

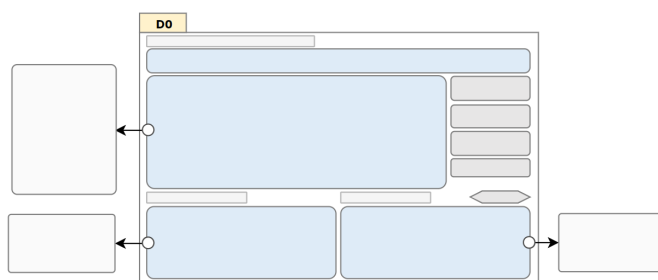


Figura B.2: Exemplo do esqueleto de um diagrama de *tooltips*, com três *tooltips*.

Os componentes da linguagem representam os diferentes elementos visualmente presentes na visualização ou *dashboard*, como por exemplo: o próprio *dashboard*, visualizações, *tooltips*, botões, filtros, parâmetros e legendas. Cada componente contém uma etiqueta que o identifica de forma única nos diagramas. Uma etiqueta é composta por uma letra e um número, sendo que a letra corresponde à inicial do tipo de componente (e.g. 'V' para visualização) e o número é único e sequencial para cada tipo de componente. Note-se que as etiquetas podem repetir-se entre *dashboards*, uma vez que estes também estão identificados por uma etiqueta, sendo que esta tem de ser única para a modelação da visualização interativa de dados de um *dataset*. Para além da etiqueta, maioria dos componentes são facilmente distinguíveis visualmente através da sua forma e cor, sendo que alguns partilham estas características. Dada quantidade extensa de componentes, estes foram divididos consoante o seu papel na linguagem de modelação, nomeadamente em componentes de visualização, de *layout* e de interação.

B.1.1 Componentes de Visualização

Os componentes de visualização estão diretamente ligados à visualização dos dados. Existem quatro componentes neste grupo, nomeadamente o *dashboard*, a visualização, a *tooltip* e a legenda, descritos de seguida.

- ***Dashboard***

O componente *dashboard* é a base de qualquer diagrama *IVML*, sendo que o seu principal objetivo é representar o *dashboard* que está a ser desenhado e/ou modelado. Para além disso, pode ser utilizado para representar um dashboard externo para onde se pode navegar através do atual.

Este componente é representado através de um retângulo branco com um pequeno retângulo amarelo acima do seu canto superior esquerdo, contendo a etiqueta do componente e o título do *dashboard* (exemplo na figura B.3). A proporção do retângulo branco deverá assimilar-se à do *dashboard*, uma vez que irá conter componentes para delinear a sua estrutura e conteúdo.



Figura B.3: Componente *Dashboard*.

- **Visualização**

O componente visualização é utilizado para representar visualizações, pelo que deverá ser colocado sempre dentro de um componente *dashboard* ou *tooltip*. É representado através de um retângulo azul de cantos arredondados, sendo que existem algumas variantes do mesmo que serão descritas na secção dos componentes de layout (B.1.2). No interior do componente deverá constar sempre a sua etiqueta seguida do título da visualização, num quadrado amarelo no canto superior esquerdo, e os dados associados à visualização, dentro de um retângulo branco no centro inferior do componente. os através de um retângulo branco com um pequeno retângulo amarelo acima do seu canto superior esquerdo, contendo a etiqueta do componente e o título do *dashboard* (exemplo na figura B.4).

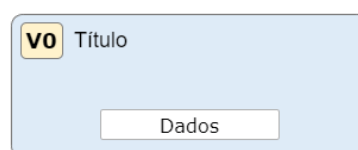


Figura B.4: Componente Visualização.

- **Tooltip**

O componente *tooltip* é utilizado para representar *tooltips* presentes no *dashboard*. Como já foi referido anteriormente, as *tooltips* estão sempre associadas a pelo menos uma visualização, sendo que a sobrepõem. Por isso, a linguagem delineou que este tipo de componente deverá ser colocado sempre fora do componente *dashboard*, associado à sua visualização pela interação correspondente. Note-se que este componente deve ser utilizado apenas no diagrama de *tooltips*, a não ser que em casos excepcionais seja simples o suficiente representar tudo num único diagrama.

Este componente é representado por um retângulo cinzento claro, de cantos arredondados, que deverá conter sempre a etiqueta correspondente, num quadrado amarelo no canto superior esquerdo, e os dados passados à *tooltip*, dentro de um retângulo branco centrado no inferior do componente (exemplo na figura B.5).



Figura B.5: Componente *Tooltip*.

- **Legenda**

O componente *legenda* é utilizado para representar as legendas presentes num *dashboard* ou *tooltip*, estando sempre associadas a uma visualização. A interação com este componente não é obrigatória dado que pode servir apenas como elemento visual, no entanto existem casos em que pode haver interações. Este componente é representado no diagrama através de um retângulo cinza, de cantos arredondados, cujo interior contém sempre a etiqueta correspondente, num quadrado amarelo no canto superior esquerdo do componente, seguida, opcionalmente, do título da legenda e finalmente, centrado no inferior do componente dentro de um retângulo branco, os dados passados à legenda (exemplo na figura B.6).

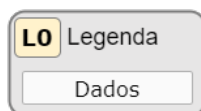


Figura B.6: Componente *Legenda*.

Os componentes de visualização podem ter dois tipos de atributos associados, nomeadamente para representar a utilização de variáveis visuais ou de diferentes tipos de gráfico. Esses atributos são colocados dentro dos componentes de visualização e são representados na linguagem através de uma variedade de ícones que será apresentados de seguida. Note-se que alguns destes ícones se basearam nos ícones do Tableau [13], dado que é a plataforma líder no mercado em visualização interativa de dados e a plataforma de referência dos autores da linguagem IVML.

- **Variável Visual**

As variáveis visuais são utilizadas para distinguir ou quantificar métricas diferentes dentro da mesma visualização ou entre visualizações no mesmo *dashboard*. Assim, para identificar as variáveis visuais principais, a linguagem disponibiliza os três ícones da figura B.8. Esses ícones devem ser utilizados como atributos de uma visualização e/ou legenda, representados, num tamanho reduzido, no canto superior direito desses componentes.

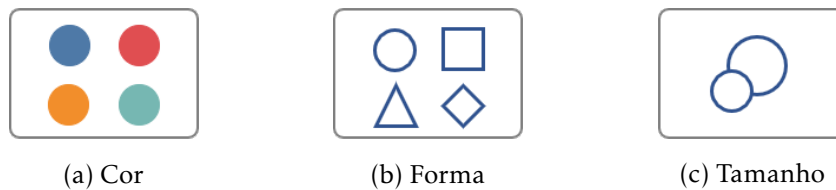


Figura B.8: Ícones para identificar variáveis visuais.

- **Tipo de Gráfico**

Cada visualização pode conter um ou vários tipos de gráfico em simultâneo. Assim, para detalhar a visualização e identificar o(s) tipo(s) de gráfico dentro desta, o *IVML* disponibiliza os ícones apresentados na figura B.10, para representar alguns dos principais tipos de gráficos. Estes ícones devem ser inseridos, em tamanho visível, no centro do componente de visualização.

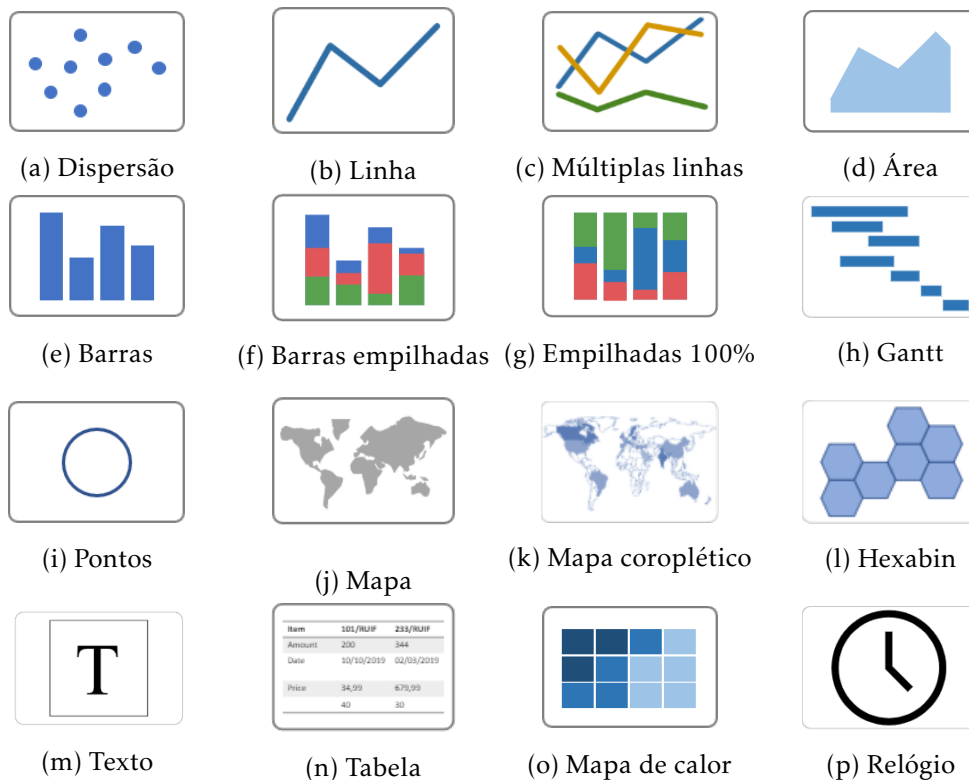
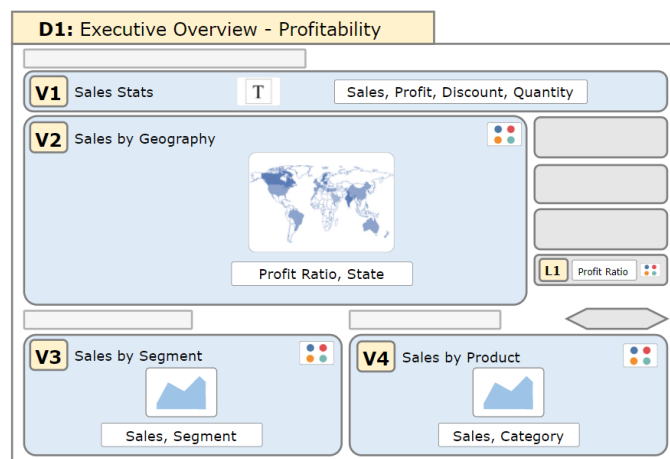


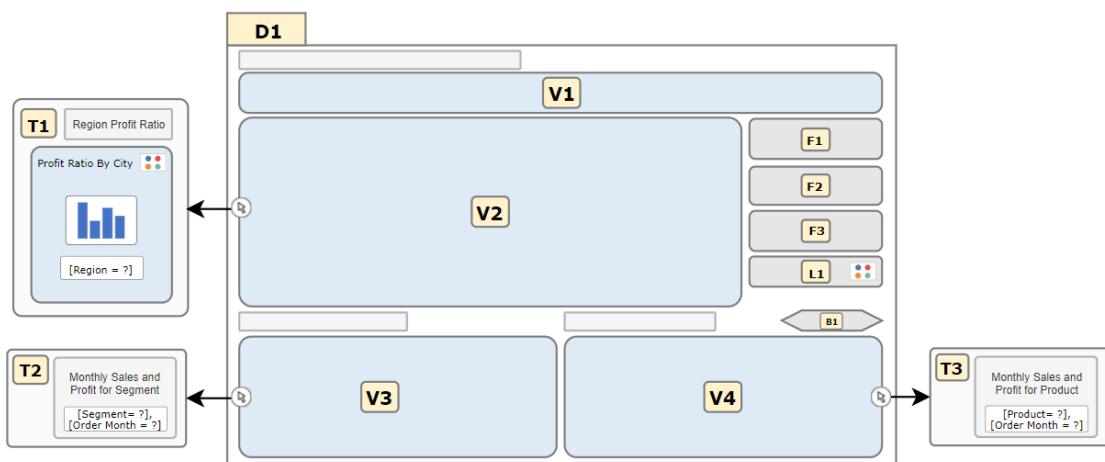
Figura B.10: Ícones para identificar o tipo de gráfico.

Quando existem vários gráficos dentro da mesma visualização e estes utilizam dados diferentes, poderá-se colocar um retângulo de dados debaixo de cada um dos ícones de gráfico, em alternativa ao retângulo de dados associado ao componente visualização.

Complementando os esqueletos dos exemplos nas figuras B.1 e B.2 com os componentes de visualização descritos acima, obtém-se os diagramas da figura abaixo (B.11), sendo que o de *tooltips* já se encontra completo. O diagrama de *tooltips* apenas mantém a estrutura de componentes do *dashboard*, colocando as etiquetas no centro de cada componente e removendo todos os títulos e os atributos não partilhados com as *tooltips* (e.g. na figura abaixo a cor da legenda L1 é partilhada com a *tooltip* T1). Note-se que quando há seleção de dados por interação com um componente, identifica-se esses dados no retângulo de dados com a mesma notação usada pela linguagem IFML. Por exemplo, no diagrama de *tooltips* abaixo, através da interação do utilizador é passada uma determinada região à *tooltip* T1, sendo que, por isso, consta nos dados do componente a notação '[Region = ?]'



(a) Diagrama de Interações incompleto



(b) Diagrama de Tooltips completo

Figura B.11: Exemplo de diagramas IVML apenas com componentes de visualização.

B.1.2 Componentes de *Layout*

Os componentes de *layout*, como o nome indica, especificam detalhes do *layout* do *dashboard*. De seguida apresenta-se os quatro componentes neste grupo, dos quais três são variantes do componente visualização descrito na secção anterior.

- **Visualização com eixo X partilhado**

O componente visualização com eixo X partilhado é uma variante do componente visualização onde, como o nome sugere, o seu eixo horizontal (X) é partilhado com outras visualizações verticalmente, nomeadamente acima ou abaixo. Este componente é usado quando se alinha várias visualizações verticalmente ao longo do mesmo eixo X (horizontal). A sua representação na linguagem IVML é feita através de um componente semelhante ao de visualização, mas cujos cantos inferiores (no caso de partilha com visualizações abaixo) ou superiores (no caso de partilha com visualizações acima) do rectângulo deixam de ser arredondados, como é apresentado nas figuras B.12a e B.12b, respetivamente.

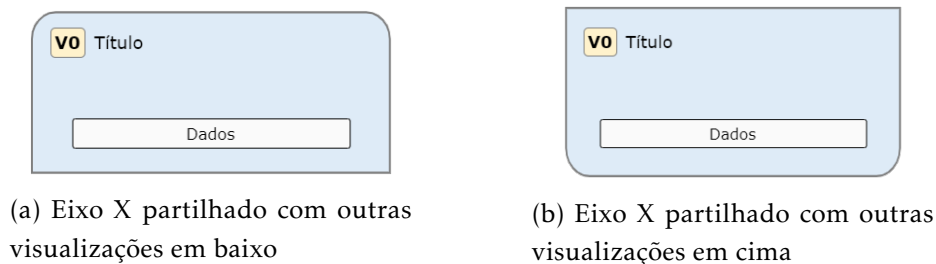


Figura B.13: Componente visualização com eixo X partilhado

- **Visualização com eixo Y partilhado**

O componente visualização com eixo Y partilhado é outra variante do componente visualização onde, como o nome sugere, o seu eixo vertical (Y) é partilhado com outras visualizações horizontalmente, nomeadamente à sua direita ou esquerda. Este componente é usado quando se alinha várias visualizações horizontalmente ao longo do mesmo eixo Y (vertical). A sua representação na linguagem IVML é feita através de um componente semelhante ao anterior, mas onde apenas os cantos do lado direito (no caso de partilha com visualizações à direita) ou do lado esquerdo (no caso de partilha com visualizações à esquerda) do rectângulo deixam de ser arredondados, como é apresentado nas figuras B.14a e B.14b, respetivamente.

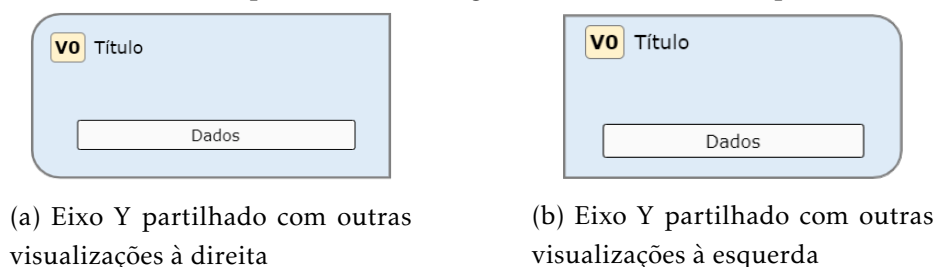


Figura B.15: Componente visualização com eixo Y partilhado

- **Visualização com eixos X e Y partilhados**

O componente visualização com eixos X e Y partilhado é a última variante do componente visualização e, como o nome sugere, tanto o eixo horizontal (X) como o vertical (Y) são partilhados com outras visualizações horizontalmente e verticalmente, respetivamente. Este componente é usado quando se alinha várias visualizações horizontalmente ao longo do mesmo eixo Y (vertical) e verticalmente ao longo do mesmo eixo X (horizontal). A sua representação na linguagem *IVML* é feita através de um componente semelhante aos anteriores, mas onde todos os cantos do rectângulo deixam de ser arredondados, como é ilustrado na figura B.16.

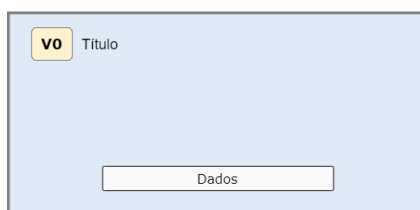


Figura B.16: Componente visualização com eixos X e Y partilhados.

- **Placeholder**

O componente *placeholder* é utilizado para representar elementos do *dashboard* que não são diretamente influenciados pelos dados ou não são relevantes para a interatividade. Deste modo, servem apenas para representar na estrutura do *dashboard* a presença de um elemento visual, como um texto ou imagem, complementar. Este componente é representado assim por um retângulo cizento claro, sem etiqueta e com apenas um título ou descrição breve (exemplo na figura B.17).



Figura B.17: Componente *Placeholder*.

Estes componentes não têm atributos adicionais, sendo que os componentes variantes do componente visualização podem conter os mesmos atributos desse componente. Por outro lado, o componente *placeholder* não pode conter nenhum atributo além do seu título ou descrição breve.

Ao adicionar os componentes de *layout* ao exemplo da figura anterior (B.11), obtém-se o diagrama de interação da figura abaixo (B.18), cuja única alteração é a adição de descrição nos componentes *placeholder*. Note-se que se omitiu o diagrama de *tooltips*, uma vez que já se encontrava completo na figura anterior.

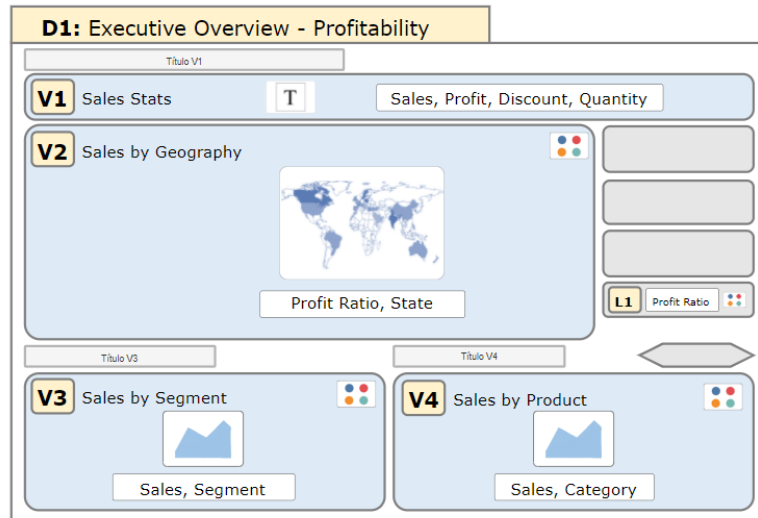


Figura B.18: Exemplo do diagrama de interações IVML incompleto, contendo apenas componentes de visualização e *layout*.

B.1.3 Componentes de Interação

Os componentes de interação, como o nome sugere, especificam a interatibilidade com elementos do *dashboard*. De seguida apresenta-se os cinco componentes deste grupo, nomeadamente o botão, o filtro, o parâmetro, a ação de dados e o *parameter binding*.

- **Botão**

O componente botão é utilizado para ilustrar botões ou outros elementos que ao clicar executem uma ação, normalmente de navegação. Este componente é representado por um hexágono cinzento escuro, alongado horizontalmente, cujo interior deve conter a sua etiqueta e um nome ou título breve, como ilustrado na figura B.19.



Figura B.19: Componente Botão.

- **Filtro**

O componente filtro é utilizado para representar filtros interativos presentes no *dashboard*. Este componente é representado por um retângulo cinzento escuro, cujo interior deve conter a sua etiqueta, um título breve e um retângulo com os dados que estão sob filtragem (exemplo na figura B.20).



Figura B.20: Componente Filtro.

- **Parâmetro**

O componente parâmetro, como o nome indica, é utilizado quando o *dashboard* contém parâmetros, isto é, elementos que permitem alternar a informação mostrada nas visualizações naquele momento conforme a opção selecionada (de numa lista ou booleano). Este componente, ilustrado na figura B.21, é visualmente semelhante ao do filtro, no entanto não contém um retângulo de dados, pois as opções nos parâmetros tanto podem pertencer ou não ao conjunto de dados.

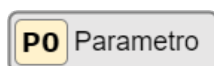


Figura B.21: Componente Parâmetro.

- **Ação de Dados**

O componente ação de dados é utilizado para representar ações executadas em segundo plano aos dados, mas que foram requisitadas pelo utilizador através de uma interação com outro componente, como por exemplo filtragens ou pesquisas. Este componente está, por isso, sempre associado a uma interação com um componente filtro ou legenda. Um componente ação de dados é representada por um paralelogramo cinzento escuro no exterior do componente *dashboard*, contendo no topo uma breve descrição da ação e em baixo um retângulo com os dados passados para executar a ação (com a notação IFML), como apresentado na figura seguinte (B.22). , Opcionalmente este componente poderá conter no centro um ícone a identificar o tipo de resultado que a ação tem nas visualizações (apresentados mais à frente), nomeadamente filtagem ou destaque dos dados.



Figura B.22: Componente Ação de Dados.

- ***Parameter Binding***

O componente *parameter binding* é utilizado de forma semelhante ao componente IFML com o mesmo nome, no entanto no caso do IVML apenas se aplicam à interação com parâmetros. Este componente representa a ligação entre os parâmetros e o conjunto de dados do *dashboard*, apresentado a ação executada sobre os dados através da selecção de parâmetros. Por isso, estes componentes estão sempre associados a uma interação com um componente parâmetro.

Na linguagem IVML, o componente *parameter binding* (figura B.23) é representado por um paralelogramo cinzento escuro no exterior do componente *dashboard*, contendo uma breve descrição no topo, as opções do parâmetro ao centro caso sejam

fora do conjunto de dados e em baixo um retângulo de dados contendo uma expressão que representa a ação executada sobre os dados (ou os próprios dados associados). Por exemplo, um parâmetro que altere a granularidade de uma métrica *date* de um conjunto de dados, poderia ser representado pela expressão $AGG(x, date)$ onde x é a granularidade selecionada, *date* é a métrica e *AGG* significa agregação. O *IVML* não especifica regras para a construção destas expressões, fornecendo liberdade na sua representação, no entanto devem ser consistentes dentro e entre diagramas e o seu significado bem documentado *à posteriori*.

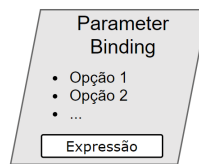


Figura B.23: Componente *Parameter Binding*.

Os componentes de interação podem ter três tipos de atributos associados, nomeadamente para representar tipos de interação, de navegação e de resultado das ações. Estes atributos são representados através de ícones, como apresentado de seguida.

- **Tipo de Interação**

Os atributos para os tipos de interação representam as ações realizadas pelos utilizadores ao interagirem com os elementos do *dashboard*. A linguagem suporta dois tipos de interação, nomeadamente clicar ou sobrepor o cursor/*hover*, cujos ícones são apresentados nas figuras B.24a e B.24b, respetivamente. Estes atributos devem ser representados na fronteira do componente onde ocorreu a interação e associadas ao início de setas contínuas (descritas mais à frente). Por vezes poderá ser necessário distinguir interações com diferentes gráficos dentro da mesma visualização, sendo que, nesses casos, permite-se a colocação dos ícones de interação na fronteira dos ícones de cada tipo de gráfico.



(a) Clique



(b) *Hover*

Figura B.25: Ícones para os tipos de interação.

- **Resultado das Ações**

Este tipo de atributos representam, como o nome indica, o resultado das ações realizadas sobre os dados através de uma interação. A linguagem suporta dois tipos de resultados, nomeadamente filtragem e destaque, cujos ícones são apresentados nas figuras x e y. Estes ícones só podem ser colocados no centro de componentes de ação de dados (entre o título e os dados), representando o resultado dessa ação na visualização.



Figura B.27: Ícones para os tipos de resultados das ações.

- **Navegação**

Os atributos de navegação permitem representar a navegação entre *dashboards* ou para páginas *web* (URLs) exteriores, cujos ícones são ilustrados nas figuras B.28a e B.28b, respetivamente. Estes ícones devem ser colocados junto a uma seta contínua (descritas mais à frente), entre um ícone de tipo de interação e, consoante o tipo de navegação, um componente *dashboard* ou um retângulo branco contendo o URL externo.



Figura B.29: Ícones para os tipos de navegação.

Os diferentes componentes de interação conetam-se a outros componentes através de setas de dois tipos: tracejadas e contínuas. A seta contínua deve ser utilizada sempre que a ligação entre os componentes é desencadeada por uma interação do utilizador, iniciando-se sempre, por isso, num ícone de interação. Por outro lado, a seta tracejada deve ser utilizada quando a ligação entre os componentes é impercetível pelo utilizador, i.e. quando é executada em segundo plano pelo sistema, sendo que, por isso, tem sempre início em componentes ação de dados ou *parameter binding*. Na figura B.31 apresenta-se os dois tipos de setas suportadas pela linguagem.



Figura B.31: Tipos de setas utilizadas para conexão de componentes.

Finalmente, com a adição ao exemplo anterior (figura B.18) dos componentes de interação e de todos os atributos e setas necessários, obteve-se o diagrama de interação anterior da figura abaixo (B.32), completando o exemplo. Note-se novamente a omissão do diagrama de *tooltips*, por já se encontrar completo.

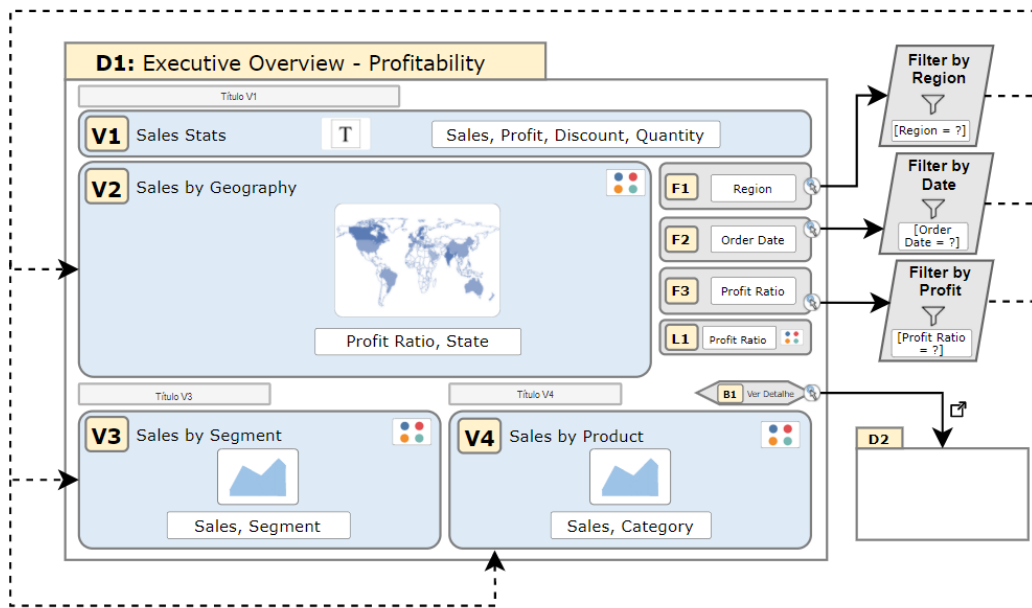


Figura B.32: Exemplo de diagrama de interações IVML completo.

Estes diagramas correspondem à modelação do *dashboard* de exemplo do Tableau [13] para o conjunto de dados "Superstore", com algumas alterações para efeitos de demonstração da linguagem de modelação. Na figura abaixo (B.33) é possível observar esse *dashboard* e confirmar que o seu desenho e modelação corresponde ao descrito pelos diagramas IVML apresentados anteriormente (de interação na figura B.32 e de *tooltips* na B.11b).

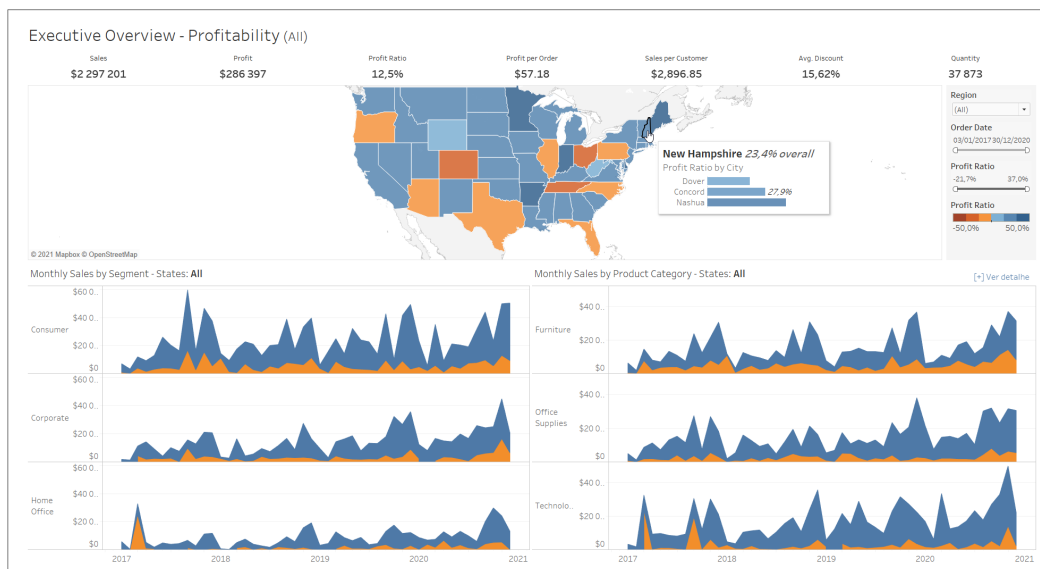


Figura B.33: *Dashboard* de exemplo do Tableau para o conjunto de dados "Superstore", com algumas alterações e uma *tooltip* ativa.

Embora o exemplo anterior demonstre a utilização de maior parte dos componentes da linguagem IVML, não foi possível apresentar todos os detalhes de modelação neste apêndice. Espera-se num futuro breve ter um relatório técnico detalhado da linguagem.

VARIANTES DO DASHBOARD GENERALISTA

O *dashboard* generalista, como explicado no capítulo 7, apresenta toda a informação fitossanitária referente ao inimigo secundário selecionado pelo utilizador. No capítulo 7 apresentou-se a variante deste *dashboard* correspondente à praga sésia, pelo que neste apêndice apresentam-se as variantes para as restantes 11 pragas secundárias estudadas pelo projeto FitoAgro.

C.1 Aranhaço Vermelho

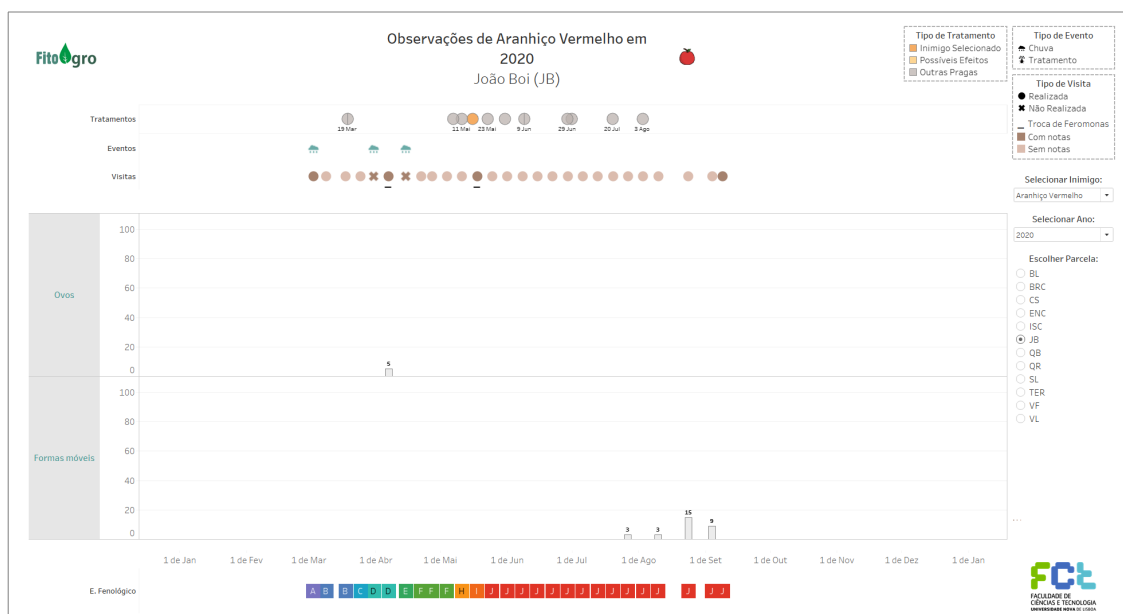


Figura C.1: *Dashboard* generalista com as observações fitossanitárias de aranhaço vermelho, bem como outros fatores relevantes que ocorreram no POB João Boi, em 2020.

C.2 Broca

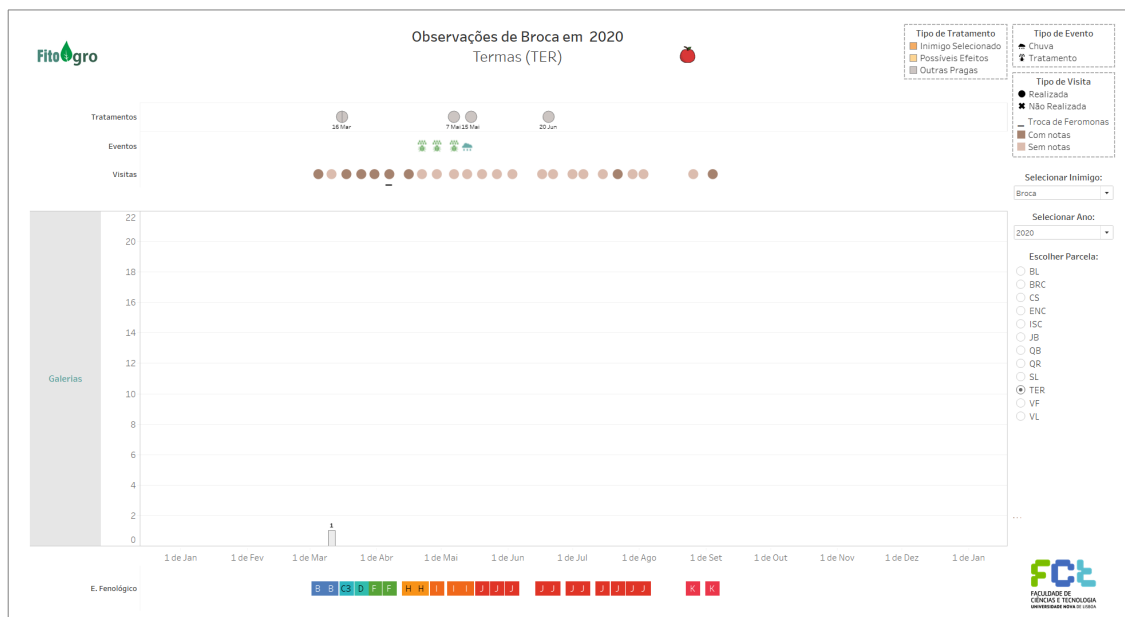


Figura C.2: *Dashboard* generalista com as observações fitossanitárias de broca, bem como outros fatores relevantes que ocorreram no POB Termas, ao longo de 2020.

C.3 Cochonilha São José

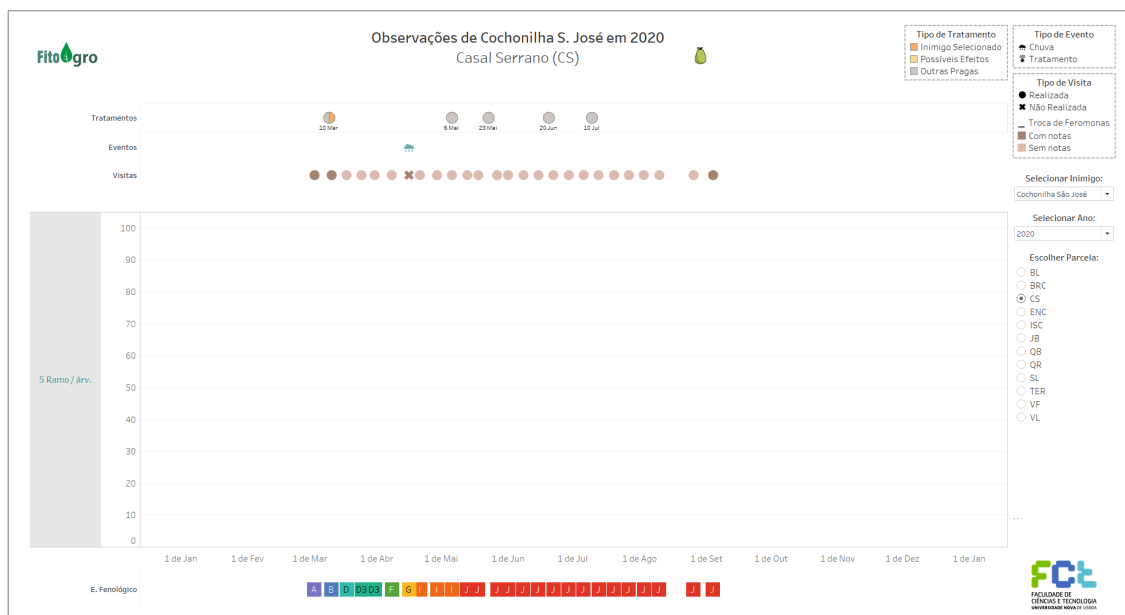


Figura C.3: *Dashboard* generalista com as observações fitossanitárias de cochonilha São José (sem registros), bem como outros fatores relevantes que ocorreram no POB Casal Serrano, ao longo de 2020.

C.4 Estenfiliose

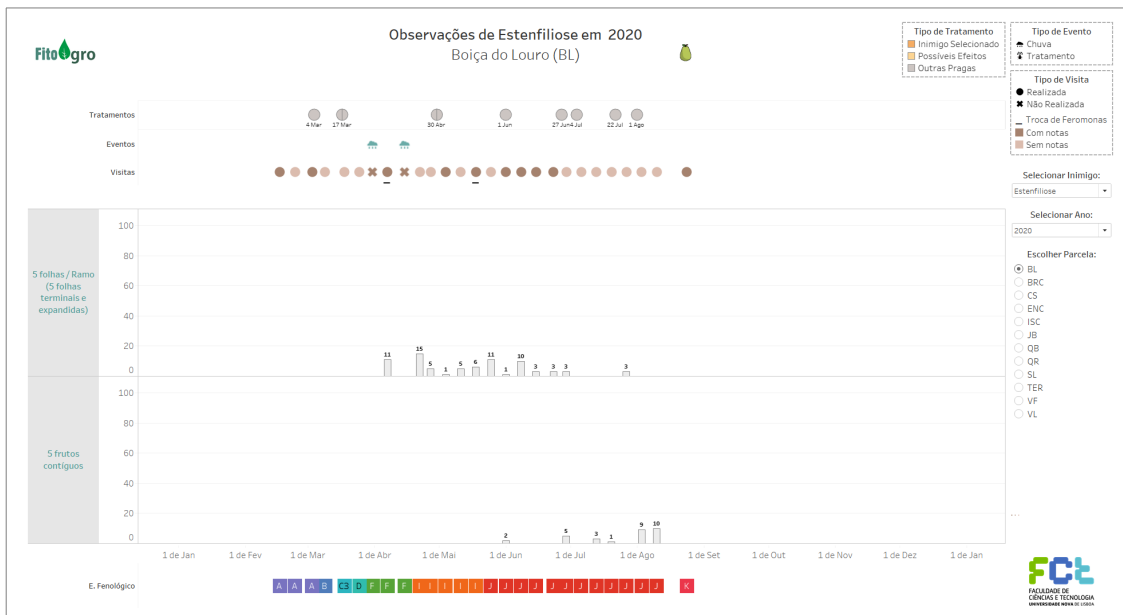


Figura C.4: *Dashboard* generalista com as observações fitossanitárias de estenfiliose, bem como outros fatores relevantes que ocorreram no **POB** Boiça do Louro, ao longo de 2020.

C.5 Hoplocampa

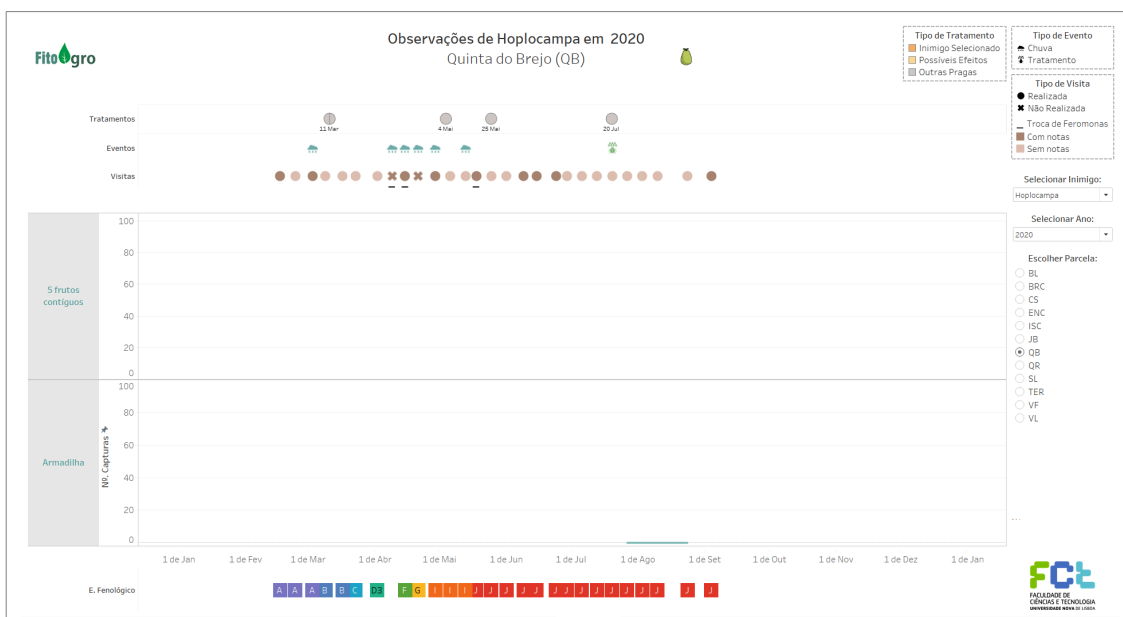


Figura C.5: *Dashboard* generalista com as observações fitossanitárias de hoplocampa (sem registos), bem como outros fatores relevantes que ocorreram no **POB** Quinta do Brejo, ao longo de 2020.

C.6 Lagarta Mineira

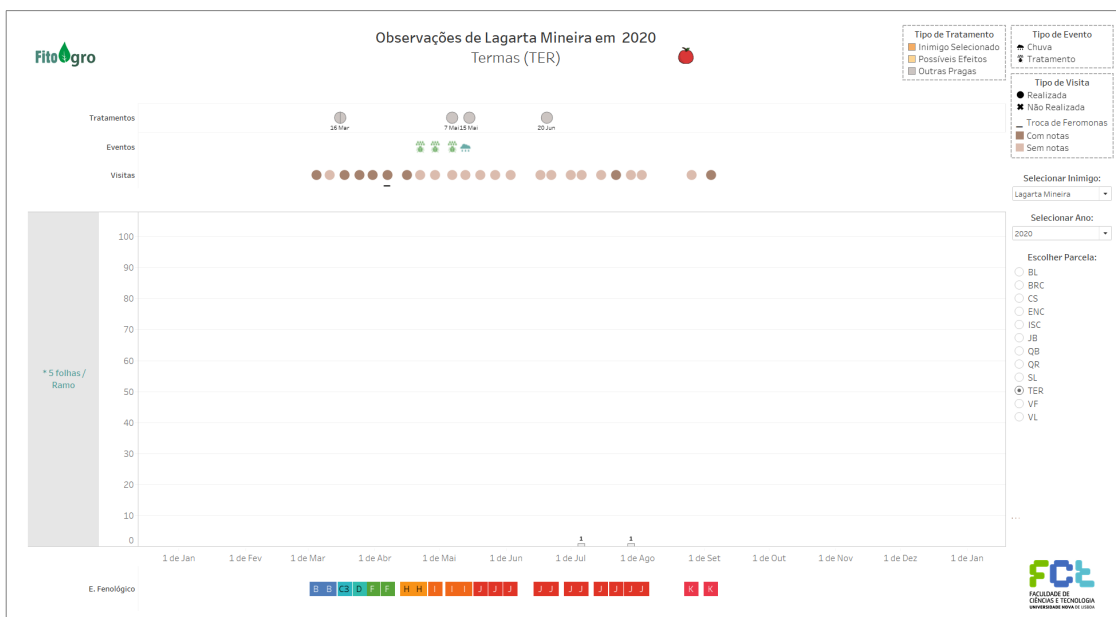


Figura C.6: Dashboard generalista com as observações fitossanitárias de lagarta mineira, bem como outros fatores relevantes que ocorreram no POB Termas, ao longo de 2020.

C.7 Mosca Mediterrânica

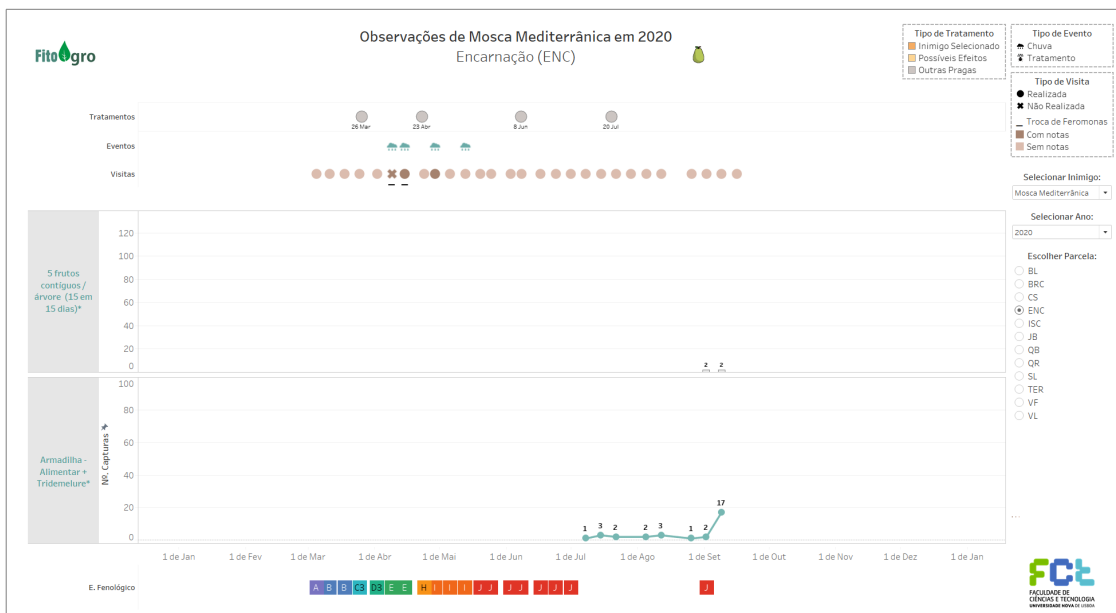


Figura C.7: Dashboard generalista com as observações fitossanitárias de mosca mediterrânica, bem como outros fatores relevantes que ocorreram no POB Encarnação, em 2020.

C.10 Psila

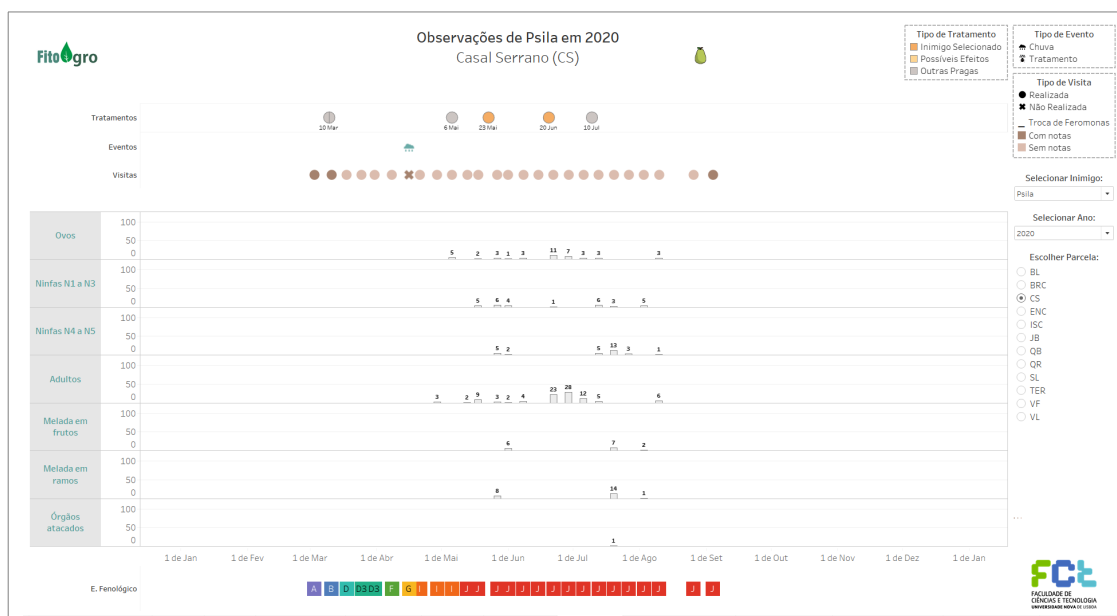


Figura C.10: *Dashboard* generalista com as observações fitossanitárias de psila, bem como outros fatores relevantes que ocorreram no **POB** Casal Serrano, ao longo de 2020.

C.11 Pulgão Lanígero

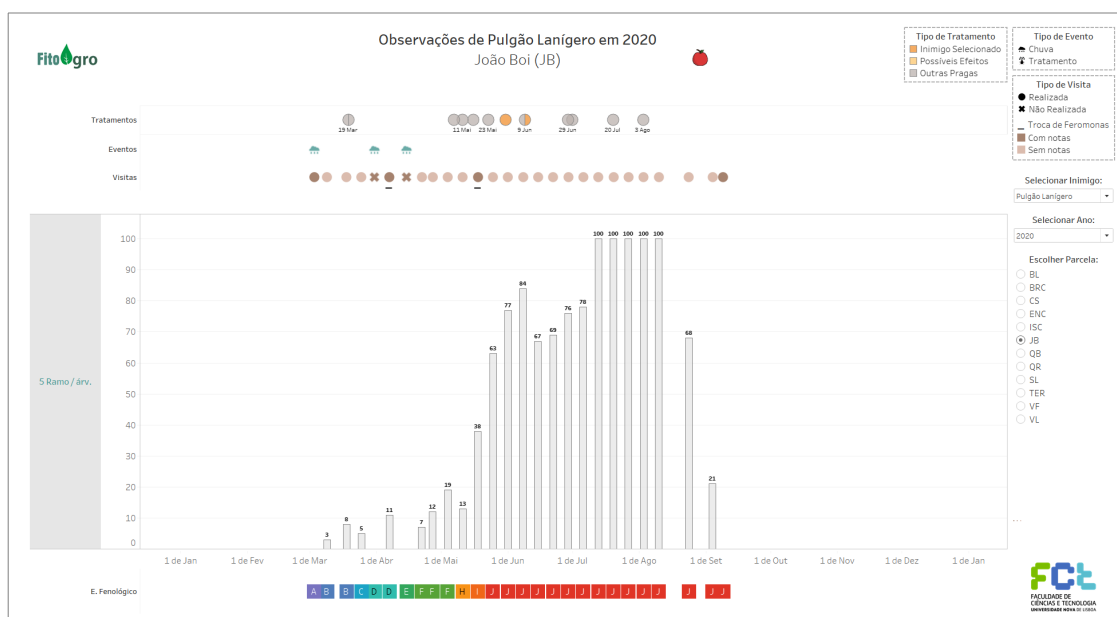


Figura C.11: *Dashboard* generalista com as observações fitossanitárias de pulgão Lanígero, bem como outros fatores relevantes que ocorreram no **POB** João Boi, ao longo de 2020.



QUESTIONÁRIO PARA AVALIAÇÃO DA COMPONENTE

• SECÇÃO 1 - MONITORIZAÇÃO E CONTROLO DOS DADOS METEOROLÓGICOS

A - Mapa de estado das estações meteorológicas (visualização 6.1) e respetiva *tooltip* (visualização 6.2)

1. Quantas estações estão *offline*?
2. Quantas estações se situam em POBs?
3. Qual o estado da maioria das estações meteorológicas?
4. Qual a percentagem de qualidade da estação Leiria (Aeródromo) no último ano?
5. Quando foi enviado o último registo da estação selecionada?
6. Em que dia, dos 30 dias anteriores à data do último registo, houve falhas em sensores?

MCQ 1. Através desta visualização e respetiva *tooltip*, sou capaz de detetar estações avariadas e falhas de sensores para que se possa sinalizar a sua reparação.

B - Qualidade Diária das Estações Meteorológicas (visualização 6.4)

7. Qual o fornecedor das estações apresentadas? Escolha múltipla.
8. Das estações apresentadas qual é a mais propícia a falhas, isto é, qual a que tem menor qualidade ao longo dos 30 dias? Escolha múltipla.
9. Qual a percentagem da qualidade diária da estação AARA Maiorga no dia 14 de janeiro de 2021? Resposta aberta.
10. Qual o sensor da estação AARA Maiorga com menor qualidade no dia 14 de janeiro de 2021? Escolha múltipla.

MCQ 2. Através desta visualização e respetivas *tooltips*, sou capaz de monitorizar a qualidade das estações e dos seus respetivos sensores ao longo do tempo, de forma a identificar estações propícias a falhas versus estações fiáveis.

C - Exemplos de dados meteorológicos originais dos sensores ao longo de um dia (visualização 6.6).

11. Qual a estação apresentada? Escolha múltipla.
12. Por volta de que hora se deu o pico da temperatura e depressão do ponto de orvalho? Resposta aberta.
13. Choveu depois do pôr do sol? Escolha múltipla.

MCQ 3. (i) Através desta visualização (exemplo sem falhas), sou capaz de observar, ao longo do tempo, os dados em bruto dos diferentes sensores meteorológicos, para diferentes estações.

MCQ 3. (ii) Através desta visualização (exemplo com falhas), sou capaz de identificar falhas e inconsistências nos dados em bruto dos diferentes sensores meteorológicos, para diferentes estações.

• **SECÇÃO 2 - ANALÍTICA DA METEOROLOGIA**

D - Dados meteorológicos normalizados dos sensores agregados para um intervalo de tempo (visualização 6.13) e respetiva *tooltip* (visualização 6.14)

14. Qual o dia no centro do intervalo de tempo? Resposta Aberta.
15. Qual o período/comprimento do intervalo de tempo? Escolha múltipla.
16. Os dados estão agregados de que forma? Escolha múltipla.
17. Nos dias 28 e 29 de novembro de 2020, a precipitação acumulada registada foi semelhante. No entanto, em média choveu mais por hora no dia 28 do que no dia 29. Verdadeiro/Falso.
18. Qual a temperatura do ar máxima no dia 30 de Novembro? A que horas ocorreu? Escolha múltipla.

E - Dados meteorológicos normalizados dos sensores horariamente para um dia (visualização 6.15).

19. A estação registou que choveu a que horas daquele dia? Resposta Aberta.
20. Qual a velocidade média do vento registada às 10 da manhã? E em que direção? Escolha múltipla.
21. Sabendo que os dados são horários (de hora a hora), a que horas se deu o pico da depressão do ponto de orvalho? Resposta Aberta.
22. O meio-dia solar ocorreu ao meio-dia (12:00). Verdadeiro/Falso.

AD 1. Através das duas visualizações anteriores, sou capaz de observar a evolução dos valores registados pelos diferentes sensores meteorológicos nas diferentes estações, para diferentes agregações e intervalos de tempo. Escala *Likert*.

F - Dados Meteorológicos nos locais dos POBs (visualização 6.16)

23. Qual a estação na proximidade deste POB? Resposta aberta.

AD 2. Através da visualização anterior, sou capaz de observar a evolução dos valores registados pelos diferentes sensores meteorológicos nos locais dos POBs, para diferentes agregações e intervalos de tempo.

• **SECÇÃO 3 - MONITORIZAÇÃO E CONTROLO DOS DADOS DE CAMPO**

G - Dados de campo recolhidos durante uma visita a um POB (visualização 7.3c)

24. Quantos EOIs (i.e. árvores) foram registados como atacados por cecidómia? Resposta aberta.
25. Quantos rebentos foram registados como atacados por cecidómia no EOI número 18? Resposta aberta.
26. No total quantas folhas foram registadas como atacadas por estenfiliose? Resposta aberta.
27. Qual o valor médio observado de folhas atacadas por estenfiliose em EOIs atacados? E nos EOIs em geral? Escolha múltipla.
28. Qual armadilha teve um decréscimo no número de capturas? Escolha múltipla.

MCQ 4. Através desta visualização, sou capaz de verificar a correção e consistência os dados recolhidos em cada visita de campo.

• **SECÇÃO 4 - FITOSSANIDADE**

H - Observações fitossanitárias de cecidómia (visualização 7.11)

29. ~~Qual a cultura do POB?~~

Nota: Esta pergunta foi removida dos resultados uma vez que o inimigo apresentado no exemplo (cecidómia) ataca apenas pereiras e a maioria das respostas dadas tiveram como base esse conhecimento agrónomo e não a utilização da componente.

30. Os tratamentos ocorreram nos dias das visitas? Escolha múltipla.
31. Nos dias em que não foi possível realizar as visitas, que tipo de evento ocorreu? Escolha múltipla.
32. Atingiu-se o nível económico de ataque em 2020 neste POB? Escolha múltipla.
33. Qual a maior percentagem de ataque observada? Resposta aberta.
34. Houve troca de feromonas no dia da visita onde se registou o estado fenológico F? Escolha múltipla.

AD 4. Através desta visualização, sou capaz de observar a evolução das visitas, tratamentos, observações de campo, estado fenológico e ocorrências meteorológicas registadas, promovendo a descoberta de correlações nos mesmos.

AD 6. Através desta visualização, sou capaz de observar a percentagem de ataque da praga e o seu respetivo nível económico de ataque (NEA).

I - Comparar observações fitossanitárias de cecidómia entre POBs (visualização 7.20) e entre anos (visualização 7.21).

35. Qual dos POBs teve mais rebentos atacados em 2020? Escolha múltipla.

36. Qual dos POBs atingiu o estado de floração (F) primeiro? Escolha múltipla.

37. Qual dos POBs sofreu mais tratamentos? Escolha múltipla.

38. Em que ano se atingiu uma percentagem de ataque de 20%? Escolha múltipla.

39. Em que ano houve mais rebentos atacados no mês de maio? Escolha múltipla.

40. Em que ano se atingiu o estado fenológico I primeiro? Escolha múltipla.

AD 5. Através destas duas visualizações, sou capaz de comparar a evolução de toda a informação fitossanitária em diferentes locais e/ou anos.

• **SECÇÃO 5 - VARIÁVEIS DERIVADAS DA METEOROLOGIA**

J - Temperatura média por hora, com indicação de horas frias (visualização 6.17) e horas de frio acumuladas (visualização 6.18).

41. A partir de que temperatura se contabiliza horas de frio? Resposta aberta.

42. Houve horas frias em Quinta do Rato no início de outubro de 2020? Escolha múltipla.

43. Em que mês houve mais horas de frio na Quinta do Rato? Escolha múltipla.

44. Qual o limite de horas de frio indicado acima para a cultura? Escolha múltipla.

45. Atingiu-se o limite de horas de frio indicado antes de 15 de fevereiro de 2021? Escolha múltipla.

AD 3. (i) Através das duas visualizações anteriores sou capaz de observar a evolução das horas de frio.

K - Graus-dia (visualização 6.19) e graus-dia acumulados (visualização 6.20) ao longo do ano para o Bichado (*Cydia pomonella*)

46. Qual é o máximo de graus-dia registados na Quinta do Rato em 2020? Escolha múltipla.

47. Em que mês se registou menos graus-dia? Escolha múltipla.

48. Quantos graus dia se acumulou até Novembro de 2020? Resposta aberta.

49. Em que data se acumulou graus-dia suficientes para o pico do voo da 1ª geração? Resposta aberta.

AD 3. (ii) Através das duas visualizações anteriores sou capaz de observar a evolução dos graus-dia.

• **SECÇÃO 6 - IMPACTO DA METEOROLOGIA NA FITOSSANIDADE**

L - Comparar observações fitossanitárias de cecidómia ao longo dos graus-dia acumulados, entre POBs (visualização 7.24) e entre anos (visualização 7.25).

50. Até ao dia 18 de março de 2019 (1ª visita) qual dos POBs acumulou mais graus-dia? Escolha múltipla.

51. Em qual dos anos se acumulou mais graus-dia até ao início de agosto? Escolha múltipla.

AD 7. Considero que estas duas visualizações promovem a descoberta de correlações entre a meteorologia e a informação fitossanitária registada, para diferentes locais e/ou anos.

M - Comparar o ajuste de modelos preditivos entre POBs (visualização 7.26).

AD 8. Através desta visualização, sou capaz de observar o ajuste dos modelos preditivos existentes às observações fitossanitárias registadas.

A N E X O



LEVANTAMENTO DE MOMENTOS CRÍTICOS

Determinação de Momentos Críticos para a Cecidómia

Este ensaio tem como objetivo determinar os momentos críticos ao longo dos últimos 3 anos para cada um dos pontos de observação biológica (POBs), onde foi executado o protocolo da cecidómia (Azambujeira dos Carros, Barreiras da Bica, Boiça do Louro, Casal Serrano, Quinta do Brejo e São Lourenço).

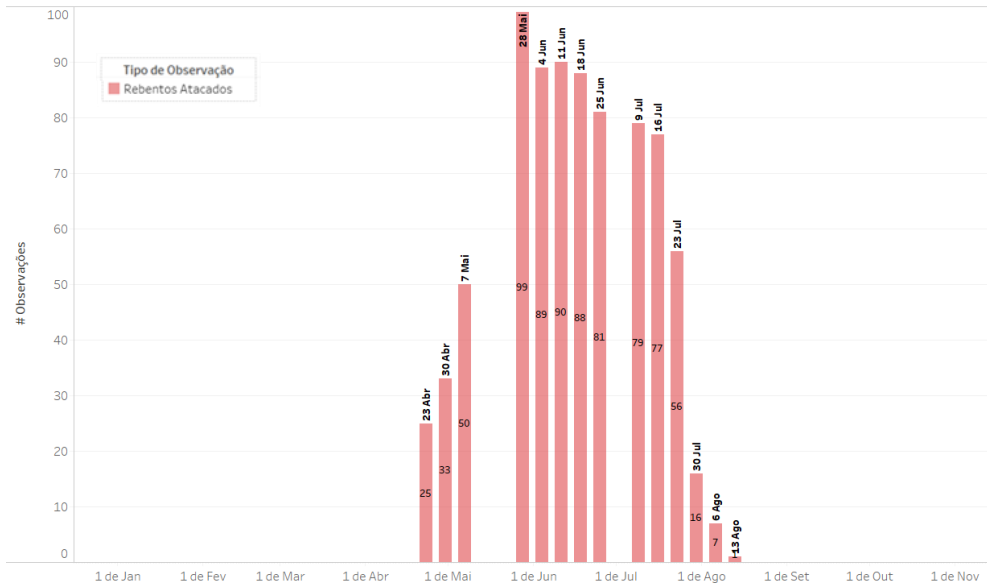
A cada um desses POBs estão associadas **2 páginas de registo**. A primeira página contém para cada ano uma visualização com as observações que foram registadas*. Deverá começar por observar os gráficos dessa página e anotar, com a melhor precisão possível, os momentos críticos que consegue identificar (e.g. ataques, tratamentos, entre outros). Os momentos críticos poderão corresponder a um instante/data fixa (e.g. início da 1ª geração) ou a intervalos de tempo (e.g. ataque da 1ª geração).

De seguida, relativamente ao que observou e anotou, deverá registar na tabela da página seguinte, da forma mais completa possível, os tipos de momentos críticos, a(s) sua(s) data(s) corresponde(s), o intervalo de imprecisão (e.g. 2 dias, 1 semana, etc.) e, opcionalmente, comentários adicionais. (Ver exemplo abaixo)

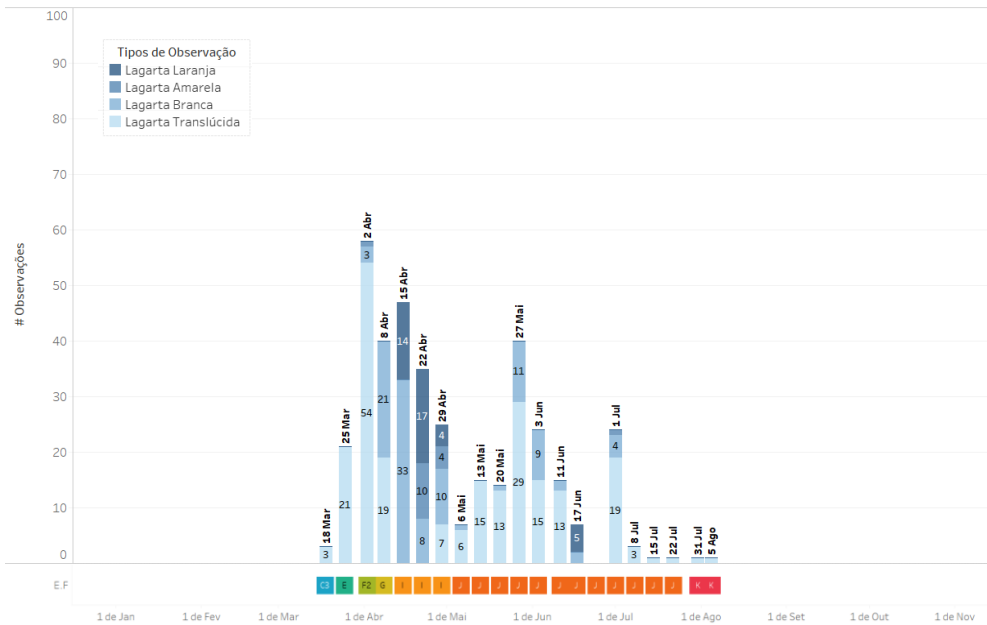
Ano	Tipo de Momento	Data(s)	Intervalo de Imprecisão	Comentários Adicionais
2019	Ataque da Geração Hibernante	27 Março – 10 Maio	3 dias	Possível sobreposição de gerações

Finalmente, certifique-se que regista o seu nome (primeiro e último) no canto superior direito de **todas** as páginas.

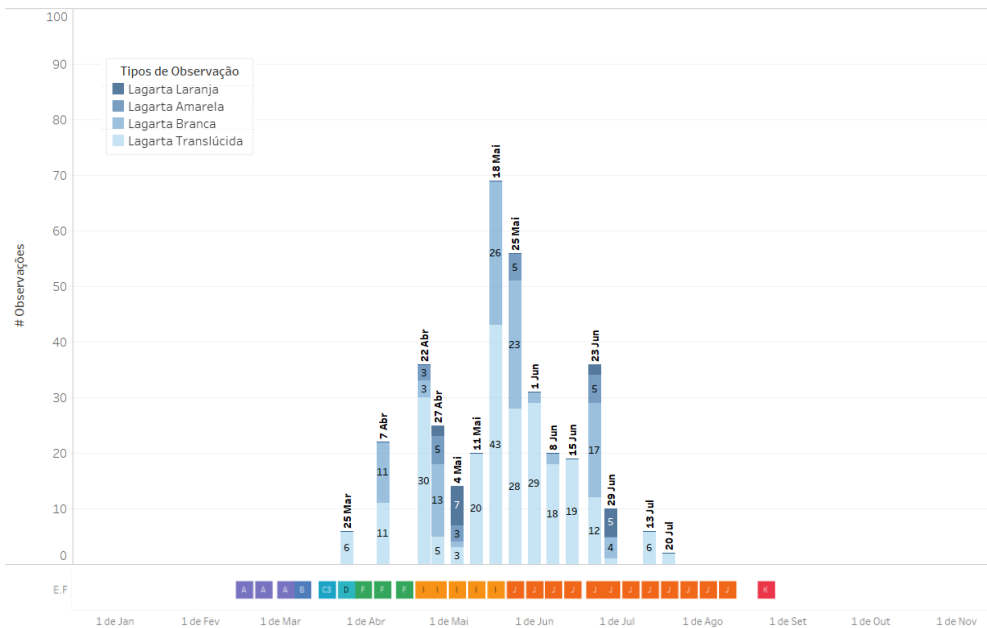
Observações de Cecidómia em 2018



Observações de Cecidómia em 2019



Observações de Cecidómia em 2020



Ano	Tipo de Momento	Data(s)	Intervalo de Imprecisão	Comentários Adicionais
2018	Início 1ª geração (Geração hibernante)	23/04/2018		
2018	Pico 1ª geração / Início da 2ª geração	28/05/2018		Pelo facto de não haver cor das larvas, não é fácil saber se se trata de início de nova geração
2018	Fim da 2ª geração	15/08/2018		Possivelmente haverá sobreposição de gerações e esta poderá ser a 3ª geração
2019	Início 1ª geração (Geração hibernante)	18/03/2019		
2019	Pico 1ª geração	02/04/2019		
2019	Fim da 1ª geração/Início da 2ª geração	06/05/2019		
2019	Pico da 2ª geração	27/05/2019		
2019	Início 3ª geração	17/06/2019	15 dias	
2019	Pico 3ª geração	01/07/2019		
2019	Fim da 3ª geração	05/08/2019		
2020	Início 1ª geração (Geração hibernante)	25/03/2020		
2020	Pico 1ª geração	22/04/2020		
2020	Início da 2ª geração	11/05/2020		
2020	Pico da 2ª geração	18/05/2020		
2020	Fim 2ª geração	20/07/2020		Possivelmente houve aqui tratamentos e esta será uma 3ª geração que teve início entre 29/06 e 13/07