



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



sid.inpe.br/mtc-m21d/2021/07.19.17.10-RPQ

RELATÓRIO DE ATIVIDADES DO PROJETO BRAZIL DATA CUBE - JANEIRO DE 2020 A MAIO DE 2021

Abner Ernani dos Anjos
Alana Kasahara Neves
Alber Hamersson Sánchez Ipia
Carlos Alberto Ferreira de Noronha
Claudinei de Camargo
Fabiana Zioti
Felipe Carvalho de Souza
Felipe Menino Carlos
Gabriel Sansigolo
Gilberto Ribeiro de Queiroz
Jeferson de Souza Arcanjo
Karine Reis Ferreira Gomes
Lorena Alves dos Santos
Luciana Mamede dos Santos
Raphael Willian da Costa
Rennan de Freitas Bezerra Marujo
Rolf Ezequiel de Oliveira Simões

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34T/454URCS>>

INPE
São José dos Campos
2021

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE - CEPPII (PORTARIA Nº 176/2018/SEI-INPE):

Presidente:

Dra. Marley Cavalcante de Lima Moscati - Coordenação-Geral de Ciências da Terra (CGCT)

Membros:

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação (CPG)
Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia, Tecnologia e Ciência Espaciais (CGCE)
Dr. Rafael Duarte Coelho dos Santos - Coordenação-Geral de Infraestrutura e Pesquisas Aplicadas (CGIP)
Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon
Clayton Martins Pereira - Divisão de Biblioteca (DIBIB)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)

EDITORAÇÃO ELETRÔNICA:

Ivone Martins - Divisão de Biblioteca (DIBIB)
Cauê Silva Fróes - Divisão de Biblioteca (DIBIB)



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



sid.inpe.br/mtc-m21d/2021/07.19.17.10-RPQ

RELATÓRIO DE ATIVIDADES DO PROJETO BRAZIL DATA CUBE - JANEIRO DE 2020 A MAIO DE 2021

Abner Ernani dos Anjos
Alana Kasahara Neves
Alber Hamersson Sánchez Ipia
Carlos Alberto Ferreira de Noronha
Claudinei de Camargo
Fabiana Zioti
Felipe Carvalho de Souza
Felipe Menino Carlos
Gabriel Sansigolo
Gilberto Ribeiro de Queiroz
Jeferson de Souza Arcanjo
Karine Reis Ferreira Gomes
Lorena Alves dos Santos
Luciana Mamede dos Santos
Raphael Willian da Costa
Rennan de Freitas Bezerra Marujo
Rolf Ezequiel de Oliveira Simões

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34T/454URCS>>

INPE
São José dos Campos
2021



Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

O Brazil Data Cube é um subprojeto do Projeto Monitoramento Ambiental dos Biomas Brasileiros, financiado com recursos do Fundo Amazônia, por meio da colaboração financeira BNDES e FUNCATE nº 17.2.0536.1"



Sumário

1	Introdução	6
2	Coleções, Cubos de Dados e Mosaicos	8
2.1	Coleções de Imagens	8
2.2	Cubos de Dados	15
2.3	Mosaicos	20
3	Plataforma Computacional	23
3.1	Aquisição, Processamento e Geração dos Cubos	24
3.1.1	Image Collection Builder	25
3.1.2	Data Cube Builder	27
3.2	Serviços	35
3.2.1	BDC-STAC	35
3.2.2	Web Time Series Service	36
3.2.3	Web Land Trajectory Service	39
3.3	Aplicações	42
3.3.1	BDC Explorer	42
3.3.2	SITS	44
3.3.3	Open Data Cube	49
3.3.4	Ambientes Interativos de Desenvolvimento	53
3.4	Ambiente de desenvolvimento e monitoramento computacional	58
3.4.1	Github	58

3.4.2	Drone CI	59
3.4.3	BDC Monitor	60
4	Classificações	63
4.1	Dados e software	65
4.1.1	Dados de Observação da Terra	65
4.1.2	Software	66
4.1.3	Coleta de amostras	68
4.1.4	Scripts das classificações	69
4.2	Metodologia	69
4.2.1	Amostragem	69
4.2.2	Classificação de Uso e Cobertura da Terra	70
4.3	Bioma Cerrado	72
4.3.1	Área classificada	72
4.3.2	Classes e amostras	72
4.3.3	Cubos de dados	73
4.3.4	Resultados	74
4.3.5	Acurácia	75
4.4	Bioma Amazônia - desmatamento	77
4.4.1	Área classificada	77
4.4.2	Classes e amostras	77
4.4.3	Cubos de dados	79
4.4.4	Resultados	79
4.4.5	Acurácia	81

4.5	Bioma Amazônia - uso e cobertura da Terra	81
4.5.1	Área classificada	81
4.5.2	Classes e amostras	81
4.5.3	Cubos de dados	83
4.5.4	Resultados	83
4.5.5	Acurácia	84
4.6	Bioma Pantanal	85
4.6.1	Área classificada	85
4.6.2	Classes e amostras	86
4.6.3	Cubos	87
4.6.4	Pós-classificação	88
4.6.5	Resultados	89
4.6.6	Acurácia	89
5	Divulgação e capacitação	92
5.1	Artigos publicados em revistas e conferências	92
5.2	Divulgação	93
5.3	Documentação	98
5.4	Eventos de capacitação	98
5.4.1	Curso SITS	98
5.4.2	Tech Talks	99
5.4.3	II Workshop do Brazil Data Cube	100
5.4.4	SBSR Interim	101

6 Recursos humanos e infraestrutura	103
6.1 Recursos humanos	103
6.2 Trabalho remoto	103
6.3 Aquisições	104
6.4 Infraestrutura	105
6.4.1 GlusterFS	106
6.4.2 Manutenção da infraestrutura computacional	106
7 Comentários Finais e Próximos Passos	107

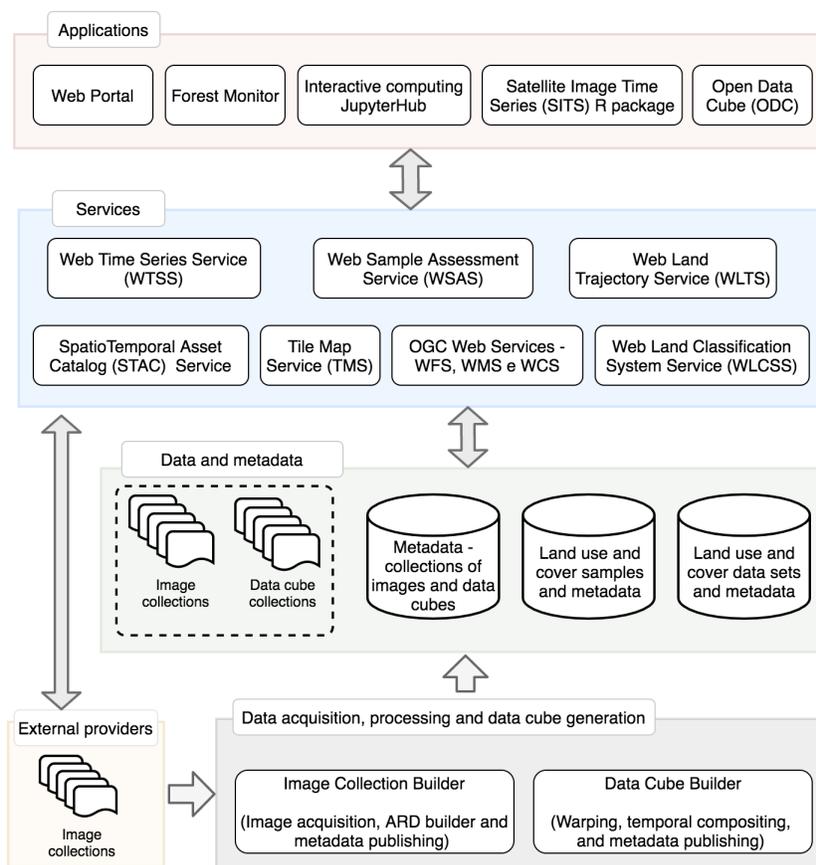
1 Introdução

Esse relatório descreve as atividades executadas e todos os produtos de dados, de software e de classificações gerados pela equipe do projeto Brazil Data Cube (BDC), no período de Janeiro de 2020 a Maio de 2021. O projeto BDC está produzindo dados a partir de grandes volumes de imagens de sensoriamento remoto para todo o território nacional e desenvolvendo uma plataforma computacional para processar e analisar esses dados usando inteligência artificial, aprendizado de máquina e análise de séries temporais de imagens. Este é um subprojeto do projeto Monitoramento Ambiental dos Biomas Brasileiros, financiado com recursos do Fundo Amazônia, por meio da colaboração financeira BNDES e FUNCATE nº 17.2.0536.1.

Nesse período, a equipe do projeto BDC continuou o desenvolvimento e evolução dos módulos da plataforma computacional, ilustrados na Figura 1, bem como a produção de coleções, cubos de dados e mosaicos a partir de imagens de sensoriamento remoto. Baseado em tecnologias de *big data* e serviços web para dados geográficos, ferramentas computacionais foram desenvolvidas para geração e consumo de cubos de dados de imagens de média resolução espacial.

Nesse período, as coleções e cubos de dados de imagens MODIS, CBERS-4/4A, Landsat-8 e Sentinel-2 foram ampliados para atingir a meta de cobrir todo o território nacional e de incluir imagens recentes. Além disso, foram criados mosaicos de imagens CBERS-4 e Landsat-8 para todo o Brasil. A família de sistemas e serviços da plataforma BDC continuou a crescer, havendo melhorias e novas implementações nos softwares *Image Collection Builder*, *Cube Builder*, integração dos produtos BDC com

Figura 1: Arquitetura de dados, softwares e serviços do Projeto Brazil Data Cube.



Source: [1].

o framework *Open Data Cube*, atualização do *BDC Explorer*, maturação dos serviços WTSS e WLTS, criação de ambientes interativos para desenvolvimento de aplicações e evolução nas aplicações, como por exemplo o SITS, entre muitos outros avanços.

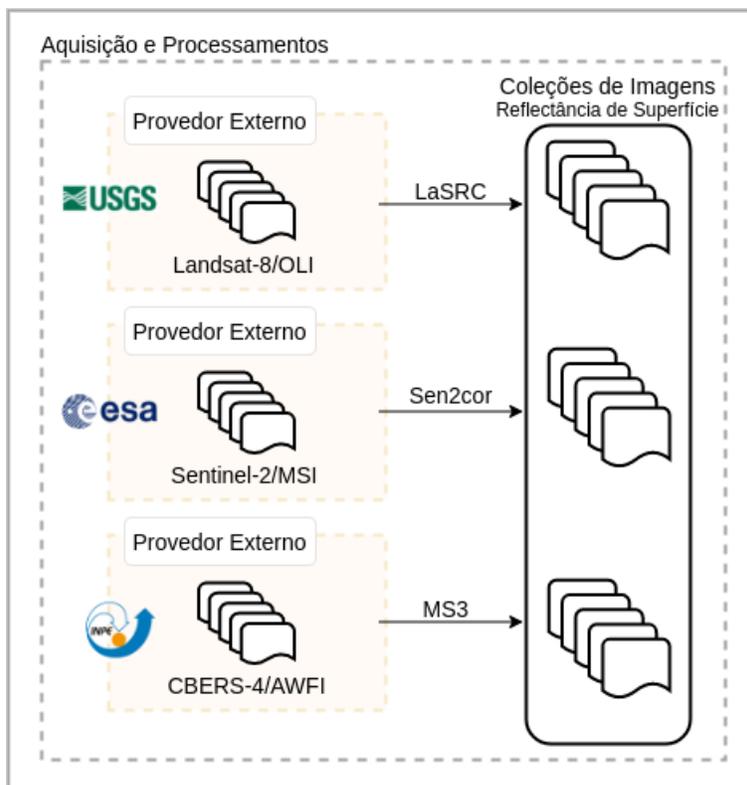
2 Coleções, Cubos de Dados e Mosaicos

Durante o ano de 2020 e primeiro semestre de 2021, continuou-se a aquisição de imagens dos satélites Landsat-8 (OLI) e Sentinel-2 (MSI) e foram adicionadas as imagens CBERS-4 (câmeras MUX e AWFII) ao catálogo do BDC. Os dados originais do satélite Landsat-8 são obtidos como produto da coleção 1 fornecido pelo USGS (*United States Geological Survey*) e as imagens Sentinel-2 são obtidas pelo portal da ESA (*European Space Agency*) e *mirrors* oficiais, como por exemplo CREODIAS e ONDADIAS. Todos os dados continuam a ser processados localmente nos servidores do projeto BDC.

2.1 Coleções de Imagens

As coleções de imagens distribuídas pelo projeto BDC são ARD (*Analysis Ready Data*), ou seja, imagens que atendem um conjunto de requisitos mínimos para serem utilizadas por aplicações [2]. Essas coleções ARD são geradas pela equipe do projeto BDC a partir de imagens obtidas de provedores de dados oficiais. Coleções de imagens em nível digital e reflectância de topo de atmosfera são utilizadas para gerar coleções ARD em nível de reflectância de superfície. Essas coleções ARD são usadas posteriormente para gerar os cubos de dados. Cada imagem é processada com os softwares distribuídos pelas agências provedoras de imagens, por exemplo para imagens Landsat-8 usa-se o processador LaSRC (*Landsat Surface Reflectance Code*). A Figura 2 ilustra o processamento de correção atmosférica feito em diversos produtos para gerar produtos reflectância de superfície.

Figura 2: Processamentos de correção atmosférica feitos em imagens disponibilizadas pelo BDC



As coleções de imagens ARD disponibilizadas pelo BDC (<https://brazildatacube.dpi.inpe.br/portal/explore>) cobrem todo o território brasileiro e são catalogadas por meio de um serviço STAC (*Spatio-Temporal Asset Catalog*). Esse serviço é utilizado para descrever os metadados, buscar e acessar as coleções e os cubos de dados do projeto BDC. No período referente a esse relatório, foi possível completar o acervo de imagens Landsat-8 Collection-1 de janeiro de 2017 até abril de 2021, bem como a inclusão dos dados CBERS-4 para todo o território nacional.

Os dados Sentinel-2 (Sentinel-2A e Sentinel-2B) apresentam um volume consideravelmente maior em comparação com as coleções de dados Landsat-8 e CBERS-4,

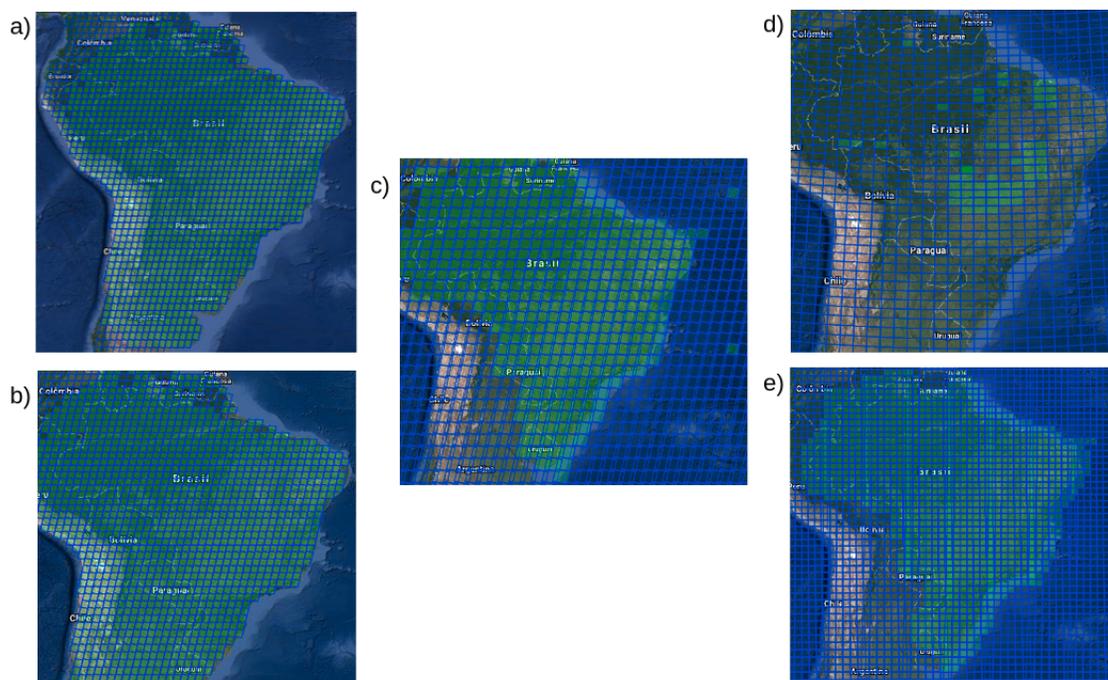
pois possui uma resolução espacial mais refinada (10 metros), uma revisita muito mais frequente (5 dias) e contém mais bandas espectrais. Além disso, a distribuição desses dados apresenta algumas restrições, por exemplo, uma conta pode fazer apenas dois downloads simultâneos. Também há a restrição de que apenas dados com até um ano da data atual são disponibilizados para download imediato. Dados com datas mais antigas do que um ano precisam ser solicitados e é necessário aguardar que os mesmos fiquem disponíveis para download. Esse processo pode levar a uma espera de até um dia para que o dado seja disponibilizado. Além disso, a quantidade de solicitações para disponibilização do dado também é limitada. Deste modo, no período do relatório foi possível obter imagens Sentinel-2 para todo o território brasileiro referentes ao ano 2020 e quase todas as imagens para o ano de 2019, restando completar apenas os meses de fevereiro e março de 2019. Além disso, as imagens do ano corrente de 2021 continuam sendo obtidas. Já para algumas áreas específicas, utilizadas para estudos pilotos, tem-se imagens Sentinel-2 já obtidas e processadas desde janeiro de 2017. A Tabela 1 apresenta os detalhes das coleções disponibilizadas pelo BDC, contendo assim os satélites e sensores dessas coleções, o nível de processamento, a resolução espacial de seus dados e o período de imagens catalogadas. A Figure 3 ilustra a completude desses para o território nacional.

Analisando as coleções de imagens geradas, conforme Tabela 2, podemos verificar que para todos os biomas brasileiros, a aquisição e processamento dos dados estão bastante adiantados, com exceção dos dados Sentinel-2, devido ao desafio do volume desses dados.

Tabela 1: Coleções de Imagens Catalogadas no Portal BDC.

Satélite/ Sensores	Nível de processamento	Resolução Espacial (m)	Período	Volume (TB)
CBERS-4/AWFI	SR (Level 4)	64	Jan 2015 - Jun 2021	30
CBERS-4/MUX	SR (Level 4)	20	Jan 2016 - Jun 2021	12
CBERS-4/MUX	SR (Level 5)	20	Jan 2017 - Dez 2019	4
Landsat-8/OLI-TIRS	DN (L1)	30	Jan 2017 - Abr 2021	50
Landsat-8/OLI	SR (L2)	30	Jan 2017 - Abr 2021	50
Sentinel-2/MSI	ToA (Level-1C)	10/20/60	Abr 2019 - Jan 2021	153
Sentinel-2/MSI	SR (Level-2A)	10/20/60	Abr 2019 - Jan 2021	203

Figura 3: Cobertura espacial das coleções de imagens ARD distribuídas pelo BDC: a) CBERS-4/MUX SR (Level 4) para o período Jan 2016 - Mai 2021; b) CBERS-4/WFI SR (Level 4) para o período Jan 2015 - Mai 2021; c) Landsat-8/OLI-TIRS para o período Jan 2017 - Abr 2021; d) Sentinel-2/MSI para o período Jan 2017 - Mar 2019 e e) Sentinel-2/MSI para o período Abr 2019 - Jan 2021.



Bioma	Coleção	Período
Amazônia	CBERS-4/AWFI (L4)	01/01/2015 - 31/05/2021
	CBERS-4/MUX (L4)	01/01/2016 - 31/05/2021
	Landsat-8/OLI-TIRS (L1)	01/08/2016 - 31/05/2021
	Landsat-8/OLI (L2)	01/08/2016 - 31/05/2021
	Sentinel-2/MSI (L1)	01/04/2019 - 31/12/2020
	Sentinel-2/MSI (L2)	Em andamento
Caatinga	CBERS-4/AWFI (L4)	01/01/2015 - 31/05/2021
	CBERS-4/MUX (L4)	01/01/2016 - 31/05/2021
	Landsat-8/OLI-TIRS (L1)	01/08/2016 - 31/05/2021
	Landsat-8/OLI (L2)	01/08/2016 - 31/05/2021
	Sentinel-2/MSI (L1)	01/04/2019 - 31/12/2020
	Sentinel-2/MSI (L2)	Em andamento
Cerrado	CBERS-4/AWFI (L4)	01/01/2015 - 31/05/2021
	CBERS-4/MUX (L4)	01/01/2016 - 31/05/2021
	Landsat-8/OLI-TIRS (L1)	01/08/2016 - 31/05/2021
	Landsat-8/OLI (L2)	01/08/2016 - 31/05/2021
	Sentinel-2/MSI (L1)	01/04/2019 - 31/12/2020
	Sentinel-2/MSI (L2)	Em andamento

Bioma	Coleção	Período
Mata Atlântica	CBERS-4/AWFI (L4)	01/01/2015 - 31/05/2021
	CBERS-4/MUX (L4)	01/01/2016 - 31/05/2021
	Landsat-8/OLI-TIRS (L1)	01/08/2016 - 31/05/2021
	Landsat-8/OLI (L2)	01/08/2016 - 31/05/2021
	Sentinel-2/MSI (L1)	01/04/2019 - 31/12/2020
	Sentinel-2/MSI (L2)	Em andamento
Pampa	CBERS-4/AWFI (L4)	01/01/2015 - 31/05/2021
	CBERS-4/MUX (L4)	01/01/2016 - 31/05/2021
	Landsat-8/OLI-TIRS (L1)	01/08/2016 - 31/05/2021
	Landsat-8/OLI (L2)	01/08/2016 - 31/05/2021
	Sentinel-2/MSI (L1)	01/04/2019 - 31/12/2020
	Sentinel-2/MSI (L2)	Em andamento
Pantanal	CBERS-4/AWFI (L4)	01/01/2015 - 31/05/2021
	CBERS-4/MUX (L4)	01/01/2016 - 31/05/2021
	Landsat-8/OLI-TIRS (L1)	01/08/2016 - 31/05/2021
	Landsat-8/OLI (L2)	01/08/2016 - 31/05/2021
	Sentinel-2/MSI (L1)	01/04/2019 - 31/12/2020
	Sentinel-2/MSI (L2)	Em andamento

Tabela 2: Status de completude das Coleções de Imagens BDC por Biomas.

2.2 Cubos de Dados

A cobertura espacial e temporal das coleções de imagens e cubos de dados foram ampliadas durante o ano de 2020 e primeiro semestre de 2021. Esses dados encontram-se disponíveis de forma aberta e de livre acesso através do portal [Brazil Data Cube Explorer](#). Para facilitar o desenvolvimento de aplicações e acesso através de interface máquina-máquina, todos os produtos de dados de imagem são disponibilizados por meio do serviço web STAC. Utilizando as coleções de imagens ARD processadas, foram gerados cubos de dados utilizando imagens em nível de reflectância de superfície para os satélites/sensores CBERS-4/WFI, CBERS-4/MUX, Landsat-8/OLI e Sentinel-2/MSI. Também foram incluídos no portal BDC os produtos MOD13Q1 e MYD13Q1, provenientes do sensor MODIS a bordo dos satélites Terra e Aqua, respectivamente.

Os cubos de dados podem ser categorizados em dois tipos: identidade e compostos temporais. Um cubo identidade representa as aquisições de imagens nas datas originais do sensor e não são regulares no tempo. Cubos compostos no tempo são regulares no tempo, ou seja, para cada período de tempo definido, por exemplo 16 dias ou 1 mês, existe uma imagem que foi produzida ou composta considerando todas as observações desse período. O processo de geração dos cubos de dados será detalhado no Capítulo 3.1.

Até a data corrente deste relatório, foram gerados diversos cubos identidade e cubos compostos no tempo, como pode ser observado na Tabela 3. O período e área espacial dos cubos de dados é diretamente dependente da disponibilidade dos dados

de entrada, no caso os produtos reflectância de superfície (coleções ARD de imagens descritas na seção anterior). Deste modo, os cubos de dados que utilizam imagens CBERS-4/WFI, CBERS-4/MUX ou Landsat-8/OLI foram gerados para todo o território brasileiro, de acordo com o período apresentado na Tabela 4. Os produtos MOD13Q1 e MYD13Q1 também foram adquiridos, preparados e disponibilizados.

Tabela 3: Cubos de dados gerados pelo BDC.

Cubo	Satélite	Sensor	Resolução Espacial (m)	Composição Temporal
CB4_64	CBERS-4	WFI	64	Identidade
CB4_64_16D_STK	CBERS-4	WFI	64	Stack de 16 dias
CB4_20_1M_STK	CBERS-4	MUX	20	Stack de 1 mês
LC8_30	Landsat-8	OLI	30	Identidade
LC8_30_16D_STK	Landsat-8	OLI	30	Stack de 16 dias
MOD13Q1	Terra	MODIS	250	Melhor pixel 16 dias
MYD13Q1	Aqua	MODIS	250	Melhor pixel 16 dias
S2_10	Sentinel 2A/2B	MSI	10	Identidade
S2_10_16D_STK	Sentinel 2A/2B	MSI	10	Stack de 16 dias

Os cubos de dados Sentinel-2 foram gerados para áreas piloto e estão sendo incrementados conforme vamos obtendo e processando as imagens de entrada. Devido as dificuldades para aquisição das imagens Sentinel-2, estamos usando a Amazon Web Services (AWS) para gerar os cubos de dados Sentinel-2 anteriores a 2019. A AWS, através de seu programa Earth on AWS (<https://aws.amazon.com/earth/>), mantém um repositório aberto e sem custos de imagens de satélites de observação o da Terra, inclusive imagens Sentinel-2. Estamos fazendo isso usando créditos de pesquisa fornecidos pela AWS ao projeto BDC. Assim, é feita uma verificação da completude das imagens Sentinel-2 no *bucket* S3 na AWS, de modo que os poucos dados faltantes são processados localmente nos servidores do BDC e enviados para

a AWS para serem utilizados em conjunto com os dados desse *bucket* para gerar os cubos de dados Sentinel-2. Até o presente momento para os biomas Amazônia e Cerrado tem-se os cubos Sentinel-2 de janeiro de 2017 até dezembro de 2020 e para os demais biomas já é possível processar as imagens para gerar os cubos referentes ao período de janeiro de 2019 até dezembro de 2020, com exceção dos meses fevereiro, março e junho de 2019 que estão sendo adquiridos.

Analisando as coleções de cubos de dados geradas, como pode ser observado na Tabela 4, nota-se que o padrão da tabela sumário de coleções de imagens se repete na tabela sumário de produtos de cubos de dados uma vez que para geração destes produtos é necessário a aquisição e processamento das coleções de imagens.

Bioma	Coleção	Período
	CB4.64	01/01/2015 - 31/05/2021
	CB4.64_16D_STK	01/01/2015 - 31/05/2021
	CB4.20_1M_STK	01/01/2016 - 31/05/2021
	LC8_30	01/08/2016 - 31/05/2021
Amazônia	LC8_30_16D_STK	01/08/2016 - 31/05/2021
	MOD13Q1	18/02/2000 - 31/03/2021
	MYD13Q1	04/07/2002 - 31/03/2021
	S2_10	Em andamento
	S2_10_16D_STK	Em andamento

Bioma	Coleção	Período
Caatinga	CB4.64	01/01/2015 - 31/05/2021
	CB4.64_16D_STK	01/01/2015 - 31/05/2021
	CB4.20_1M_STK	01/01/2016 - 31/05/2021
	LC8_30	01/08/2016 - 31/05/2021
	LC8_30_16D_STK	01/08/2016 - 31/05/2021
	MOD13Q1	18/02/2000 - 31/03/2021
	MYD13Q1	04/07/2002 - 31/03/2021
	S2_10	Em andamento
	S2_10_16D_STK	Em andamento
Cerrado	CB4.64	01/01/2015 - 31/05/2021
	CB4.64_16D_STK	01/01/2015 - 31/05/2021
	CB4.20_1M_STK	01/01/2016 - 31/05/2021
	LC8_30	01/08/2016 - 31/05/2021
	LC8_30_16D_STK	01/08/2016 - 31/05/2021
	MOD13Q1	18/02/2000 - 31/03/2021
	MYD13Q1	04/07/2002 - 31/03/2021
	S2_10	Em andamento
	S2_10_16D_STK	Em andamento

Bioma	Coleção	Período
	CB4.64	01/01/2015 - 31/05/2021
	CB4.64_16D_STK	01/01/2015 - 31/05/2021
	CB4.20_1M_STK	01/01/2016 - 31/05/2021
	LC8_30	01/08/2016 - 31/05/2021
Mata Atlântica	LC8_30_16D_STK	01/08/2016 - 31/05/2021
	MOD13Q1	18/02/2000 - 31/03/2021
	MYD13Q1	04/07/2002 - 31/03/2021
	S2_10	Em andamento
	S2_10_16D_STK	Em andamento
	CB4.64	01/01/2015 - 31/05/2021
	CB4.64_16D_STK	01/01/2015 - 31/05/2021
	CB4.20_1M_STK	01/01/2016 - 31/05/2021
	LC8_30	01/08/2016 - 31/05/2021
Pampa	LC8_30_16D_STK	01/08/2016 - 31/05/2021
	MOD13Q1	18/02/2000 - 31/03/2021
	MYD13Q1	04/07/2002 - 31/03/2021
	S2_10	Em andamento
	S2_10_16D_STK	Em andamento

Bioma	Coleção	Período
	CB4.64	01/01/2015 - 31/05/2021
	CB4.64_16D_STK	01/01/2015 - 31/05/2021
	CB4.20_1M_STK	01/01/2016 - 31/05/2021
	LC8_30	01/08/2016 - 31/05/2021
Pantanal	LC8_30_16D_STK	01/08/2016 - 31/05/2021
	MOD13Q1	18/02/2000 - 31/03/2021
	MYD13Q1	04/07/2002 - 31/03/2021
	S2_10	Em andamento
	S2_10_16D_STK	Em andamento

Tabela 4: Status de completude das Coleções de Cubos de Dados BDC por Biomas.

2.3 Mosaicos

No decorrer do projeto surgiram demandas institucionais e também externas para geração de mosaicos espaciais de imagens. A geração desses mosaicos apresenta três desafios principais:

1. Seleção das melhores imagens livres de nuvens em um grande acervo de dados.
2. Se a extensão territorial do mosaico for muito grande, temos a questão do volume de dados a ser manipulado e o tamanho do arquivo final resultante que pode ser da ordem de dezenas de Gigabytes.

3. Apresentação uniforme de cores e sem artefatos do produto final por conta da variabilidade no histograma das imagens de entrada.

Boa parte dos softwares desenvolvidos no BDC auxiliam na geração dos mosaicos, principalmente nas etapas de seleção de cenas e composição dos tiles livres de nuvens. Outras tarefas como a junção dos tiles e recorte do geotiff de acordo com um shapefile, por exemplo com o contorno do mapa do Brasil, ainda são feitos manualmente. Cabe salientar que ainda cabem melhorias no processo de geração de mosaicos no escopo do projeto BDC, que serão discutidas no futuro.

A seguir são listados os mosaicos gerados e já disponíveis através do BDC Explorer bem como algumas características de cada um deles e links para acesso, visualização, acesso a metadados e download dos arquivos:

- **Mosaico CBERS-4 Brazil:** Mosaico composto por imagens CBERS4/AWFI com 64m de resolução espacial dos meses de abril, maio e junho de 2020 e composição colorida nas bandas RGB vermelho (B15), infra-vermelho próximo (B16) e azul (B13). Link: https://brazildatacube.dpi.inpe.br/portal/explore?collection=CB4-MOSAIC-BRAZIL_64_3M_STK-1
- **Mosaico CBERS-4A Paraíba:** Mosaico composto por imagens CBERS-4A/WFI com 55m de resolução espacial dos meses de julho, agosto e setembro de 2020 e composição colorida nas bandas RGB Infra-vermelho próximo (B16), vermelho (B15) e verde (B14). Link: https://brazildatacube.dpi.inpe.br/portal/explore?collection=CB4A-MOSAIC-PARAIBA_55_3M_STK-1
- **Mosaico Landsat-8 Brazil:** Mosaico composto por imagens Landsat/OLI

com 30m de resolução espacial em dois períodos (1) julho a dezembro de 2017 e (2) janeiro a junho de 2018 e composição colorida nas bandas RGB SWIR1 (B6), infra-vermelho próximo (B5) e vermelho (B4). Link: https://brazildatacube.dpi.inpe.br/portal/explore?collection=LC8-MOSAIC-BRAZIL_30_6M_MEDSTK-1

- **Mosaico Sentinel-2 Paraíba:** Mosaico composto por imagens Sentinel-2/MSI com 10m de resolução espacial dos meses de novembro de dezembro de 2019 e janeiro de 2020 e composição colorida nas bandas RGB vermelho (B4), infra-vermelho próximo (B8) e verde (B3). Link: https://brazildatacube.dpi.inpe.br/portal/explore?collection=S2-MOSAIC-PARAIBA_10_3M_STK-1

A Figura 4 mostra o mosaico Landsat-8 gerado pelo BDC, considerando as imagens desde satélite dos últimos seis meses do ano de 2017.

Figura 4: Mosaico BDC Landsat-8/OLI de 2017



3 Plataforma Computacional

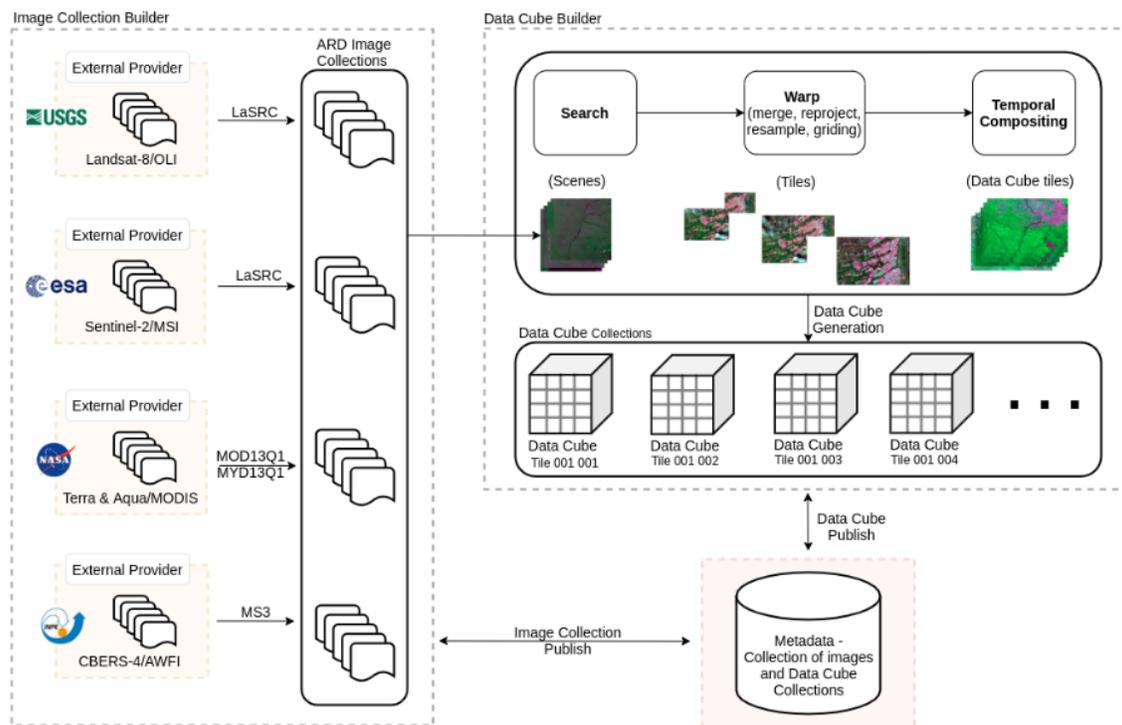
A plataforma computacional que está sendo desenvolvida no projeto BDC é composta por serviços web, aplicativos e pacotes de software, interfaces e ambientes de computação iterativa. Todos esses softwares são gratuitos e possuem código aberto. Como apresentado na Figura 1, eles podem ser divididos entre softwares para aquisição de imagens, geração e processamentos dos cubos de dados, serviços e aplicações. Os conjuntos de dados e metadados do projeto são acessados e processados por meio de serviços web. As aplicações referem-se a produtos de software para usuários finais, dentre eles sistemas com interfaces gráficas e APIs (interfaces de programação de aplicativos) para linguagens de programação de alto nível, como R e Python. Além disso, os serviços web fornecem a interface entre os conjuntos de

dados do projeto e os aplicativos.

3.1 Aquisição, Processamento e Geração dos Cubos

Um cubo de dados de observação da Terra pode ser definido como um conjunto de séries temporais associadas a pixels alinhados espacialmente, prontos para análise. Cubos de dados são criados principalmente para suportar análise de séries temporais de imagens [3]. Seguindo a arquitetura de softwares e serviços do BDC, dois softwares foram desenvolvidos para aquisição e processamento das imagens e construção dos cubos de dados, o *Image Collection Builder* e o *Data Cube Builder*, respectivamente, ilustrados na Figura 5.

Figura 5: Processo de coleta de dados, geração de ARD e de cubos.



Source: [1]

3.1.1 Image Collection Builder

O Image Collection Builder é um software desenvolvido com a finalidade de obter e processar as imagens de satélite por meio de APIs e ferramentas existentes para processamento desses dados. No projeto BDC, considerando o período de 2016 à 2021 e todo o território brasileiro, são adquiridas CBERS-4/MUX-WFI, Landsat-8/OLI-TIRS e Sentinel-2/MSI, sendo que as imagens CBERS-4/MUX-WFI são obtidas já processadas em nível de reflectância de superfície, enquanto que as imagens Landsat-8/OLI-TIRS e Sentinel-2/MSI precisam ser processadas. Deste modo, são utilizados

os softwares oficiais utilizados pelos mesmos provedores dessas imagens para realizar correção atmosférica nesses produtos, gerando assim para também o produto reflectância de superfície. Assim, as imagens Landsat-8/OLI-TIRS, que são obtidas em nível de número digital, são processadas para reflectância de superfície utilizando o software *Landsat Surface Reflectance Code* (LaSRC) [4], enquanto que as imagens Sentinel-2/MSI que são obtidas em nível reflectância de topo de atmosfera (ToA) são processadas utilizando o software Sen2cor [5].

Além do processamento para reflectância de superfície, também estão sendo agregados aos produtos Landsat-8 e Sentinel-2, uma máscara de nuvens complementar. Nesses casos, tem-se gerado o produto Fmask (versão 4.2) [6], que a partir de uma comparação com imagens da Amazônia brasileira e com os produtos de máscara de nuvens MAJA, SCL (Sen2cor) e s2cloudless apresentou melhor acurácia global [7].

Vale ressaltar que a busca e download desses dados é feita preferencialmente utilizando os provedores oficiais do dado. No caso do Landsat-8 por exemplo, essa busca é feita diretamente no EarthExplorer da USGS, enquanto que para os dados Sentinel-2 diretamente no SciHub da ESA. No período de janeiro de 2020 à maio de 2021, o Image Collection Builder foi incrementado para suportar a coleta de dados a partir de diferentes provedores de dados que contenham uma cópia do dado. Caso o coletor encontre alguma dificuldade em obter o dado em uma fonte primária, como por exemplo, instabilidade na rede ou indisponibilidade momentânea, os provedores secundários podem ser utilizados para realizar o download. No caso dos dados Sentinel-2, há suporte para os provedores SciHub, CreoDIAS, ONDADIAS e aos dados presentes na AWS, enquanto que para os dados Landsat-8 há suporte para os

provedores EarthExplorer (Collection-1 e Collection-2) e Google (Collection-1).

Após fazer o download ou processar as imagens, elas são publicadas em um banco de dados de sua respectiva coleção, e podem ser consultadas por meio do serviço STAC (3.2). As coleções adquiridas e processadas são distribuídas gratuitamente através do BDC Explorer.

Além das coleções de imagens mencionadas, o Image Collection Builder foi incrementado para ser capaz de obter os produtos MOD13Q1 e MYD13Q1, dos satélites sensores Terra/MODIS e Aqua/MODIS, os quais adéquam-se as definições de cubo de dados. No período referente a este relatório, também foi implementada a opção de publicação de dados não associados a uma grade definida, como foi o caso dos dados CBERS-4/WFI. Ainda há previsão de continuar incrementando o Image Collection Builder, de modo que as coletas possam ser feitas por meio de interface gráfica e também pretende-se ampliar o conjunto de dados para também incluir dados auxiliares, como por exemplo LADS que são utilizados para correção atmosférica dos dados Landsat-8, além da inclusão de suporte para aquisição de imagens Landsat-5 e Landsat-7.

3.1.2 Data Cube Builder

O Data Cube Builder é um software desenvolvido com a finalidade de utilizar as imagens coletadas pelo Image Collection Builder e organiza-las de forma a compor cubos de dados. No projeto BDC, considerando o período entre 2016 à 2021 e todo o território brasileiro, estão sendo gerados cubos de dados que utilizam, cada um, uma das seguintes coleções de imagens de reflectância de superfície CBERS-4/MUX,

CBERS-4/WFI, Landsat-8/OLI e Sentinel-2/MSI.

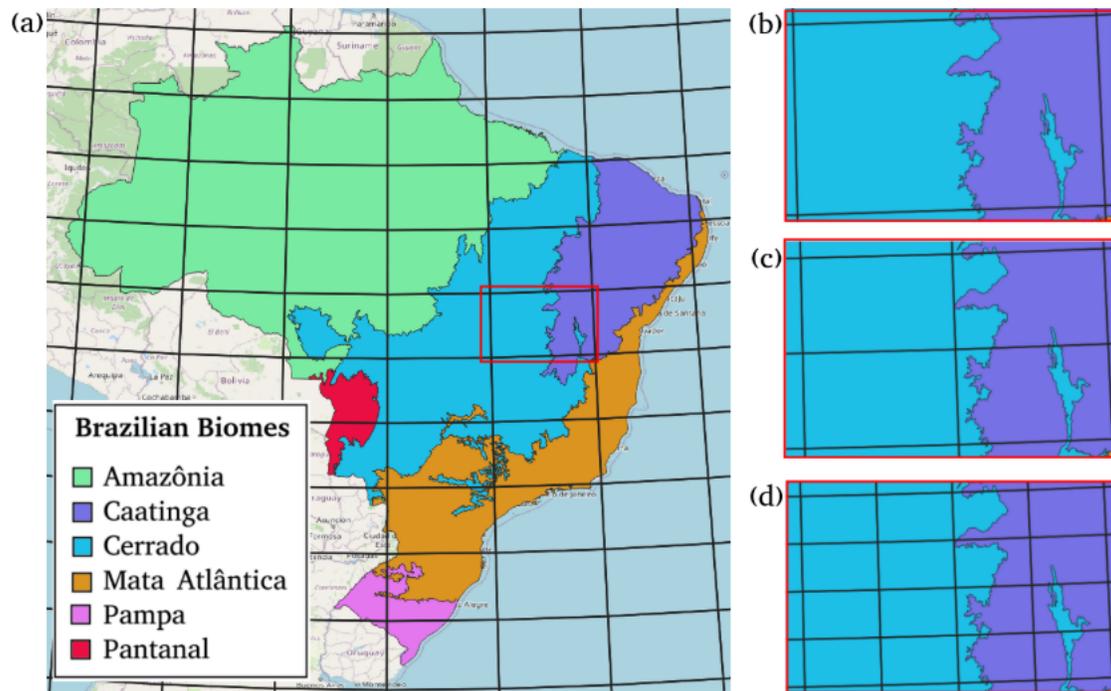
A organização dos cubos de dados é baseada em uma partição espacial do território nacional. Essa partição é representada por uma grade composta por tiles (células), de modo que os cubos de dados são gerados tile a tile fornecendo para cada banda espectral um único arquivo em disco. Considerando a resolução espacial de cada dado e visando gerar arquivos que possam ser facilmente manipulados, três grades, que seguem uma ordem hierárquica, foram criadas. Essas grades utilizam a projeção Albers Equal-Area cônica e o Datum SIRGAS 2000.

As grades BDC apresentam tiles de $1^\circ \times 1.5^\circ$, $2^\circ \times 3^\circ$, $4^\circ \times 6^\circ$ e são denominadas, BDC_SM, BDC_MD, BDC_LG, respectivamente. Os cubos de dados Sentinel-2/MSI utilizam a grade BDC_SM, os cubos de dados Landsat-8/OLI e CBERS-4/MUX utilizam a grade BDC_MD e o cubo CBERS-4/WFI utiliza a grade BDC_LG.

Uma das ferramentas contidas no Data Cube Builder e implementada durante o período deste relatório, é um gerador de grades, o mesmo capaz de gerar as grades BDC. Para gerar uma grade, inicialmente posiciona-se um ponto, chamado semente, na intersecção entre equador e um meridiano (meridiano 54° oeste, no caso da grade BDC). Esse ponto semente é adotado como o canto superior direito de um tile, de modo que seguindo um plano cartesiano, sempre definido pelas coordenadas na projeção Albers Equal-Area cônica os demais tiles são definidos seguindo um tamanho definido pelo usuário. Assim, para cada tile, os cantos superiores apresentam perfeitamente os mesmos valores dos cantos inferiores dos tiles acima e os cantos direitos apresentam perfeitamente os mesmos valores dos cantos esquerdos dos tiles vizinhos lateralmente. A hierarquia entre as grades BDC pode ser observada na

Figura 6.

Figura 6: Grades do BDC (a) Grade BDC_LG sobreposta nos Biomas Brasileiros. (b) Detalhe de um tile da grade BDC_LG. (c) Um tile BDC_LG corresponde a 4 tiles da grade BDC_MD. (d) 16 tiles BDC_SM correspondem a um tile BDC_LG ou 4 tiles BDC_MD.

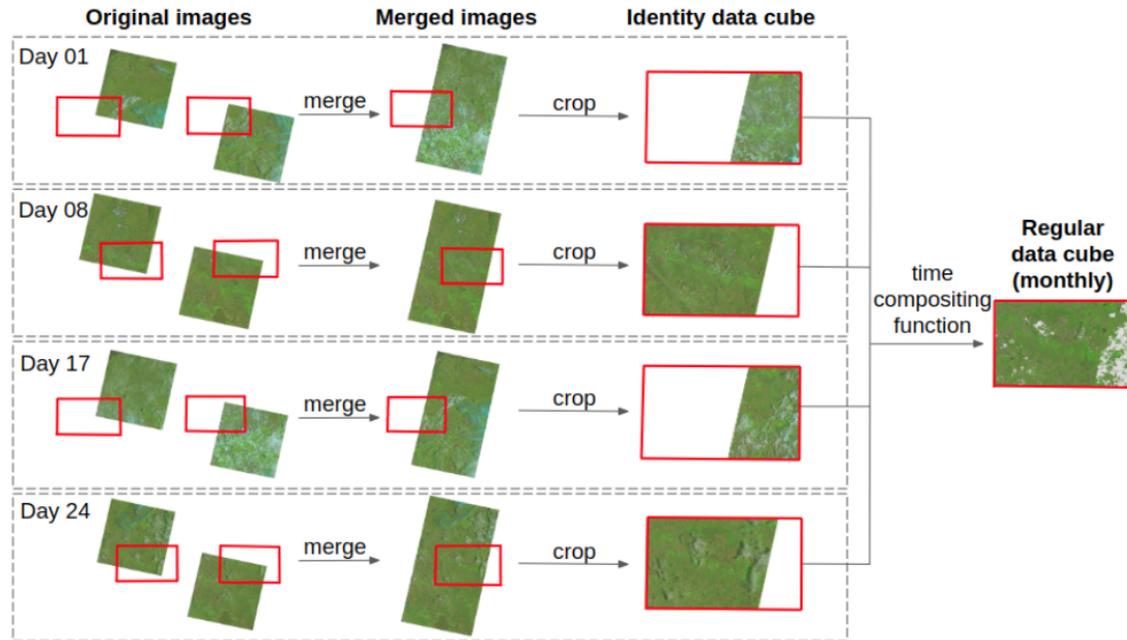


Source: [1]

Os cubos de dados são gerados considerando essas grades do BDC. Os cubos de dados podem ser categorizados em dois tipos, cubos identidade e cubos compostos no tempo. A Figura 7 ilustra o processo de geração dos cubos.

A primeira etapa no processo de geração de cubos de dados pelo Data Cube Builder, consiste em consultar todas as imagens disponíveis para uma determinada área, no caso um tile BDC. Essa consulta é feita utilizando o serviço STAC, que lista todas as imagens publicadas para uma determinada coleção. As imagens resultantes

Figura 7: Processo de geração de um cubo de dados mensal, considerando um tile de uma grade.



Source: [1].

dessa busca são então reprojctadas para a projeção da grade do cubo sendo gerado, caso existam diferentes resoluções espaciais no dado de entrada, é feito uma reamostragem bilinear para a resolução espacial mais fina, nas imagens de mesma data são feitos mosaicos espaciais e por fim as imagens são recortadas para a área do tile que esta sendo processado. Esse processo gera os chamados cubos identidade, por vezes também chamados cubos irregulares. Assim, cubos identidade representam os dados originais de entrada em sua data de aquisição original, reprojctados, recortados, reamostrados e mosaicados para a grade BDC. Entretanto, séries temporais extraídas a partir de cubos identidade podem não apresentar observações equidistantes no tempo, dificultando sua utilização por algoritmos de classificação. Assim, uma

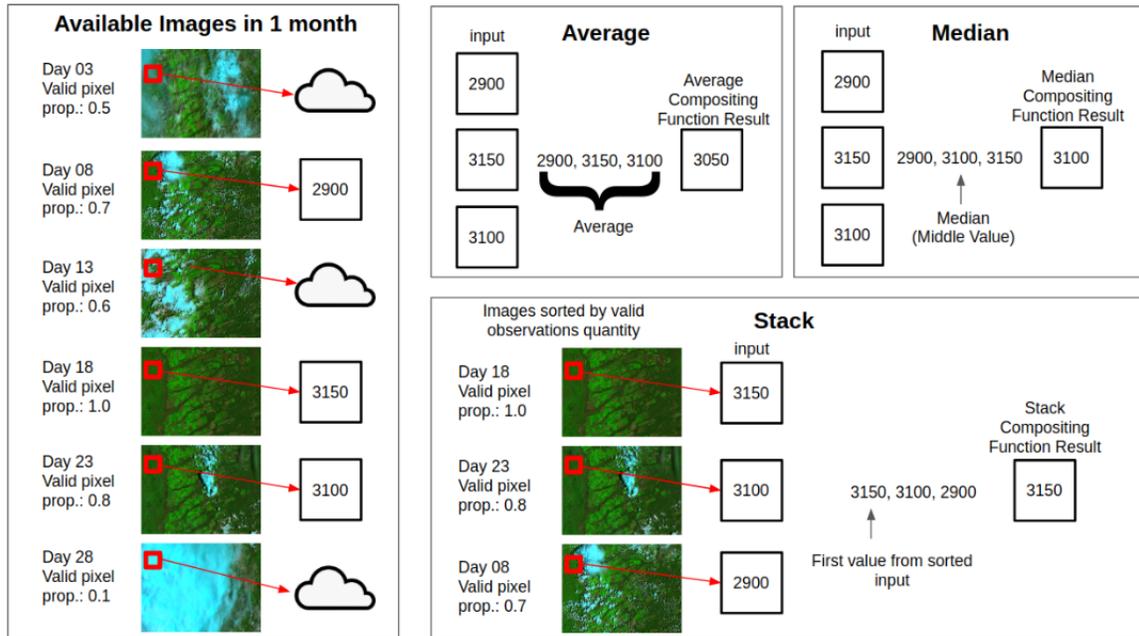
vez em posse dos cubos identidade é possível aplicar funções de composição temporal para gerar cubos regulares no tempo, por vezes também chamados cubos regulares.

A função de composição temporal calcula ou seleciona um valor de pixel que representa os dados contidos em um intervalo de tempo. Ao aplicar uma função de composição temporal à um cubo identidade, tem-se um cubo composto no tempo. A Figura 8 ilustra algumas das funções de composição temporal, sendo elas média, mediana e STACK. Dado um intervalo de tempo, por exemplo, 16 dias ou 1 mês, a composição temporal por média avalia todos os valores livre de nuvens e sombra de nuvens obtidos utilizando a banda de máscara de nuvens para calcular o valor médio para cada uma das bandas espectrais. A função de composição temporal por mediana, funciona de forma similar só que obtendo o valor central dentre os disponíveis, ao invés do valor médio. Já a função de composição temporal STACK, consiste em uma seleção de melhor pixel. No caso, da STACK, inicialmente é feito uma ordenação das imagens com mais área útil, ou seja menor quantidade de pixels contendo nuvens, sombra de nuvens e maior quantidade de pixels sem estes alvos, assim, o valor resultante da composição STACK consiste no primeiro valor encontrado, livre de nuvens e sombras de nuvens, considerando essa ordenação.

No período deste relatório, diversas melhorias foram integradas ao software Cube Builder. Tais melhorias incluem uma interface gráfica, contendo todas as ferramentas para criação de grades, criação e processamento dos cubos de dados. Essa interface é ilustrada na Figura 9.

Além de ser capaz de gerar os produtos de cubos de dados, no período referente a este relatório implementou-se a geração de bandas auxiliares para esses cubos

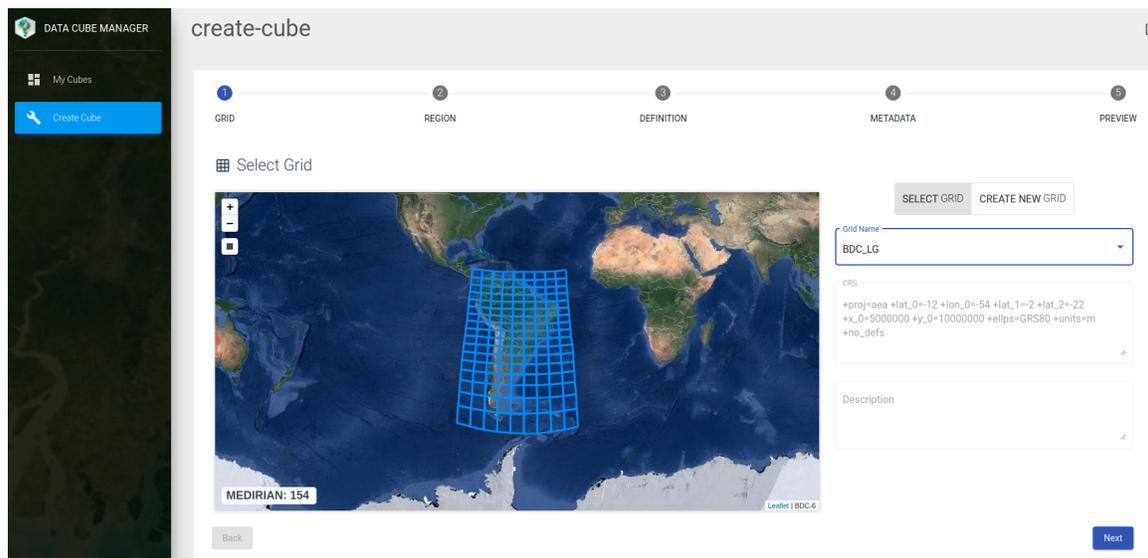
Figura 8: Composição temporal.



Source: [1].

compostos no tempo. Essas bandas são *Provenance*, *TotalOb* e *ClearOb*. Como mencionado previamente, a função de composição temporal STACK, consiste em uma seleção de melhor pixel, assim, a banda *Provenance*, contem a informação da data de origem da imagem, em formato de dia juliano, ou seja com valores que variam entre 1 e 366. Já a banda *TotalOb*, indica para cada pixel dos cubos o número total de observações feitas dentro do período agrupado, por exemplo, a coleção de imagem Sentinel-2 apresenta uma resolução temporal de 5 dias, assim, um cubo composto para um período de 16 dias é composto geralmente por 3 observações dessa coleção (exceções para áreas de sobreposição de órbitas ou quando há algum problema na aquisição das imagens). Por fim, a banda *ClearOb* utiliza as máscaras de nuvens e

Figura 9: Interface Data Cube Builder.



sombra de nuvens das imagens utilizadas para compor o cubo para informar quantas observações dentre as disponíveis para o período de composição não apresentavam as feições de nuvem ou sombra de nuvens.

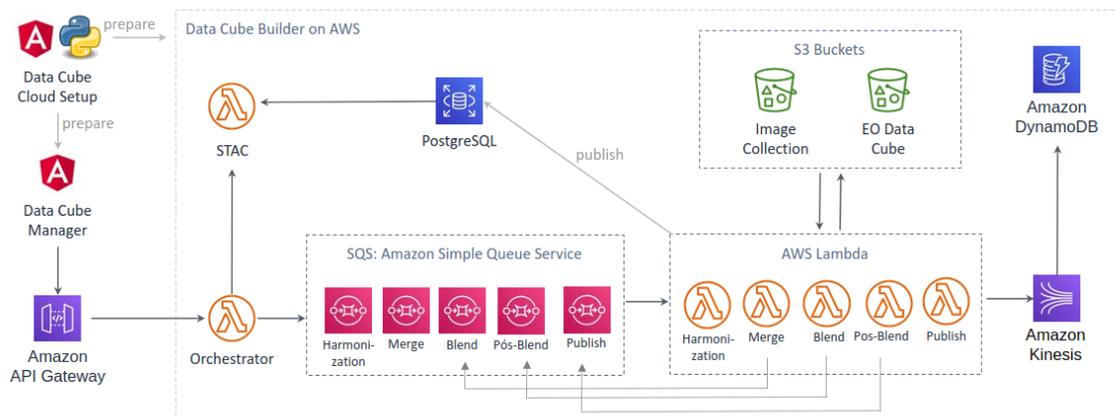
Adicionalmente as bandas de informação auxiliar, a ferramenta Data Cube Builder também foi incrementada para ser capaz de gerar índices de vegetação, como por exemplo o Índice de Vegetação da Diferença Normalizada (NDVI) e o Índice de vegetação melhorada (EVI). Também foram adicionados parâmetros que permitem a verificação de valores saturados ou faltantes em alguma bandas espectrais para que estes não sejam utilizados para compor os cubos, sendo substituídos pelo valor válido de outra data no produto final.

A ferramenta também foi adaptada para ser capaz de gerar cubos de dados a partir de múltiplas coleções, ou seja, uma vez que essas coleções sejam compatíveis entre si, podem ser utilizadas em conjunto para gerar um único cubo de dados. Logo,

essa implementação torna capaz a geração de cubos multi-sensores. Entretanto, deve se ter cautela ao unir dados de fontes diversas, deste modo também foi adicionado o suporte a correção de BRDF utilizando a técnica do c-fator [8], que consiste em um método para compatibilização de sensores como Landsat-8 e Sentinel-2. Assim, dois estudos estão previstos para serem desenvolvidos no decorrer do projeto: o primeiro testando tais correções de BRDF em produtos Landsat-8 e Sentinel-2, e o segundo envolvendo a harmonização de diversos sensores do programa Landsat, no caso Landsat-5, Landsat-7 e Landsat-8.

Além de ser capaz de gerar os produtos de cubos de dados utilizando os servidores do projeto BDC, adaptações também foram feitas para que houvesse uma versão da ferramenta capaz de funcionar em ambiente de nuvem na *Amazon Web Service* (AWS). Assim, os produtos de cubos de dados podem ser gerados tanto na infraestrutura local do BDC como também no ambiente de nuvem da AWS, como pode ser observado na Figura 10.

Figura 10: AWS BDC Cube Builder.

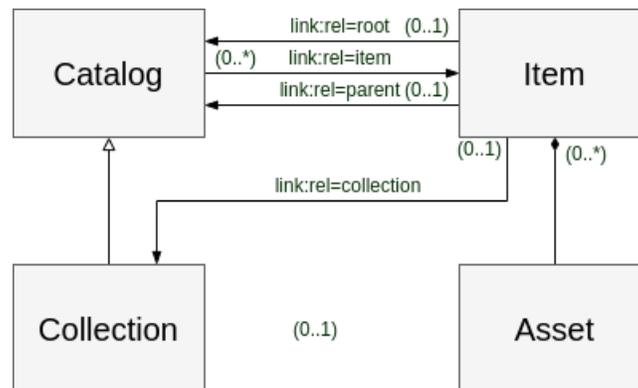


3.2 Serviços

3.2.1 BDC-STAC

Para descoberta dos dados do BDC, foi desenvolvido um serviço *web* REST na linguagem Python que utiliza o *microframework* Flask. Este serviço cataloga os metadados do BDC seguindo a especificação STAC¹. A especificação do STAC diz como os metadados devem ser descritos e acessados. O serviço implementa uma STAC API, que possibilita a realização de requisições através métodos GET e POST para consultar os metadados dos cubos. Os resultados das consultas utilizam o padrão JSON e GeoJSON.

Figura 11: Modelo de representação dos dados através do STAC



A especificação do STAC oferece os seguintes conceitos:

- **Catalog**: provê uma estrutura de links contendo os diversos Items do STAC, coleções ou mesmo para outros catálogos.

¹<https://github.com/radiantearth/stac-spec>

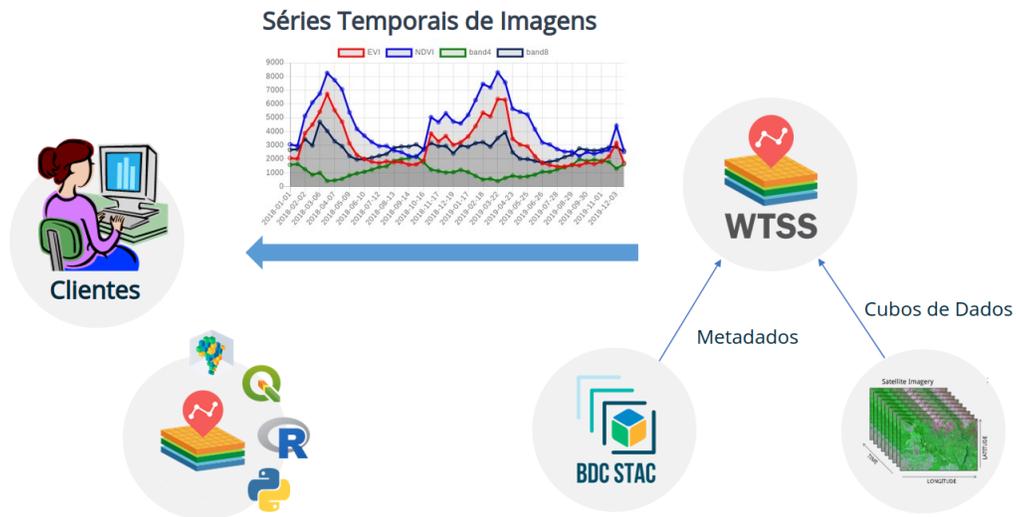
- **Collection:** é uma especialização do catalogo que oferece informação sobre os metadados de uma coleção de produtos de dados.
- **Item:** é uma *feature* descrita em GeoJSON contendo os metadados de uma cena de um produto de dados, tais como retângulo envolvente espacial, tempo e outros links vinculados nos Assets. Conceitualmente, um item pode conter um ou vários Assets.
- **Asset:** representa um arquivo contendo os dados espaço temporais. Um asset também pode representar os arquivos auxiliares vinculados ao item.

Além do serviço STAC, foram desenvolvidos dois clientes baseados na API STAC, o `stac.py` pacote em linguagem Python, e pacote `rstac` em R [9].

3.2.2 Web Time Series Service

Web Time Series Service (WTSS) é um serviço web para recuperação das series temporais das imagens de sensoriamento remoto. O serviço preenche a lacuna entre as aplicações de sensoriamento remoto e seus requisitos de padrão de acesso a dados por meio de API simples e robusta. Uma API simples e bem definida facilita a construção de novos aplicativos capazes de consumir as séries temporais e, assim, construir novos produtos derivados. Para o funcionamento com os cubos de dados do projeto BDC, o WTSS utiliza o serviço do STAC para consulta dos metadados e o acesso aos dados físicos nos servidores do projeto, conforme apresentado na Figura 12.

Figura 12: Web Time Series Service



Essencialmente, o serviço oferece três operações:

- **list_coverages**: Recupera a lista com os nomes das cubos de dados ou *coverages* disponíveis. A Figura 13 apresenta o documento de saída no formato JSON.

Figura 13: WTSS: Operação *list_coverages* apresentados os cubos de dados do projeto BDC

```
{
  "coverages": [
    "MOD13Q1-6",
    "MYD13Q1-6",
    "S2_10_16D_STK-1",
    "S2-SEN2COR_10_16D_STK-1",
    "S2_10-1",
    "CB4MUX_20-1",
    "CB4MUX_20_1M_STK-1",
    "LC8_30-1",
    "LC8_30_16D_STK-1",
    "CB4_20_1M_STK-1",
    "LC8_30_6M_MEDSTK-1",
    "CB4_64_16D_STK-1",
    "CB4_64-1",
    "LANDSAT-MOZ_30_1M_STK-1"
  ]
}
```

- **describe_coverage**: Esta operação recupera os metadados de uma coverage identificada pelo parâmetro *name*. Através dessa operação, é possível identificar os limites espaciais do cubo de dados selecionado, bem como o limite espacial para então recuperar as séries temporais utilizando a operação *time_series*.
- **time_series**: Possibilita recuperar as séries temporais dos cubos de dados do projeto. Os parâmetros de requisição estão descritos na Tabela 5

Este serviço pode ser facilmente integrado em tecnologias de código aberto, como Python através da biblioteca `wtss.py`², R com a biblioteca `Rwtss`³ e também aplicações web para análise e visualização de dados de maneira interativa como JupyterHub. A definição da especificação do serviço está disponível no [github](#) do BDC.

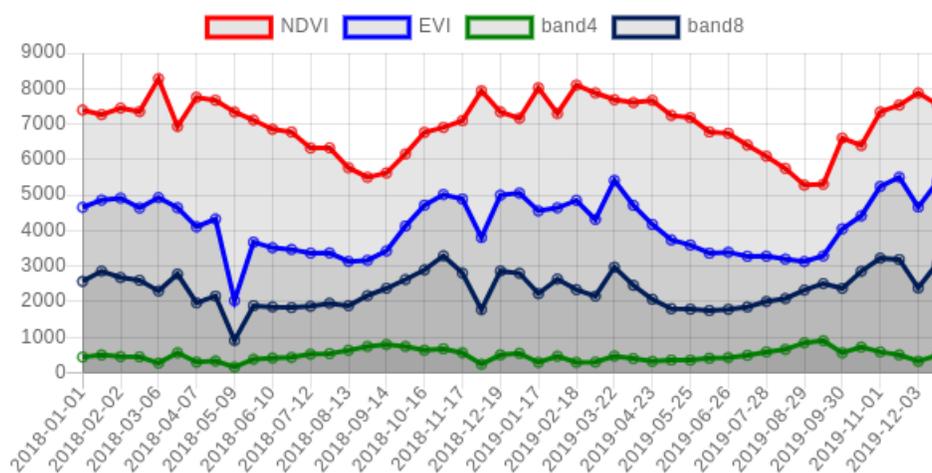
²<https://pypi.org/project/wtss/>

³<https://github.com/e-sensing/Rwtss>

Tabela 5: Parâmetros associados à operação *time_series*

Parâmetro	Descrição
coverage	nome da coverage
attributes	lista com o nome dos atributos
latitude	latitude (em EPSG:4326)
longitude	longitude (em EPSG:4326)
start_date	Tempo inicial da série temporal
end_date	Tempo final da série temporal

Figura 14: WTSS: Operação *time_series*



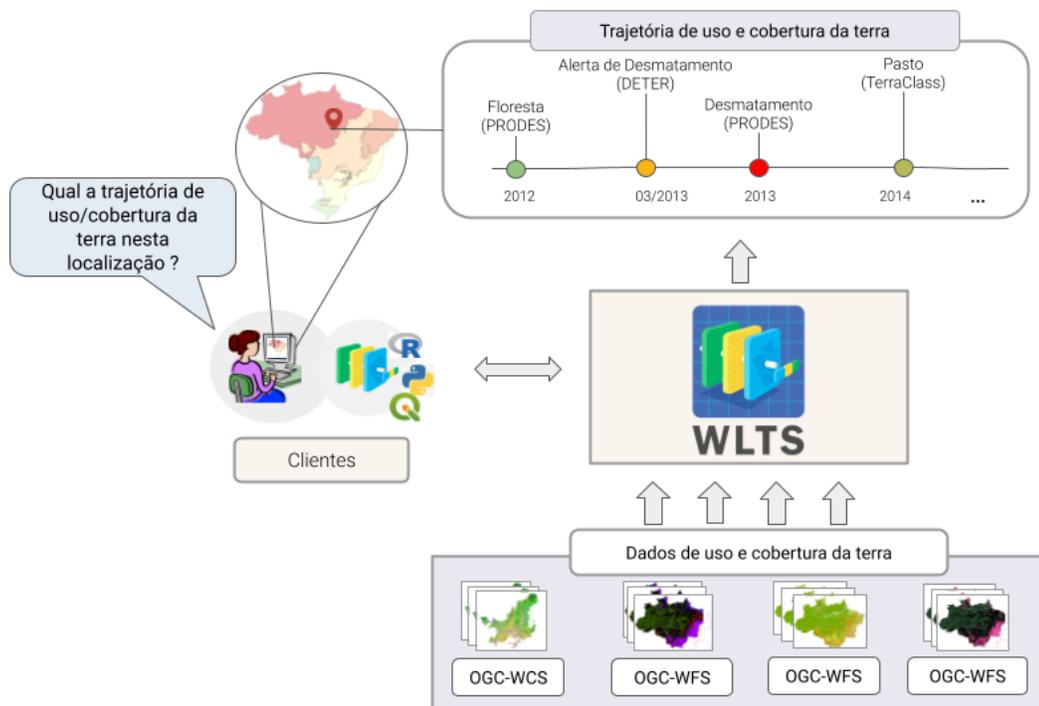
3.2.3 Web Land Trajectory Service

O WLTS é um serviço web com o propósito de facilitar o acesso às diversas coleções de dados de uso e cobertura da Terra através de uma API projetada com base no conceito de trajetórias. A abordagem WLTS é usar um modelo de dados que define um conjunto mínimo de informações temporais e espaciais para representar diferentes fontes e tipos de dados.

A Figura 15 apresenta a visão geral sob a ótica de um usuário, que requisita uma

trajetória de UCT associada a uma dada localização no espaço geográfico. O usuário realiza uma requisição aos serviços por intermédio de um cliente, desenvolvidos na linguagem Python e R, nomeados respectivamente de `wlts.py`⁴ e `rwlts`⁵. O usuário também pode consultar o serviço WLTS utilizando o plugin para o QGIS. Com a informação de uma localização espacial fornecida pelo usuário, o serviço consulta o conjunto de dados disponíveis retornando a trajetória de UCT.

Figura 15: Web Land Trajectory Service



A API definida pelo serviço WLTS é composta de três operações:

- `wlts/list_collections`: retorna a lista com os nomes de todas as coleções disponíveis em uma instância do serviço. Os nomes retornados por essa operação

⁴<https://github.com/brazil-data-cube/wlts.py>

⁵<https://github.com/brazil-data-cube/rwlts>

podem ser utilizados nas operações subseqüentes para identificar uma coleção de interesse.

- **wlts/describe_collection**: recupera os metadados de uma coleção em particular identificada pelo parâmetro `collection_id`. Esses metadados incluem uma descrição da coleção, as informações espaciais e temporais. Os metadados também incluem informações sobre o sistema de classificação da coleção.
- **wlts/trajectory**: retorna a trajetória de uso e cobertura da terra de uma ou mais coleções dada uma localização espacial. A propriedade `result` contém as informações associado ao item de dados: `class`, `date` e `collection`.

Os parâmetros de requisição estão descritos na Tabela 6.

Tabela 6: Parâmetros associados à operação *trajetória*

Parâmetro	Descrição
<code>collections</code>	lista com o nome das collection
<code>latitude</code>	latitude (em EPSG:4326)
<code>longitude</code>	longitude (em EPSG:4326)
<code>start_date</code>	Tempo inicial da trajetória
<code>end_date</code>	Tempo final da trajetória

Os clientes encapsulam as consultas à API do serviço WLTS. Permitindo aos pesquisadores trabalharem mais próximo do domínio de aplicação. Os clientes oferecem suporte à construção de aplicações e uso em ambientes de computação interativa como *IPython/Jupyter*.

O WLTS pode ser utilizado em diversas aplicações, como na validação de conjuntos de dados de cobertura da terra, na seleção de amostras de treinamento para dar suporte aos algoritmos de aprendizado de máquina supervisionados usados na

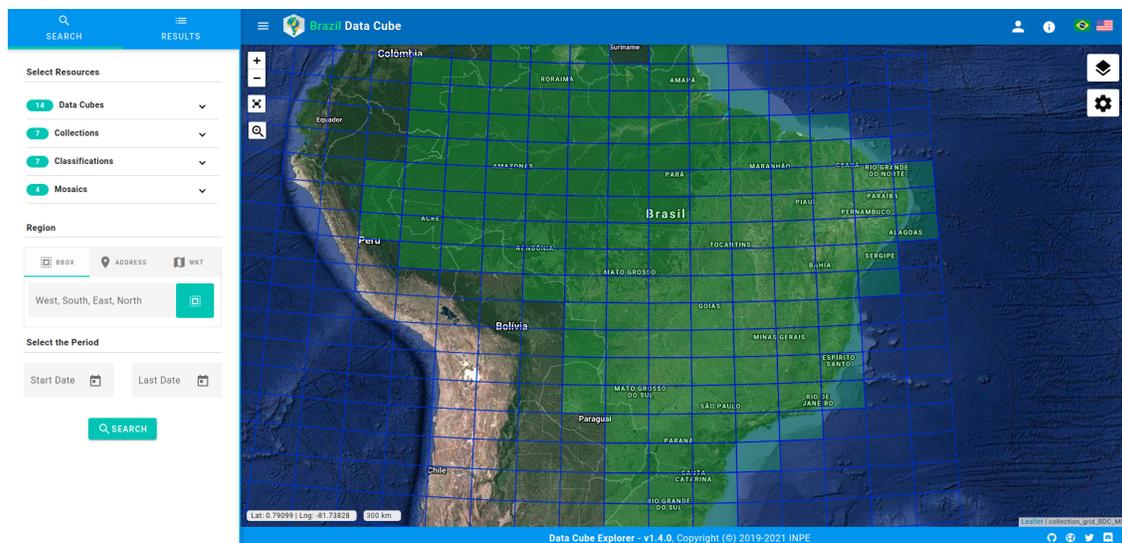
geração de novos mapas de classificação. A definição da especificação do serviço está disponível em no [github](#) do BDC.

3.3 Aplicações

3.3.1 BDC Explorer

O BDC Explorer, mostrado na Figura 16, é uma aplicação web responsável pela visualização das coleções imagens, cubos de dados, classificações e mosaicos gerados pelo projeto BDC (link: <http://brazildatacube.dpi.inpe.br/portal/explore>). Através dessa aplicação web, usuários podem acessar e visualizar os dados disponibilizados pela equipe do projeto em um mapa iterativo, usando um componente de controle temporal.

Figura 16: BDC Explorer, v1.4.0.



Na Figura 16, os tiles em verde da grade representam as regiões com dados

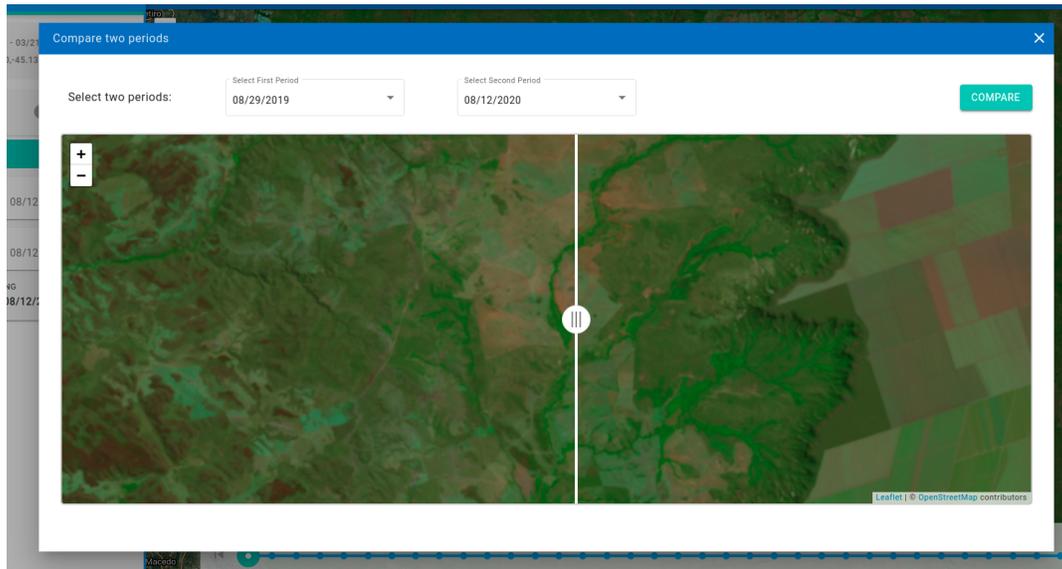
disponíveis na plataforma para um determinado conjunto de imagens, nesse exemplo, temos a visualização para o cubo de dados do Landsat-8 de 16 dias e a função de composição STACK.

Essa aplicação, além das funcionalidades de controle temporal, também fornece outros componentes com o intuito de auxiliar o usuário na navegação e recuperação dos dados, como:

- **Componente de Search:** permite a consulta de imagens disponíveis no sistema, através de um filtro temporal e espacial.
- **Componente de Comparação de imagens:** permite uma análise visual de diferentes imagens em dois períodos de tempo para um determinado conjunto de dados. Como pode-se observar na Figura 17.
- **Componente de Visualização de Series Temporais:** possibilita a extração da series temporais de um ponto, usando um cubo de dados e suas bandas desejadas. Para o funcionamento desse componente, é utilizado o serviço WTSS (3.2.2).
- **Componente para composição de Imagens coloridas:** Permite a composição e visualização das imagens RGB no mapa, de acordo com as bandas escolhidas. Além disso, essa interface também permite a edição da opacidade, saturação e contraste da imagem, em tempo de execução, evitando assim que o usuário precise realizar o download.
- **Componente de Download:** permite o download dos dados disponíveis na plataforma em larga escala para que o usuário possa baixar imagens específicas ou

o conjunto completo de uma determinada coleção de imagem, cubo de dados, classificação ou mosaico.

Figura 17: Componente para comparação de imagens.



Esse sistema está em sua versão 1.4.0 e realiza todo o controle de acesso dos usuários, através de um sistema de autenticação, desenvolvido pela equipe do projeto BDC.

3.3.2 SITS

O `sits` é um pacote R de código aberto para análise de séries temporais de imagens de satélite utilizando aprendizagem de máquina que adota uma abordagem time-first, space-later [10]. Sua principal contribuição é fornecer um fluxo de trabalho completo para a classificação de cobertura da terra de grandes conjuntos de dados de Observação da Terra. O `sits` oferece um conjunto de facilidades aos usuários,

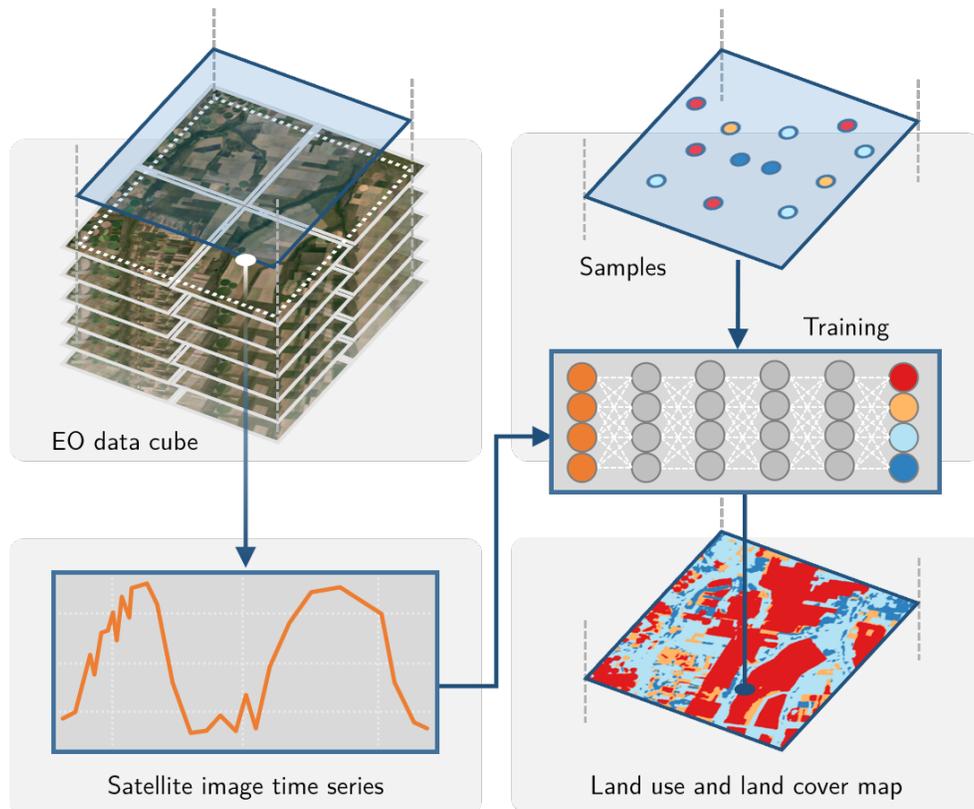
incluindo:

- acessar os cubos de dados do projeto BDC;
- recuperar séries temporais a partir desses cubos;
- melhorar a qualidade das amostras de treinamento;
- treinar modelos de aprendizado de máquina para realizar classificações de cobertura da terra;
- remover *outliers* de classificação usando suavizações espaciais;
- avaliar os produtos de cobertura da terra gerados.

O pacote `sits` incorpora os mais recentes avanços em catálogos de imagens para serviços de computação em nuvem oferecendo suporte a especificação STAC. Esse suporte é implementado através do pacote em R `rstac` [9]. O `sits` implementa uma especialização de cubo de dados, que juntamente com o pacote `rstac`, reúne todos os metadados necessários para a geração de mapas de uso e cobertura da terra. O pacote `sits` também inclui diferentes métodos de aprendizagem de máquina e de *deep learning* para classificação de séries temporais incluindo novos métodos propostos na literatura de séries temporais de imagens de satélite [11, 12], e métodos para o controle de qualidade de amostras de treinamento [13]. O gerenciamento dos recursos computacionais é resolvido pelo `sits`, e ele é capaz de realizar processamento paralelo dos cubos de dados, o que permite obter um bom desempenho de tempo nas classificações.

A concepção da API `sits` considera o fluxo de trabalho típico para a classificação de áreas utilizando séries temporais de imagens de satélite (ver figura 18). Os usuários definem um cubo de dados selecionando um subconjunto de uma coleção de imagens ARD. Eles obtêm os dados de treinamento a partir de um conjunto de pontos no cubo de dados cujos rótulos são conhecidos. Após realizar o controle de qualidade das amostras de treinamento, os usuários constroem um modelo de aprendizagem da máquina e o utilizam para classificar o cubo de dados completo. Os resultados passam por uma fase de suavização espacial que remove os *outliers*.

Figura 18: Uso de séries temporais para classificação do uso e cobertura da terra



Fonte: [10]

O `sits` armazena séries temporais em uma tabela de relacional de objetos. Essa tabela contém dados e metadados (por exemplo, informações espaciais e temporais, o rótulo atribuído à amostra, o cubo de dados de onde os dados foram extraídos e os dados das séries temporais para cada localização espacial-tempo) e pode ser manipulada por funções nativas da linguagem. A localização espacial é dada em coordenadas de longitude e latitude para o elipsoide WGS84.

A experiência com métodos de aprendizagem de máquinas mostra que o fator limitante na obtenção de bons resultados é o número e a qualidade das amostras de treinamento. Conjuntos de dados grandes e precisos são melhores, não importa o algoritmo utilizado [14,15], enquanto amostras ruidosas e imperfeitas têm um efeito negativo no desempenho da classificação [16]. O software que utiliza o aprendizado de máquinas para análise de imagens de satélite precisa de bons métodos para o controle da qualidade das amostras.

O pacote `sits` fornece uma técnica inovadora de controle de qualidade de amostras baseada em mapas auto-organizados (SOM) [13,17]. O SOM é uma técnica de redução de dimensionalidade. Os dados de alta dimensão são mapeados em duas dimensões, mantendo as relações topológicas entre padrões similares [18]. Os dados de entrada para avaliação da qualidade são um conjunto de amostras de treinamento, obtidas conforme descrito acima. Ao representar um conjunto de dados de alta dimensão de amostras de treinamento em um mapa 2D auto-organizado, as unidades do mapa (chamadas “neurônios”) competem por cada amostra. Espera-se que amostras de boa qualidade de cada rótulo estejam próximas umas das outras no mapa resultante. Os vizinhos de cada neurônio de um mapa SOM fornecem informações

sobre a variabilidade intraclasse e interclasse.

Depois de selecionar amostras de boa qualidade, o próximo passo é treinar um modelo de aprendizagem de máquina. O pacote fornece suporte para a classificação das séries temporais, preservando a resolução temporal completa dos dados de entrada. Ele suporta os seguintes algoritmos de aprendizagem de máquina:

- *Linear discriminant analysis.*
- *Quadratic discriminant analysis.*
- *Multinomial logit.*
- *Support vector machines.*
- *Random forests.*
- *Extreme gradient boosting.*
- *Deep learning (DL) using multi-layer perceptrons.*
- *Deep Residual Networks.*
- *TempCNN.*

O pacote `sits` pode ser utilizado em qualquer ambiente de computação que suporte a linguagem R. Ao trabalhar com grandes dados de Observação da Terra, o ambiente ideal para o `sits` é uma máquina virtual localizada perto do repositório de dados. Para alcançar eficiência, o `sits` implementa seu próprio processamento paralelo. Os usuários não são obrigados a aprender a fazer multiprocessamento e, portanto, sua curva de aprendizado é reduzida.

A classificação de imagens no `sits` é feita por um *cluster* de processamentos independentes ligados a uma ou mais máquinas virtuais. Para evitar a sobrecarga de comunicação, os dados são lidos e armazenadas de forma independente; a interação direta entre o processo principal e os processos é mantida em um nível mínimo.

Quando se trabalha com grandes conjuntos de dados de Observação da Terra, há uma considerável variabilidade de dados em cada classe. Como resultado, alguns pixels poderão ser classificados erroneamente. Estes erros são mais prováveis de ocorrer em áreas de transição entre classes ou quando se lida com pixels mistos. Para compensar estes problemas, o `sits` inclui um método de suavização de pós-processamento baseado na probabilidade Bayesiana que utiliza informações da vizinhança de um pixel para reduzir a incerteza sobre seu rótulo.

3.3.3 Open Data Cube

Open Data Cube (ODC) é um *framework* composto por diversas ferramentas e serviços que permite a catalogação e análise de grandes volumes de dados de observação da Terra. Para isso, o *framework* disponibiliza diversos componentes, dentre eles [19]:

- `datacube-explorer`: aplicação *web* para busca e exploração de dados;
- `datacube-stats`: ferramenta de linha de comando (CLI, do inglês *Command Line Interface*) para análise temporal de grandes volumes de dados de Observação da Terra, utilizando técnicas de processamento distribuído e paralelo;
- `datacube-ui`: aplicação *web* para visualização e execução de algoritmos de

análise de dados;

- **datacube-ows**: aplicação que fornece acesso aos dados através dos serviços *web* OGC *Web Map Service* (WMS), *Web Map Tile Service* (WMTS), *Web Coverage Service* (WCS).

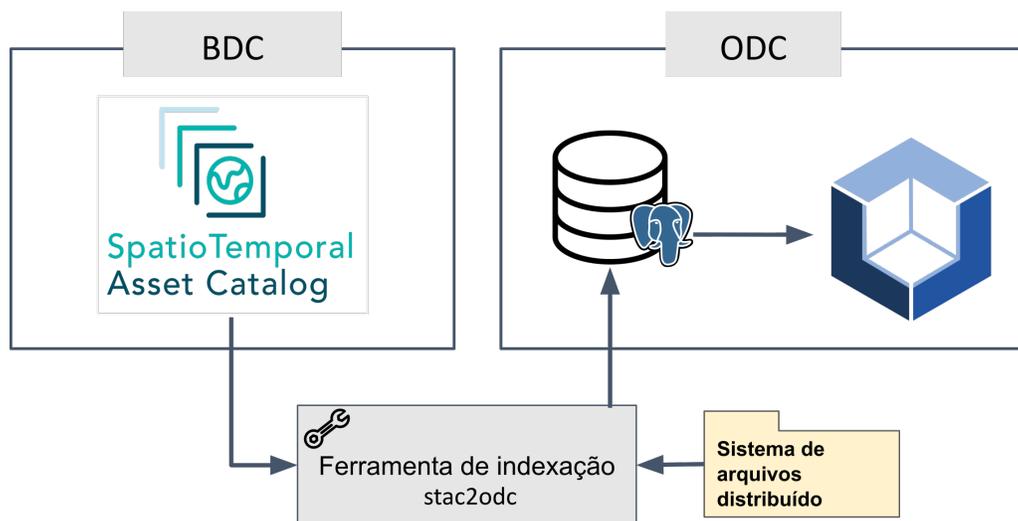
Além desses, há também o **datacube-core**, o principal componente do ecossistema ODC. Ele disponibiliza funcionalidades que permitem as operações de catalogação, busca e recuperação de dados. Utiliza em seu funcionamento um Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL, no qual é feita a indexação dos dados. Neste processo, tem-se o registro dos metadados, com as informações espaço-temporais dos dados e o local onde estão armazenados. As operações de busca e recuperação utilizam essas informações em suas respectivas funcionalidades. Esse componente é implementado na linguagem de programação Python, sendo utilizado por todos os demais componentes do ecossistema ODC para a manipulação dos dados.

Com o objetivo de ampliar os serviços e ferramentas que podem ser utilizados pelos usuários da Plataforma BDC para acesso, visualização e análise de dados, fez-se a integração entre os componentes do ODC e os produtos de dados do BDC. A integração foi realizada em um processo dividido em duas partes: (i) criação da instância ODC; (ii) configuração e personalização dos componentes ODC para funcionamento junto aos dados do BDC.

Na primeira etapa, de criação da instância ODC na infraestrutura BDC (**bdc-odc**), inicialmente fez-se a configuração de um servidor PostgreSQL. Em seguida, o **datacube-core** foi configurado. Para a indexação dos dados do BDC no **bdc-odc**, foi feita a

criação da ferramenta CLI `stac2odc` [20]. Esta ferramenta permite que os metadados dos dados do BDC, disponibilizados através da interface `bdc-stac`, possam ser utilizados nas tarefas de indexação do `datacube-core`. Com base nesta ferramenta, seguindo o fluxo ilustrado na Figura 19, os cubos de dados AWF/IBERS-4, OLI/Landsat-8 e MSI/Sentinel-2 do BDC foram indexados no `bdc-odc`.

Figura 19: Fluxo de indexação com o `stac2odc`.



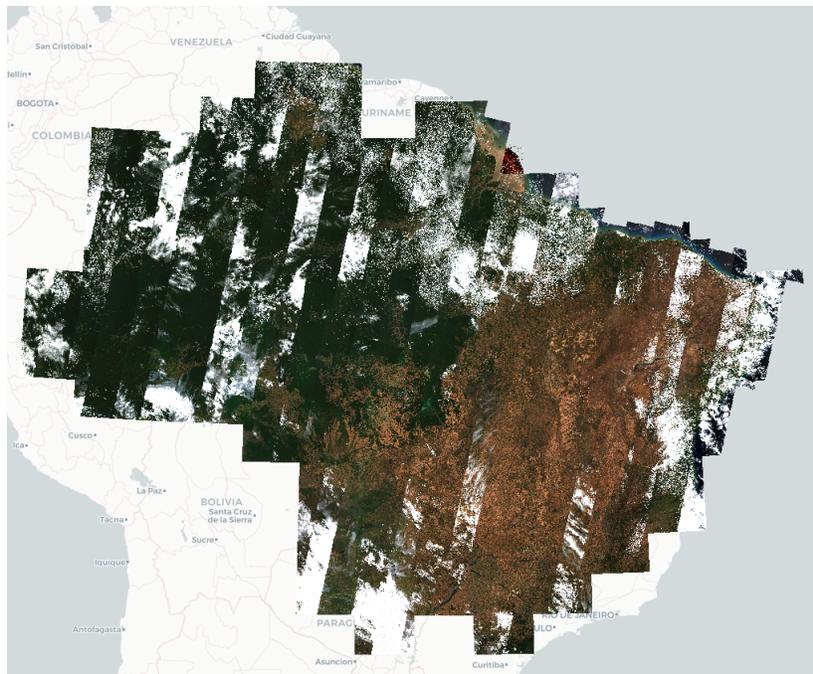
Fonte: [20]

De modo a potencializar o uso dessa abordagem criada no BDC, o `stac2odc` foi desenvolvido de forma genérica [21], permitindo que outros pesquisadores e instituições façam seu uso em qualquer interface STAC, tendo esta relação ou não ao serviço `bdc-stac`.

Após a criação do `bdc-odc` e indexação dos dados, fez-se a segunda etapa da integração. Nessa etapa, os componentes do ODC `datacube-explorer`, `datacube-stats`, `datacube-ui` e `datacube-ows` foram configurados no `bdc-odc`. Para isso, foram ne-

cessárias adaptações no código fonte desses componentes de modo a comportar as características e necessidades da Plataforma BDC. Ao final deste processo, os produtos de dados do BDC passaram a estar disponíveis para uso através das ferramentas e serviços do ODC. Para exemplificar o uso desse serviços, na Figura 20 é ilustrado a recuperação de cenas do cubo de dados OLI/Landsat-8 através do serviço OGC WMS disponibilizado pelo componente `datacube-ows`.

Figura 20: Uso de cenas do cubo de dados OLI/Landsat-8 através do serviço OGC WMS.



Fonte: [21]

Além desses componentes, para usufruir das funcionalidades disponibilizadas com a integração, fez-se também a criação da ferramenta `datacube-classification` [22]. Esta ferramenta estende as funcionalidades do componente `datacube-stats` para facilitar

a geração de mapas de uso e cobertura da terra com o uso de séries temporais de Observação da Terra e algoritmos de Aprendizado de Máquina (ML, do inglês *Machine Learning*) no ecossistema ODC. A abordagem metodológica de classificação implementada nessa ferramenta é a mesma descrita no pacote `sits`.

Na versão atual do `datacube-classification`, estão implementadas funcionalidades para a extração de séries temporais de cubos de dados, treinamento e uso de algoritmos de ML para classificação de uso e cobertura da terra. Métodos de interpolação linear temporal, geração de cubos de dados de imagens fração e suavização espacial também estão disponíveis. Por estender o componente `datacube-stats`, a ferramenta também permite que as classificações realizadas sejam feitas utilizando técnicas de processamento distribuído e paralelo.

Por fim, para facilitar o uso da integração realizada, foram criadas também documentações técnicas e exemplos de uso. Esses estão disponíveis no repositório do GitHub [bdc-odc](#) e na [página de documentação](#).

3.3.4 Ambientes Interativos de Desenvolvimento

Para a realização das atividades de processamento e análise dos cubos de dados feitos no projeto BDC, é demandado o uso de recursos computacionais que comumente não estão disponíveis nas máquinas dos colaboradores. Assim, faz-se necessário a disponibilização desses recursos para viabilizar o desenvolvimento dos trabalhos de análise. Na plataforma BDC, essa necessidade foi suprida com a configuração, personalização e disponibilização de ambientes de computação interativa aos usuários.

Tais ambientes são criados com base nas tecnologias Jupyter Hub ⁶ e RStudio⁷, sendo alocados na infraestrutura computacional do projeto [20, 21].

O Jupyter Hub é uma ferramenta que permite o gerenciamento e disponibilização de ambientes de computação interativa Jupyter para múltiplos usuários. Esses ambientes são gerados para cada usuário e permitem a criação dos chamados *notebooks*, que contém narrativas computacionais compostas por códigos, anotações textuais, gráficos e diversos outros elementos, como figuras, vídeos e fórmulas matemáticas em \LaTeX . Esses documentos podem ser editados através das interfaces *web* interativas Jupyter Notebook e Jupyter Lab, sendo esta segunda a utilizada na Plataforma BDC. Na Figura 21, é apresentado um exemplo de *notebook* utilizando o pacote `sits` editado através da interface Jupyter Lab.

Esses ambientes são criados pelo Jupyter Hub através do uso da tecnologia Docker ⁸. Desta forma, uma vez que o usuário acessa o sistema e requisita um ambiente interativo, é feita a criação de um espaço de trabalho específico para aquele usuário. Isso evita conflitos de configuração e acesso não autorizado aos arquivos dos utilizadores. Além disso, cada espaço de trabalho aloca para uso, uma quantidade específica de recurso computacional (memória RAM e núcleos de processamento).

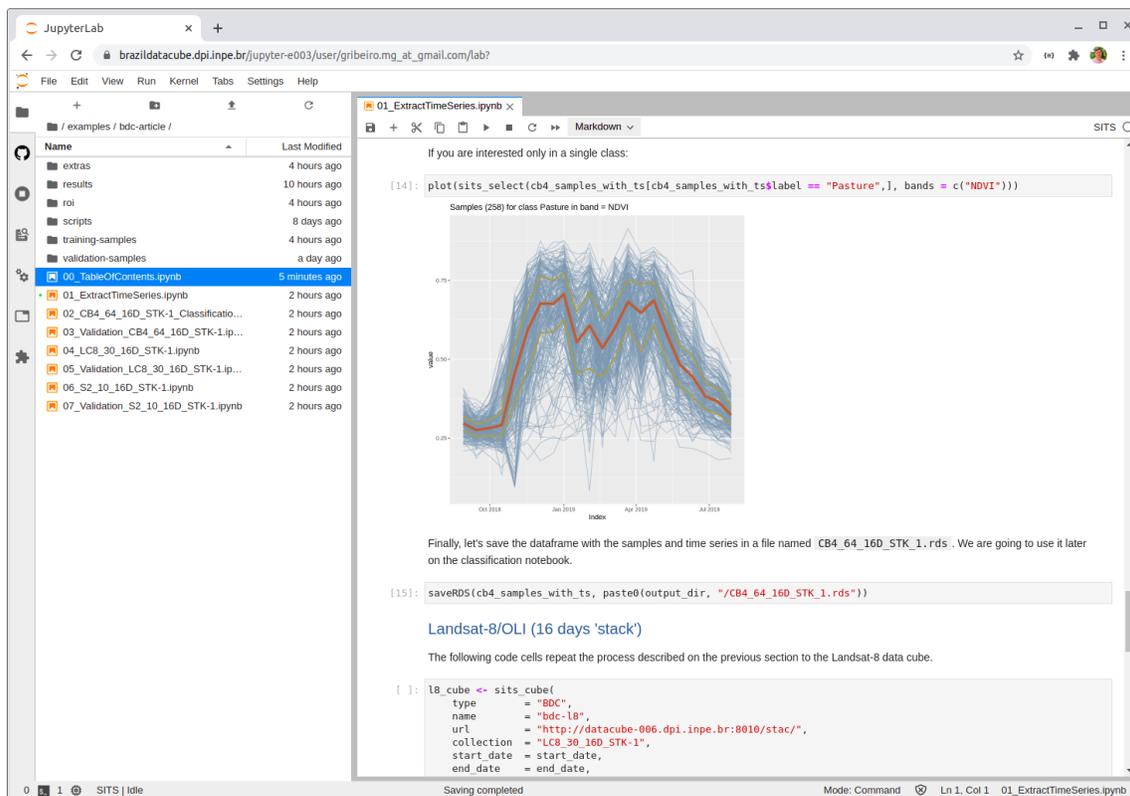
Por utilizar o Docker, o Jupyter Hub permite também que seja feita a customização dos ambientes que serão disponibilizados aos utilizadores, tornando possível a criação de ambientes especializados para diferentes atividades. Por exemplo, pode-se ter um ambiente com bibliotecas de código que permitam o processamento e

⁶Jupyter Hub <https://jupyterhub.readthedocs.io/en/stable/>

⁷RStudio <https://www.rstudio.com/>

⁸Docker <https://www.docker.com/>

Figura 21: Exemplo de um *notebook* sendo editado na interface Jupyter Lab.



análise de dados geoespaciais utilizando a linguagem de programação Python. Outro exemplo é o de um ambiente em que há pacotes da linguagem de programação R específicos para a análise de séries temporais.

Na plataforma BDC, o Jupyter Hub foi configurado para permitir a realização das atividades de análise e processamento de dados. Para isso, fez-se a criação de ambientes personalizados que suprem diferentes necessidades. Atualmente estão disponíveis para os usuários, os seguintes ambientes:

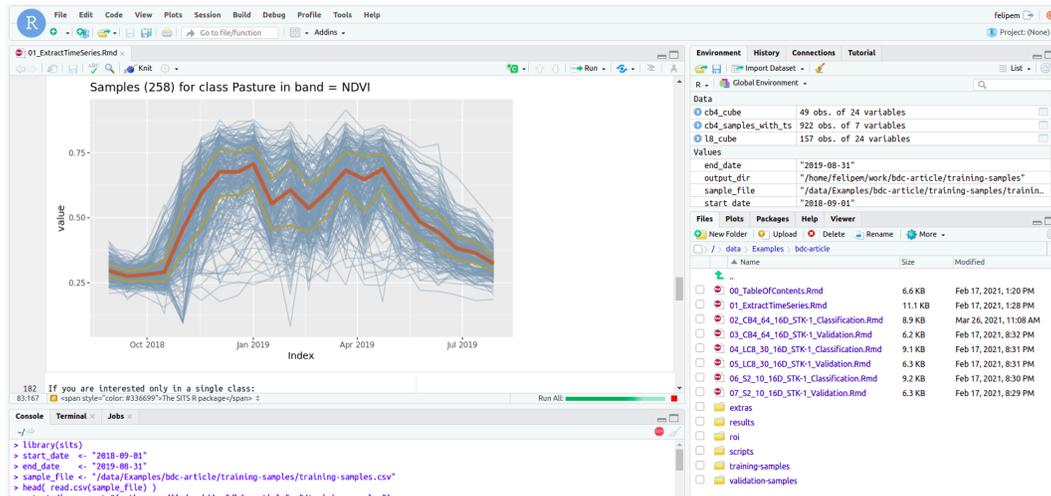
- **Satellite Image Time Series:** ambiente pronto para uso e aplicação do pacote `sits`;

- **Python Geospatial**: ambiente para desenvolvimento de análises de dados geoespaciais na linguagem de programação Python;
- **R Geospatial**: ambiente para desenvolvimento de análises de dados geoespaciais na linguagem de programação R;
- **Open Data Cube**: ambiente pronto para uso e aplicação das ferramentas e funcionalidades do *framework* ODC;
- **Sentinel Toolboxes**: ambiente para desenvolvimento de análises de dados geoespaciais com a ferramenta ESA SNAP.

Os códigos utilizados para a criação e configuração desses ambientes para trabalho integrado à infraestrutura do BDC, estão disponíveis no repositório do GitHub [jupyterhub-docker](#) e no [Docker Hub](#).

Para a disponibilização dos ambientes de computação interativos aos usuários, como citado anteriormente, além do Jupyter Hub, na Plataforma BDC faz-se uso do RStudio. Este é um ambiente integrado de desenvolvimento (IDE, do inglês *Integrated Development Environment*) específico para a linguagem de programação R. Por se tratar de uma IDE, oferece funcionalidades que facilitam a construção de pacotes de *software* na linguagem R. Seu funcionamento tem características análogas ao Jupyter Hub, disponibilizando um ambiente de trabalho, com quantidade específica de recurso computacional, para cada um dos utilizadores. O acesso e uso do RStudio é feito através de uma interface *web*. Como apresentado na Figura 22, ao acessar o ambiente, o usuário tem disponível funcionalidades para a edição e execução interativa de código, além de opções para a visualização dos resultados gerados.

Figura 22: Exemplo de interface do RStudio.



Na Plataforma BDC, a disponibilização de ambientes RStudio é feita aos colaboradores que trabalham com as atividades de geração de mapas de uso e cobertura da terra utilizando os cubos de dados. Assim, os ambientes são customizados e preparados para esta atividade, fornecendo todas as ferramentas necessárias, o que inclui o pacote `sits` e suas dependências de uso.

Os códigos e customizações realizados para a configuração desses ambientes e integração com a infraestrutura do projeto BDC, estão disponíveis no repositório do GitHub [sits-docker](#) e no [Docker Hub](#).

3.4 Ambiente de desenvolvimento e monitoramento computacional

3.4.1 Github

O GitHub é uma plataforma que permite a hospedagem de código fonte e controle de versão. No projeto BDC, essa plataforma está sendo utilizada para a disponibilização do código fonte das tecnologias de *software Open Source* desenvolvidas. Além disso, de modo a comportar o time de desenvolvimento do projeto, o GitHub é utilizado para o gerenciamento das atividades de desenvolvimento, equipes e responsáveis por cada uma das tecnologias e funcionalidades que estão sendo implementadas.

Os repositórios com os códigos dos *softwares* seguem um *template* definido pelos desenvolvedores do projeto BDC, que padroniza a organização dos códigos. Isto facilita o desenvolvimento e mantém a consistência entre os projetos, potencializando a utilização por desenvolvedores terceiros. Este *template* é disponibilizado no repositório do GitHub [cookiecutter-bdc](#).

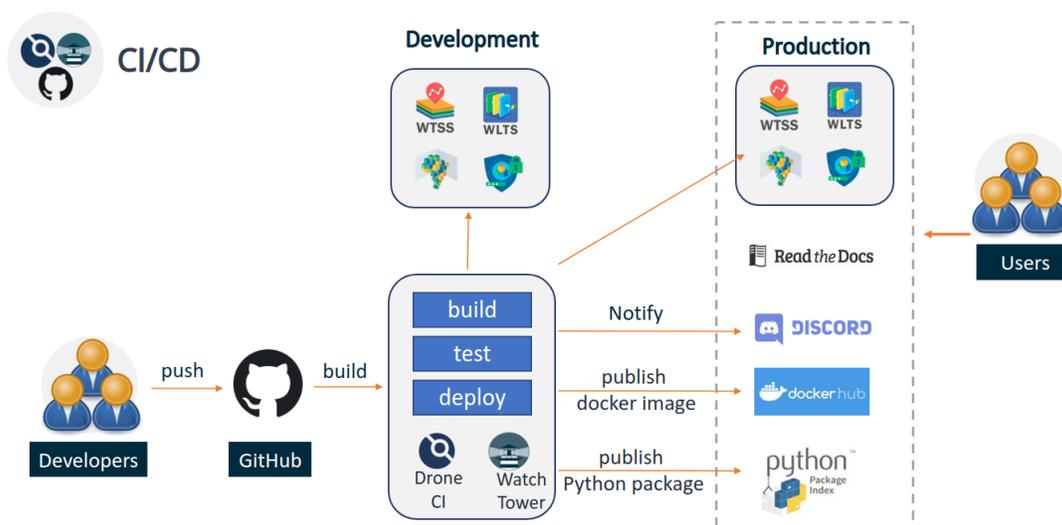
Além da disponibilização gratuita desses códigos e do controle de versão, ao fazer o uso do GitHub, o projeto BDC estabelece uma interface de comunicação direta entre os utilizadores do *software* e os desenvolvedores. Isto já que, através do GitHub, a requisição de funcionalidades e o aviso sobre erros encontrados podem ser informados através de *issues*.

3.4.2 Drone CI

Drone CI é um serviço de integração e entrega contínua para automatização dos *build*, testes e também o *deploy* dos repositórios de códigos fonte no GitHub. Por meio desta ferramenta é possível atingir uma maior confiança e segurança nas aplicações desenvolvidas pelo BDC a fim de garantir que tecnologias fiquem mais estáveis. Através dos pedidos de integração de código (*Pull Request*) nos repositórios dos GitHub é disparado uma pipeline de execução no Drone⁹ hospedado no servidor BDC, conforme a Figura 23. Ao final da execução, o Drone faz a notificação para o GitHub com o status final da execução, informando se passou ou se teve erros. A pipeline também faz a notificação pelo Discord para os desenvolvedores e envolvidos acompanharem. Por fim, o drone faz o *deploy* da aplicação no servidor de produção BDC para que esta fique disponível para acesso aos usuários.

⁹<https://drone.dpi.inpe.br>

Figura 23: Rotina de integração e entrega contínua pelo Drone



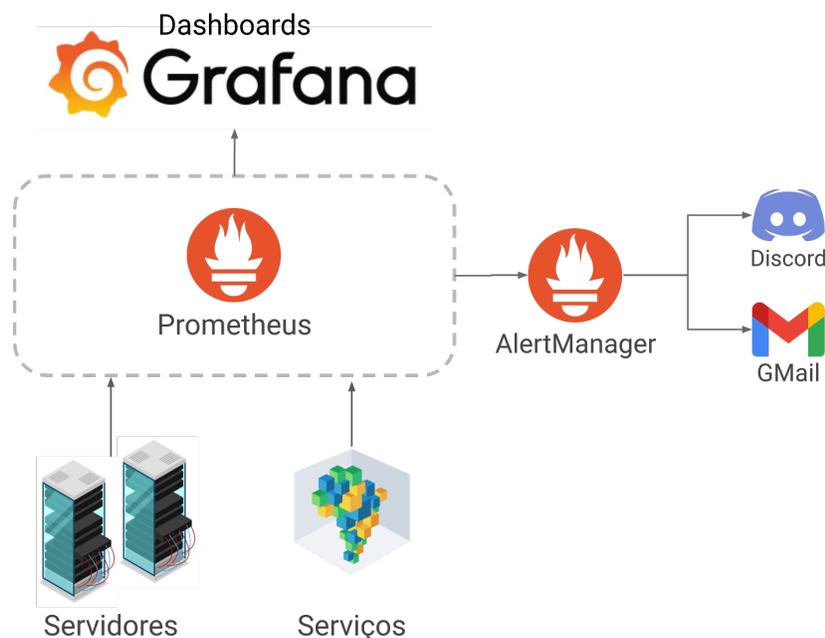
3.4.3 BDC Monitor

O controle e gerenciamento dos ativos de *software* e *hardware* de uma infraestrutura computacional são atividades essenciais para a provisão ininterrupta de seus serviços. Conseqüentemente, serve também para evitar que as operações que estão sendo realizadas com base nesses serviços sejam afetadas. Dentre as atividades para o auxílio a esse controle e gerenciamento está o monitoramento da disponibilidade e uso dos ativos. Na plataforma BDC, para a realização dessa atividade, criou-se a ferramenta BDC Monitor, que fornece páginas *web* de alto nível com informações gerais e específicas do *software* e *hardware* da plataforma. Também é feita a geração e envio de alertas que informam através de serviços de mensagem as mudanças e incidentes ocorridos. O desenvolvimento do BDC Monitor foi feito com base nas

tecnologias Prometheus¹⁰, AlertManager¹¹ e Grafana¹².

Na Figura 24, é apresentado o fluxo de funcionamento do BDC Monitor. Como pode ser observado, o Prometheus, ferramenta que fornece recursos para a monitoração dos ativos, é o responsável em recuperar e armazenar as informações dos serviços e servidores da Plataforma BDC.

Figura 24: Visão geral do BDC Monitor



Para a visualização das informações coletadas, o BDC Monitor disponibiliza páginas criadas com a ferramenta Grafana, que possuem funcionalidades de alto nível, como gráficos interativos e consultas dinâmicas aos dados. Na versão atual do BDC Monitor, estão disponíveis páginas para a visualização de uso dos recursos de

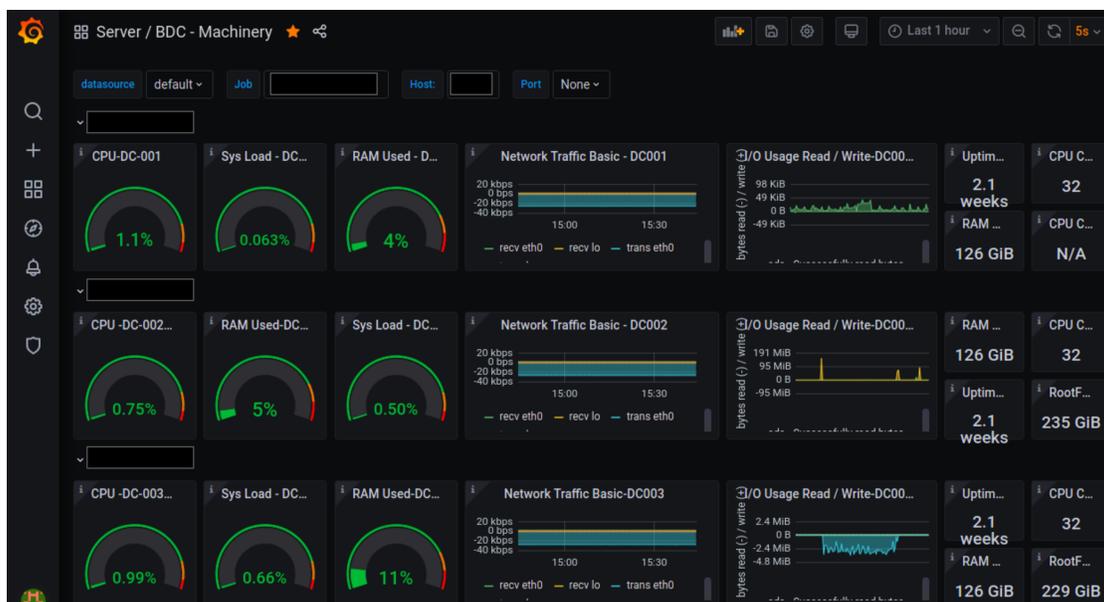
¹⁰Prometheus <https://prometheus.io/>

¹¹AlertManager <https://prometheus.io/docs/alerting/latest/alertmanager/>

¹²Grafana <https://grafana.com/>

hardware, disponibilidade de serviços e funcionamento da rede. Além dessas, páginas com a visão geral dos produtos de dados e espaço em disco disponível, também foram criadas. Para exemplificar uma dessas páginas, a Figura 25 faz a apresentação da página de visão geral dos servidores disponíveis na infraestrutura do projeto BDC. Neste exemplo, tem-se informações do uso de CPU, carga no sistema, uso de memória, tráfego de rede e I/O dos discos.

Figura 25: Exemplo de página Grafana com informações dos servidores do projeto BDC

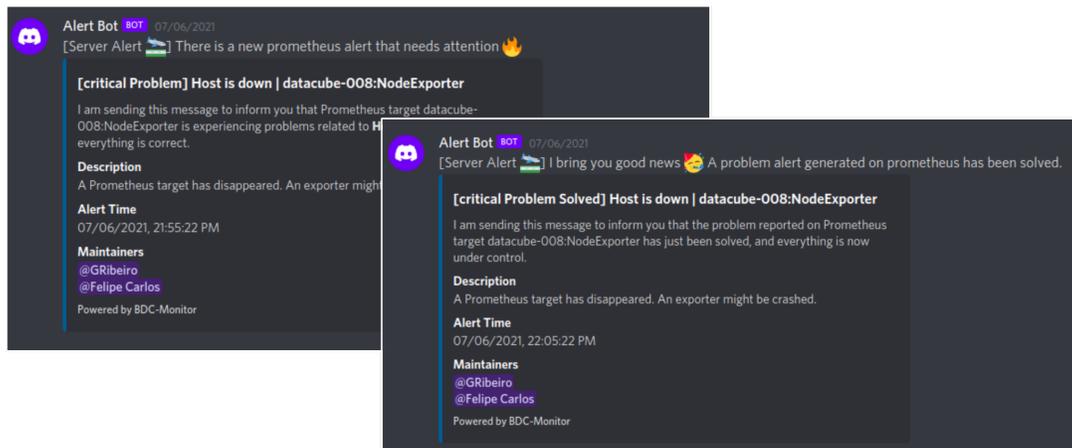


Além disso, com base nas informações coletadas, alertas de incidentes e mudanças são gerados pelo Prometheus. Quando gerados, eles são comunicados aos respectivos responsáveis, através dos serviços de Email e Discord¹³. Para que esses serviços pudessem ser utilizados na comunicação desses alertas, no BDC Monitor fez-se o uso do AlertManager, uma extensão do Prometheus que permite o envio dos alertas a

¹³<https://discord.com/>

diferentes serviços. Na Figura 26, é apresentado um exemplo de alerta enviado pelo BDC Monitor ao serviço Discord. Pode-se notar que, tanto o incidente, quanto a sua solução são reportados aos responsáveis, facilitando o acompanhamento.

Figura 26: Exemplos de mensagens geradas pelo BDC Monitor



Os códigos utilizados para a composição do BDC Monitor, estão disponíveis no repositório do GitHub [bdc-monitor](#).

4 Classificações

O Brasil é um dos maiores produtores mundiais de alimentos e a sua crescente demanda produz mudanças constantes no uso e cobertura da terra no país. Ao mesmo tempo, o país precisa reduzir suas taxas de desmatamento e emissões de gases de efeito estufa (GEE) no contexto da atual emergência climática [23]. Para atender este desafio, o país precisa balancear ganhos econômicos com práticas agrícolas sustentáveis mediante políticas eficientes de comando e controle orientadas a conter o desmatamento [24].

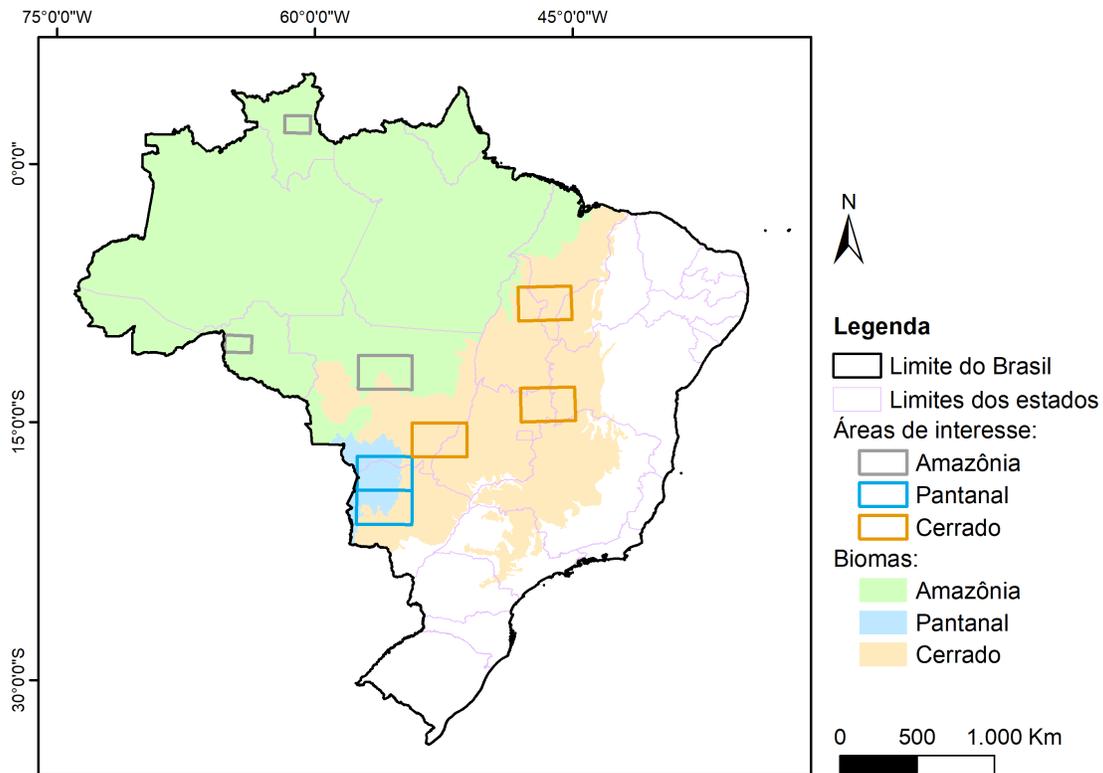
O Sensoriamento Remoto tem se mostrado uma ferramenta eficiente para monitorar as mudanças no uso e cobertura da terra [25]. No entanto, o seu uso para identificar mudanças de forma sistemática ainda é limitado, em grande parte por força dos atuais métodos e plataformas de análise de dados disponíveis.

Uma alternativa para aprimorar as capacidades das análises de dados de sensoriamento remoto é o uso de cubos de dados de imagens, que consistem em matrizes multidimensionais de dados multi-espectrais [1]. O projeto BDC propõe o uso de cubos de dados obtidos a partir de sensores embarcados nos satélites Aqua, Terra, Landsat 8, Sentinel-2 e CBERS 4 para melhorar significativamente a extração de informação sobre as dinâmicas de uso da terra no Brasil, usando métodos de inteligência artificial, aprendizado de máquina e análise de séries temporais.

Complementando o relatório do ano anterior, este relatório apresenta os avanços obtidos na área de aplicação dos cubos de dados desenvolvidos pelo projeto BDC. Os resultados deste período incluem classificações usando cubos de dados Landsat-8 e Sentinel-2 nos biomas Amazônia (Capítulos 4.4 e 4.5), Cerrado (Capítulo 4.3) e Pantanal (Capítulo 4.6). A Figura 27 apresenta as áreas das classificações do projeto nos biomas mencionados anteriormente.

Os resultados da classificação ficam disponíveis no [portal](#) do projeto Brazil Data Cube, na aba de classificações.

Figura 27: Áreas classificadas usando cubos de dados de imagens gerados pelo projeto BDC.



4.1 Dados e software

4.1.1 Dados de Observação da Terra

Para realizar as classificações, foram utilizadas cubos de dados de imagens dos satélites Landsat-8 e Sentinel-2 disponíveis com os nomes LC8_30_16D_STK-1 e S2-SEN2COR_10_16D_STK-1, respectivamente. Desde Fevereiro de 2013, o Landsat-8 dá continuidade ao registro científico de Observação da Terra do Projeto Landsat por meio do sensor *Operational Land Imager* (OLI). Ele tem como objetivo coletar

e armazenar dados globais de média resolução multi-espectral (ópticos e termais) e fabricar e distribuir produtos (imagens) padronizados sem custo para os usuários [26] (ver Tabela 7).

Tabela 7: Bandas espectrais do sensor OLI do satélite Landsat-8 [27].

Banda	Descrição	Resolução (m)	Comprimento de onda (μm)
1	Coastal aerosol	30	0.43–0.45
2	Blue	30	0.45–0.51
3	Green	30	0.53–0.59
4	Red	30	0.64–0.67
5	Near infrared	30	0.85–0.88
6	Shortwave infrared	30	1.57–1.65
7	Shortwave infrared	30	2.11–2.29
8	Panchromatic	15	0.50–0.68
9	Cirrus	30	1.36–1.38
10	Thermal Infrared	100	10.60–11.19
11	Thermal Infrared	100	11.50–12.51

Operando desde junho de 2015, a missão Sentinel-2 é composta por dois satélites na mesma órbita e diametralmente opostos (Sentinel-2A). O objetivo da missão é monitorar a variabilidade das condições da superfície da Terra com um período de revisita de 10 dias (ou 5, considerando os 2 satélites juntos). O sensor a bordo é o *Multi-Spectral Instrument* (MSI) que conta com 13 bandas espectrais (ver Tabela 8).

4.1.2 Software

As classificações de uso e cobertura da Terra feitas no BDC usam principalmente dois pacotes do R [29]: o `sits` (tag v0.12.0) para fazer as classificações e o `rstac` para fazer intercâmbio de dados pela Internet.

O pacote `sits` ¹⁴ possui um conjunto de ferramentas de análise, visualização

¹⁴Pacote `sits` <https://github.com/e-sensing/sits/>

Tabela 8: Resoluções do produto L2A feito com bandas espectrais do sensor MSI a bordo do satélite Sentinel-2 [28].

Banda	Descrição	Resolução (m)	Comprimento de onda (nm)
1	Ultra blue	60	443
2	Blue	60, 20, 10	490
3	Green	60, 20, 10	560
4	Red	60, 20, 10	665
B5	Visible and near infrared	60, 20	705
6	Visible and near infrared	60, 20	705
7	Visible and near infrared	60, 20	783
8	Visible and near infrared	10	842
8A	Visible and near infrared	60, 20	865
9	Short wave infrared	60, 20	945
11	Short wave infrared	60, 20	1610
12	Short wave infrared	60, 20	2190

e classificação de séries temporais de imagens de Observação da Terra. Ele permite acessar cubos de dados usando dados tanto locais como remotos (internet), extrair séries temporais de pontos amostrais, treinar modelos de aprendizagem de máquinas, classificar cubos de dados e realizar o pós-processamento dos resultados de classificação (ver secção 3.3.2).

Como complemento ao `sits`, o pacote `rstac`¹⁵ fornece funções para pesquisar, acessar e baixar dados espaço-temporais de Observação da Terra usando a especificação *SpatioTemporal Asset Catalog (STAC)*. STAC é um linguagem padrão para descrever informações geoespaciais, permitindo uma fácil indexação e acesso¹⁶ (ver Seção 3.2).

¹⁵Pacote `rstac` <https://cran.r-project.org/package=rstac>

¹⁶STAC <http://stacspec.org>

4.1.3 Coleta de amostras

Para coletar amostras de treinamento, a equipe do projeto BDC usou a ferramenta *Collect Earth* e os serviços WLTS. O *Collect Earth* ¹⁷ é uma ferramenta integrada ao *Google Earth* usada para realizar a coleta de amostras. Ela provê acesso a imagens multi-temporais de alta resolução através do *Google Maps*, *Bing Maps* e coleções de imagem do *Google Earth Engine*. Uma das funcionalidades da ferramenta é a integração com os bancos de dados *SQLite* e *PostgreSQL*.

O *Collect Earth* roda em sistemas operacionais Windows, Mac e Linux e faz parte do *OpenForis*, um conjunto de ferramentas de código aberto para monitoramento do meio ambiente. O seu catálogo conta com ferramentas para: coleta de amostras, gerenciamento de dados espaciais, análise colaborativa, disseminação, processamento, entre outras funções. São as seguintes: (a) Collect, (b) Collect Mobile, (c) Calc, (d) Collect Earth, (e) Collect Earth Online, (f) o SEPAL e (g) Earth Map. Outra ferramenta importante no processo de coleta de amostras é o *QGIS* ¹⁸, um software de sistema de informação geográfica (SIG) livre e de código aberto. Ele permite visualizar, editar, e analisar dados geoespaciais.

O serviço WLTS também foi utilizado na seleção de amostras de treinamento para dar suporte aos algoritmos de aprendizado de máquina supervisionados usados na classificação (ver secção 3.2.3). Foram utilizados os clientes *wlts.py* e *rwlts*, desenvolvidos respectivamente na linguagem *Python* e linguagem *R*, para a construção de aplicações e uso em ambientes de computação interativa como *IPython/Jupyter*.

¹⁷Collect Earth <http://www.openforis.org/tools/collect-earth>

¹⁸QGIS <https://www.qgis.org>

4.1.4 Scripts das classificações

O código dos scripts usados durante as classificações estão no repositório do projeto BDC chamado *BDC Classification*¹⁹. No repositório, cada classificação tem uma pasta denominada *scripts* com o código e outra chamada *samples* com as amostras usadas no treinamento das classificações.

4.2 Metodologia

4.2.1 Amostragem

No projeto BDC, existem amostras que foram coletadas: (1) *in situ*, ou seja, coleta feita no campo por especialistas; (2) através de interpretação visual de imagens de sensoriamento remoto; e (3) coleta baseado na concordância entre mapas de base feitos por terceiros.

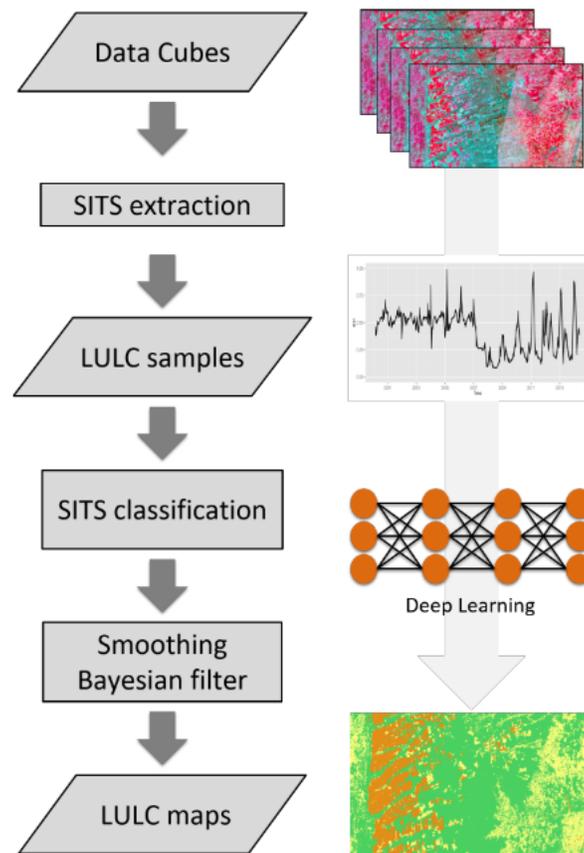
A coleta *in situ* aproveita o trabalho de campo feito por especialistas. Estas coletas acontecem com regularidade e são feitas pelos cientistas (ver [30]) ou organizações do governo como INPE ou EMBRAPA. As outras coletas tiram proveito de fontes de dados pré-existentes, como imagens de observação da Terra ou dados de qualidade conhecida, como os mapas dos sistemas PRODES e DETER [31], e mais recentemente os dados Mapbiomas [32]. Estas coletas podem ser realizadas de forma manual, usando o software QGIS, ou usando o serviço WLTS (ver Secção 4.1.3).

¹⁹<https://github.com/brazil-data-cube/bdc-classifications>

4.2.2 Classificação de Uso e Cobertura da Terra

O projeto BDC usa cubos de dados de imagens para gerar mapas anuais de uso e cobertura da Terra para as áreas de interesse nos diferentes biomas brasileiros. A metodologia utiliza técnicas de aprendizagem de máquina para classificar séries temporais de imagens de satélite extraídas do cubos de dados, como mostrado na Figura 28.

Figura 28: Fluxograma do método utilizado para classificar uso e cobertura da Terra a partir de cubos de dados de imagens. Fonte: [1]



As entradas do método são os cubos de dados (ver secção 4.1.1) e amostras de

treinamento. As amostras de treinamento consistem na informação espaçotemporal de uma classe de uso e cobertura da Terra e tipicamente incluem o nome da cobertura ou uso, a longitude, a latitude (em sistema de coordenadas *World Geodetic System* - WGS84), e as datas de começo e fim da respectiva classe.

O próximo passo é recuperar as séries de tempo para cada ponto das amostras usando um cubo de dados. Este processo é chamado *SITS extraction* e é feito usando o pacote `sits`. No caso de um cubo *on-line*, o pacote `sits` faz uso do pacote `rstac` (ver Seção 4.1.2). O resultado deste processo são séries temporais com um uso ou cobertura da Terra conhecido.

Estas séries temporais são usadas posteriormente para treinar um modelo de aprendizagem de máquina. Os algoritmos de aprendizagem de máquina mais comumente usados nas classificações do BDC são *Random Forest* e *Deep Learning*; ambos disponíveis no pacote `sits`.

Depois da etapa de treinamento, realiza-se a etapa de classificação, onde o modelo treinado anteriormente é aplicado no cubo de dados. O resultado é um mapa de uso e cobertura da terra que potencialmente pode ser pós-processado para melhorar a sua qualidade.

O pós-processamento mais usado pelo BDC é o *Bayesian Smoother*. Ele consiste em usar as probabilidades de cada classe da vizinhança para atualizar as probabilidades dos pixels da classificação usando o teorema de Bayes. O resultado é uma redução do efeito sal e pimenta.

Finalmente, avalia-se a acurácia dos mapas de uso e cobertura da Terra. Entre os vários métodos disponíveis para tal, o projeto BDC aplica o uso das boas práticas

da *Food and Agriculture Organization of the United Nations* (FAO) [33, 34].

4.3 Bioma Cerrado

4.3.1 Área classificada

O Cerrado é o segundo maior bioma da América do Sul, ocupando uma área de aproximadamente 2 milhões de km², cerca de 22% do território brasileiro. Considerado um *hotspot* global de biodiversidade, este bioma apresenta extrema abundância de espécies endêmicas, mas sofre excepcional perda de habitat. Por sua diversidade biológica, o Cerrado é reconhecido como a savana mais rica do mundo, abrigando pelo menos 11600 espécies de plantas nativas catalogadas [35].

4.3.2 Classes e amostras

O conjunto de dados de treinamento foi constituído por três subconjuntos de dados coletados tanto no escritório como em campo (ver Seção 4.1.3).

No primeiro subconjunto de amostras, foram coletados 5307 pontos para todo o Cerrado através de interpretação visual de imagens, selecionados através de uma grade de 20 quilômetros usando a ferramenta *Collect Earth*. No segundo subconjunto de amostras, 85026 pontos foram coletados com uma grade de 5 quilômetros utilizando mapas de uso e cobertura da Terra, produzidos pelos projetos MapBiomas, Pastagem.org (Lapig/UFG) e pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para 2018. No terceiro subconjunto de amostras, foram usados os dados de campo da equipe do projeto TerraClass Cerrado, coletados em 2018, e da equipe do Pastagem.org (Lapig/UFG), coletados entre 2017 e 2018.

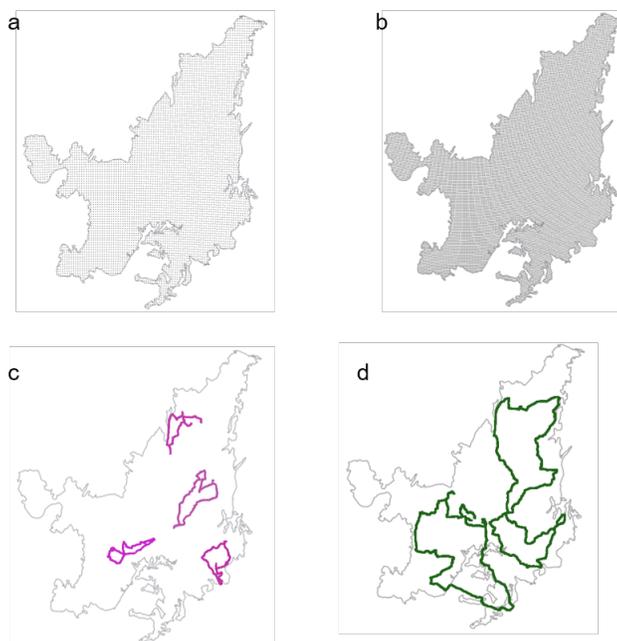
Para filtrar essas amostras, primeiramente foi feita uma seleção de quais classes de uso e cobertura da Terra seriam utilizadas na classificação. Posteriormente, fez-se uma análise de concordância em relação ao uso e cobertura Terra quando para um mesmo ponto amostral existia mais de uma informação. Por exemplo, ao coletarmos 85026 pontos dos mapas produzidos pelo MapBiomias, Pastagem.org e IBGE (segundo subconjunto), foi feita uma validação cruzada para verificar quando esse ponto entraria ou não para o conjunto de dados de treinamento. Então, se para um mesmo ponto as três fontes de dados concordavam, o ponto entrava para a coleção de dados de treinamento. Quando só havia duas fontes de dados (isso acontecia para áreas de não pastagem), a classe de uso e cobertura da terra do MapBiomias deveria concordar com a classe do IBGE. Caso não concordassem, essas amostras eram descartadas.

O grid das amostras coletadas utilizando a Ferramenta Collect Earth, os mapas de uso e cobertura da terra do MapBiomias, Pastagem.org e IBGE e os dados de campo podem ser observado na Figura 29. As amostras usadas para a classificação estão disponíveis no arquivo RDS *samples.rds* na pasta *bdc_cerrado/data/samples* do repositório *BDC Classifications* (ver Seção 4.1.4).

4.3.3 Cubos de dados

A classificação foi feita para o cubo de dados Landsat-8 (nome: LC8_30_16D_STK-1), composição “stack” 16 dias, para três tiles do projeto BDC: 042050, 044046 e 044049, utilizado o pacote `sits` em R. O Algoritmo utilizado foi o Random Forest com 2000 árvores. E as bandas do Landsat 8 utilizadas foram: “BAND1”, “BAND2”,

Figura 29: Conjunto amostral do Cerrado: (a) grid de 20 km usado para coleta de amostras utilizando Collect Earth; (b) grid de 5 km usado para coleta de amostras utilizando mapas de uso e cobertura da terra do MapBiomas, Pastagem.org e IBGE (usando a ferramenta WLTS); (c) amostras de campo coletadas pela equipe do projeto TerraClass Cerrado em 2018; (d) amostras de campo coletadas pela equipe do Pastagem.org em 2017 e 2018.



"BAND3", "BAND4", "BAND5", "BAND6", "BAND7", "EVI" e "NDVI".

4.3.4 Resultados

A classificação dos três tiles Landsat para o ano safra 2017/2018 é apresentada na Figura 30. Os resultados da classificação estão disponíveis no portal do projeto Brazil Data Cube: https://brazildatacube.dpi.inpe.br/portal/explore?collection=LCC_L8_30_1M_STK_Cerrado-1

4.3.5 Acurácia

Ao compararmos a classificação dos três tiles com os dados de pasto produzidos pelos Pastagem.org, obtivemos acurácias globais iguais a 0,86, 0,89 e 0,86 para os tiles 044046, 044049 e 042050 respectivamente (Tabela 9).

Tabela 9: Tabela 4: Acurácia do produtor (PA), usuário (UA) e global (OA) das classificações do Cerrado do cubo de dados Landsat-8 (LC8_30_16D_STK-1) (tiles 44046, 44049, e 42050) quando comparado com os dados do Pastagem.org.

	BDC 44046			BDC 44049			BDC 42050		
	PA	UA	OA	PA	UA	OA	PA	UA	OA
Non Pasture	0,87	0,98	0,86	0,91	0,95	0,89	0,93	0,84	0,86
Pasture	0,717	0,27		0,76	0,65		0,78	0,89	

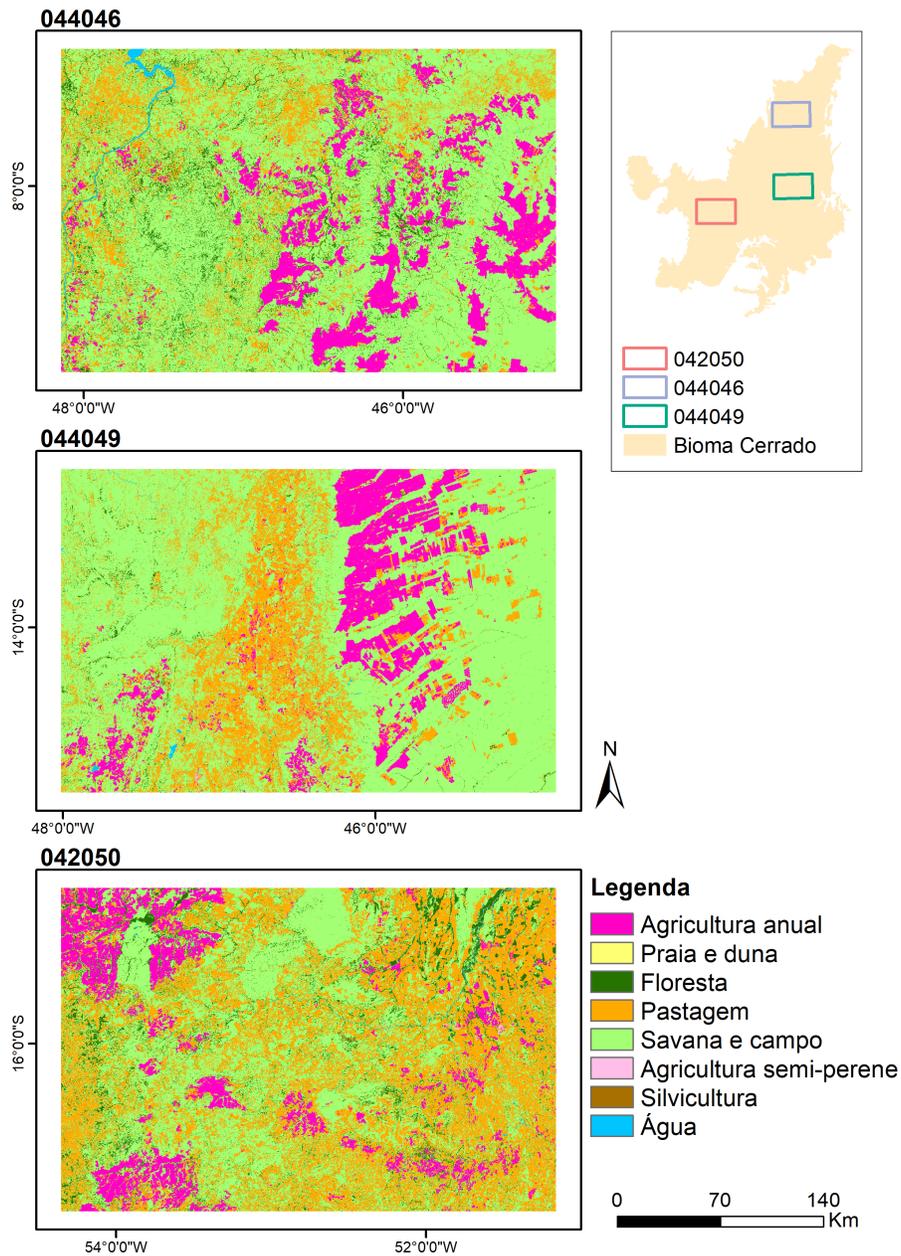
Quando comparamos as classificações com os dados do PRODES Cerrado, obtemos as seguintes acurácias globais: 0,85, 0,87 e 0,85 para os tiles 044046, 044049 e 042050 respectivamente (Tabela 10).

Tabela 10: Acurácia do produtor (PA), usuário (UA) e global (OA) das classificações do cubo de dados Landsat-8 (LC8_30_16D_STK-1) quando comparado com os dados do PRODES Cerrado.

	BDC 44046			BDC 44049			BDC 42050		
	PA	UA	OA	PA	UA	OA	PA	UA	OA
Anthropic	0,77	0,72	0,85	0,78	0,83	0,87	0,79	0,94	0,85
Water	0,72	0,8		0,44	0,68		0,46	0,77	
Natural	0,88	0,91		0,92	0,89		0,94	0,76	

As estimações de acurácia foram feitas usando o método *Bootstrapping*.

Figura 30: Classificação dos tiles 042050, 044049 e 044046 do cubo de dados Landsat-8 (LC8_30_16D_STK-1) utilizando o algoritmo Random Forest, de setembro de 2017 a agosto de 2018.



4.4 Bioma Amazônia - desmatamento

4.4.1 Área classificada

As florestas tropicais têm um papel importante no sistema terrestre regulando os ciclos da água e do carbono e também protegendo a biodiversidade. Existe uma tendência global de mudança de uso e cobertura da Terra onde áreas de floresta natural são substituídas por outros usos, como a agricultura [36]. O desmatamento contribui nas emissões de carbono e perda de biodiversidade, portanto mapas precisos de desmatamento, degradação florestal e uso e cobertura da terra são fundamentais para fundamentar a tomada de decisões [37].

A floresta Amazônica é a maior floresta tropical do mundo. Este bioma tem importante papel como reserva de biomassa e refúgio da biodiversidade [38]. Desde os anos 1990s, mais de 540 milhões de hectares de floresta natural foram desmatadas na Amazônia Brasileira de acordo com os dados do PRODES [39].

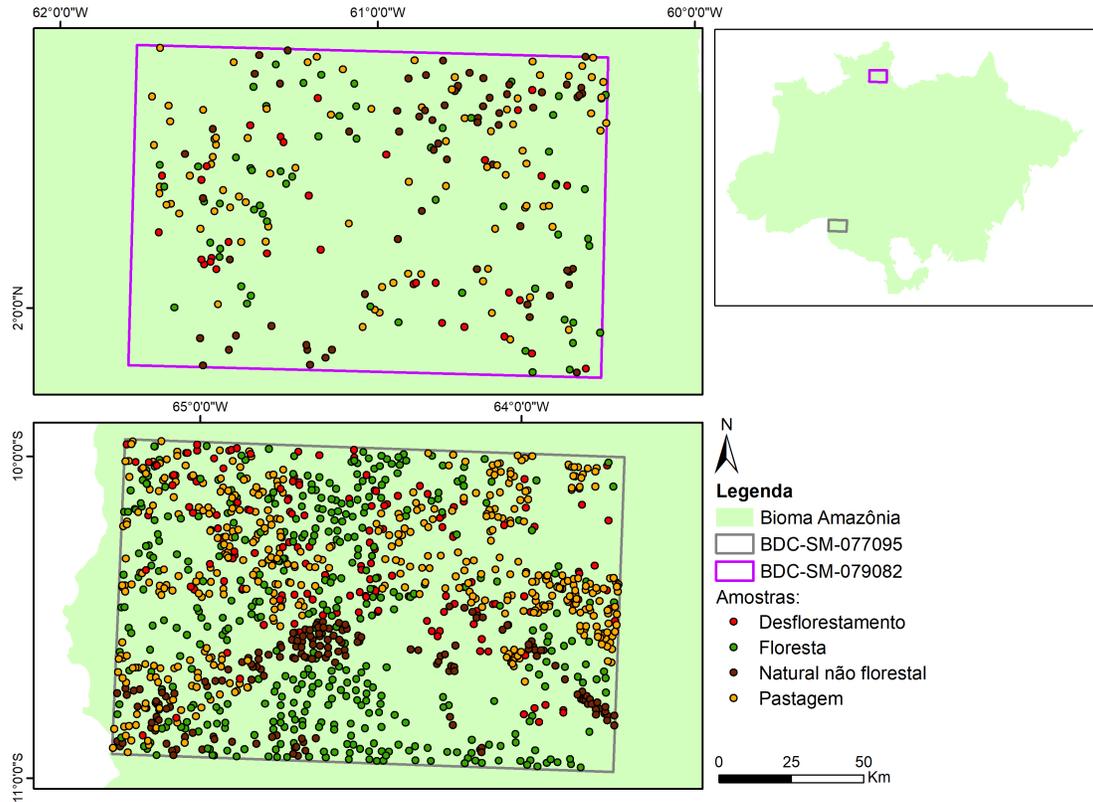
Para atestar as capacidades dos seus cubos de dados e ferramentas de classificações, o projeto BDC fez classificações voltadas à identificação de áreas desmatadas nos estados de Rondônia e Roraima, nos cubos de dados identificados como *BDC-SM-077095* e *BDC-SM-079082* (ver Figura 27).

4.4.2 Classes e amostras

Foram coletados pontos interpretados visualmente com base em imagens Sentinel-2 do ano agrícola 2019 (Agosto 2018 - Julho 2019) para serem usados como amostras na classificação (Figura 31). Esse conjunto de pontos foi interpretado pelo time de

classificação do projeto BDC (ver Secção 4.2.1). A Tabela 11 apresenta as classes e respectivas quantidades de amostras coletadas em cada tile.

Figura 31: Amostras usadas para o treinamento na classificação da Amazônia com foco em desflorestamento.



As amostras usadas durante a classificação estão disponíveis no arquivo RDS *samples_tb.rds* na pasta *bdc_amazonia/data/samples* do repositório *BDC Classifications* (ver secção 4.1.4).

Tabela 11: Número de amostras usadas para treinamento das classificações da Amazônia.

Classe	Tile 077095	Tile 079082
Desflorestamento	157	62
Floresta	414	96
Natural não florestal	171	100
Pastagem	460	142

4.4.3 Cubos de dados

A classificação foi feita usando o cubo de dados do sensor MSI/Sentinel-2 (S2-SEN2COR_10_16D_STK-1), na composição "stack" [1], com 16 dias de resolução temporal e 10m de resolução espacial (ver Secção 2.2). Dois tiles do projeto BDC foram classificados: 077095 e 079082. Utilizou-se o algoritmo Random Forest, com 2000 árvores, e as bandas espectrais 1, 2, 3, 4, 5, 6, 7, 8, 8A, 11, 12, e os índices de vegetação EVI e NDVI.

4.4.4 Resultados

O resultado das classificações dos tiles 077095 e 079082 podem ser vistos nas Figuras 32 e 33, respectivamente. Eles também ficaram disponíveis no portal do projeto Brazil Data Cube: https://brazildatacube.dpi.inpe.br/portal/explore?collection=LCC_S2_10_16D_STK_Amazonia-1

Figura 32: Resultado de classificação da Amazônia focada em desmatamento no tile 077095 (Rondônia).

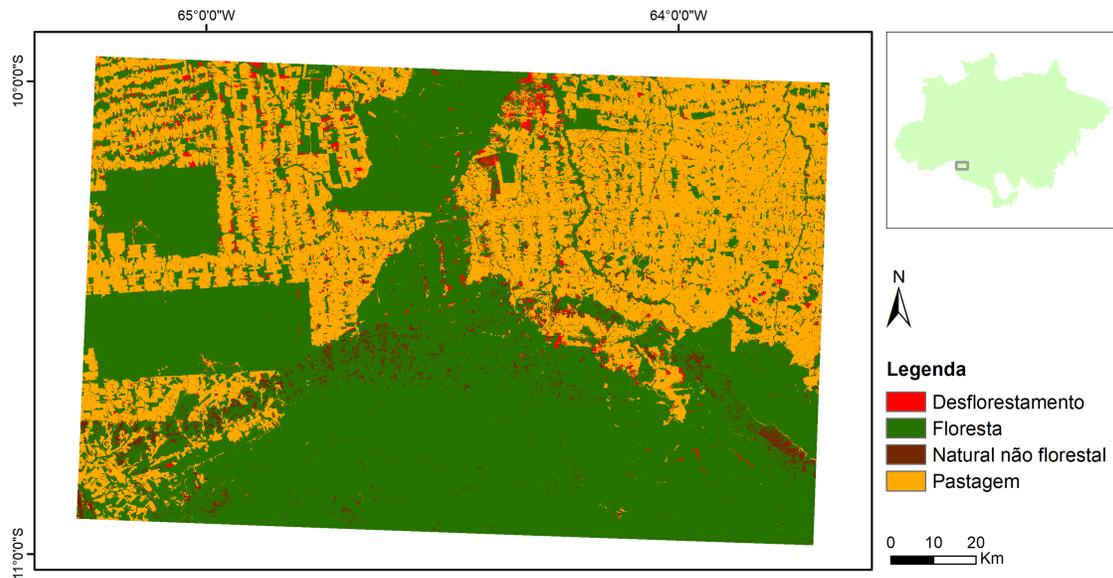
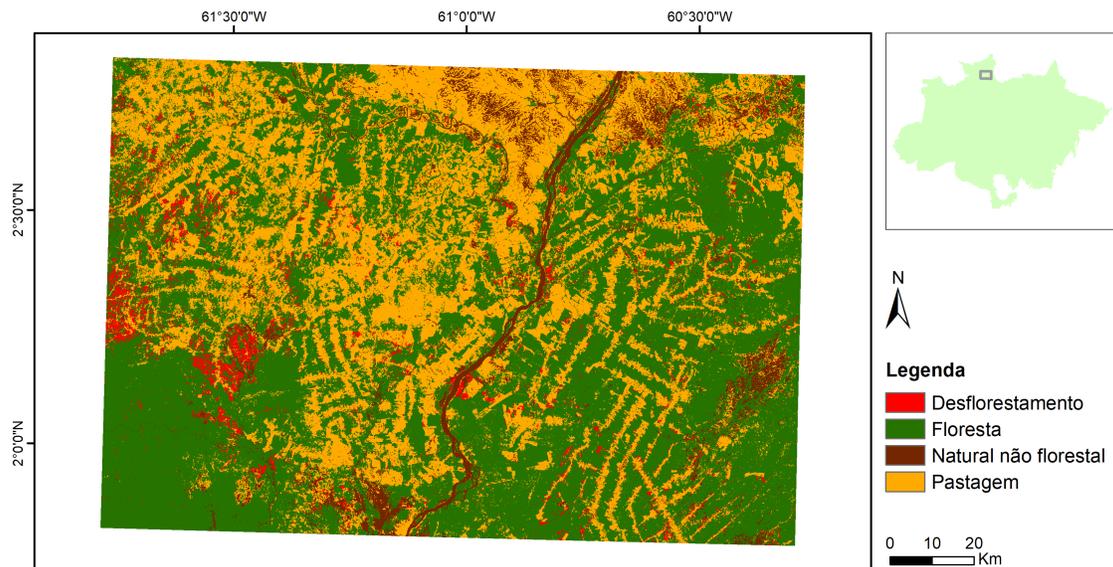


Figura 33: Resultado de classificação da Amazônia focada em desmatamento no tile 079082 (Roraima).



4.4.5 Acurácia

A acurácia da classificação do tile 077095, em Rondônia, é apresentada na Tabela 12; o cálculo da acurácia do tile 079082 (Roraima) ainda esta sendo processada.

Tabela 12: Acurácia do produtor (PA) e do usuário (UA) da classificação do cubo 077095 (Rondônia) usando boas práticas [34]. A acurácia global é 0,9.

		Referência				UA
		Desmat.	Floresta	Não Flor.	Past.	
BDC	Desmatamento	256	3	4	18	0,91
	Floresta	0	358	4	3	0,98
	Não Florestal	3	48	148	30	0,65
	Pastagem	4	3	3	315	0,97
PA		0,97	0,87	0,93	0,86	

4.5 Bioma Amazônia - uso e cobertura da Terra

4.5.1 Área classificada

Além dos mapeamentos voltados para a identificação de áreas desmatadas (Capítulo 4.4), uma classificação adicional foi feita para a caracterização do uso e cobertura da Terra no estado do Mato Grosso, mais especificamente no cubo de dados Landsat-8 (LC8_30_16D_STK-1) para o tile identificado como BDC-MD-041048 (ver Figura 27).

4.5.2 Classes e amostras

Para esta classificação, pontos interpretados visualmente com base em imagens Landsat do período seco e chuvoso de 2018 foram utilizados. Esses pontos foram interpretados e cedidos pelo Laboratório de Processamento de Imagens e Geoprocessamento

da Universidade Federal de Goiás (LAPIG/UFG). Para aumentar a quantidade de amostras disponíveis para treinamento, foram consideradas as amostras localizadas no tile a ser classificado e nos tiles vizinhos: BDC-MD-041047, BDC-MD-041048 e BDC-MD-042048 (Figura 34). As classes consideradas foram aquelas que possuíam mais de 100 amostras, resultando no explicitado na Tabela 13.

Figura 34: Amostras usadas para o treinamento na classificação da Amazônia com foco em uso e cobertura da terra.

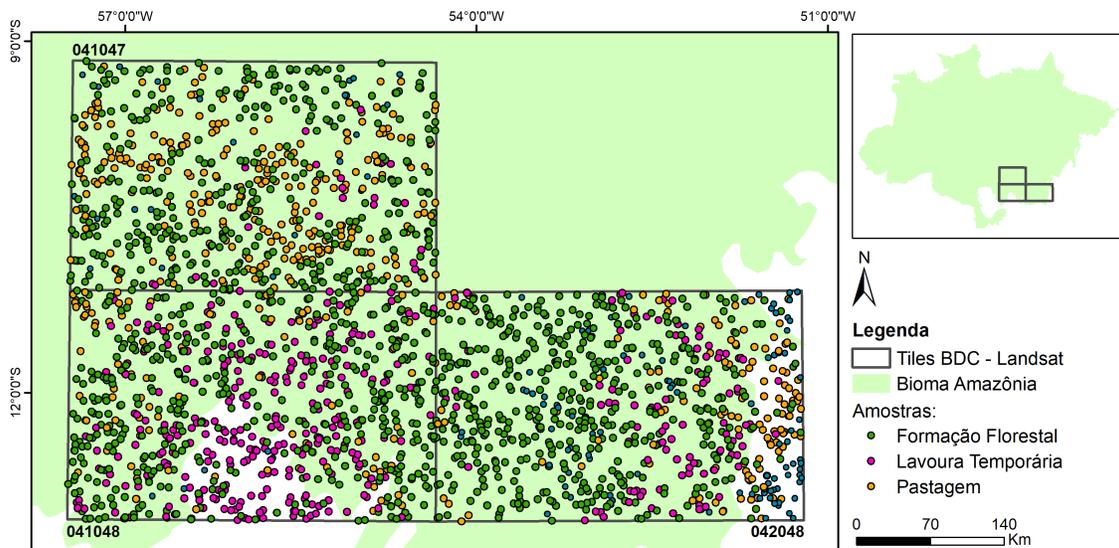


Tabela 13: Classes e quantidade de amostras usadas no treinamento da classificação de uso e cobertura da terra na Amazônia.

Classes	Amostras
Formação Florestal	1284
Lavoura Temporária	405
Pastagem	482
Total	2171

Para representar a classe de Água na classificação, utilizou-se um produto de

abrangência global, o Global Surface Water, produzido por Pekel *et al.* [40]²⁰. Dentre os dados disponíveis nesse produto global, optou-se pela utilização do Maximum Water Extent, que identifica todas as localizações que já foram ocupadas por algum corpo d'água durante o período analisado (1984-2020).

As amostras usadas durante a classificação estão disponíveis no arquivo RDS *samples_amazonia_tc.rds* na pasta *bdc-amazonia_tc/data/samples* do repositório *BDC Classifications* (ver secção 4.1.4).

4.5.3 Cubos de dados

A classificação foi feita usando o cubo de dados do sensor OLI/Landsat 8, na composição "stack" [1], com 16 dias de resolução temporal e 30m de resolução espacial (LC8_30_16D_STK-1). Foi classificado o seguinte tile do projeto: BDC-MD-041048. Utilizou-se o algoritmo Random Forest, com 2000 árvores, e as seguintes bandas espectrais: "BAND1", "BAND2", "BAND3", "BAND4", "BAND5", "BAND6", "BAND7", "EVI" e "NDVI".

4.5.4 Resultados

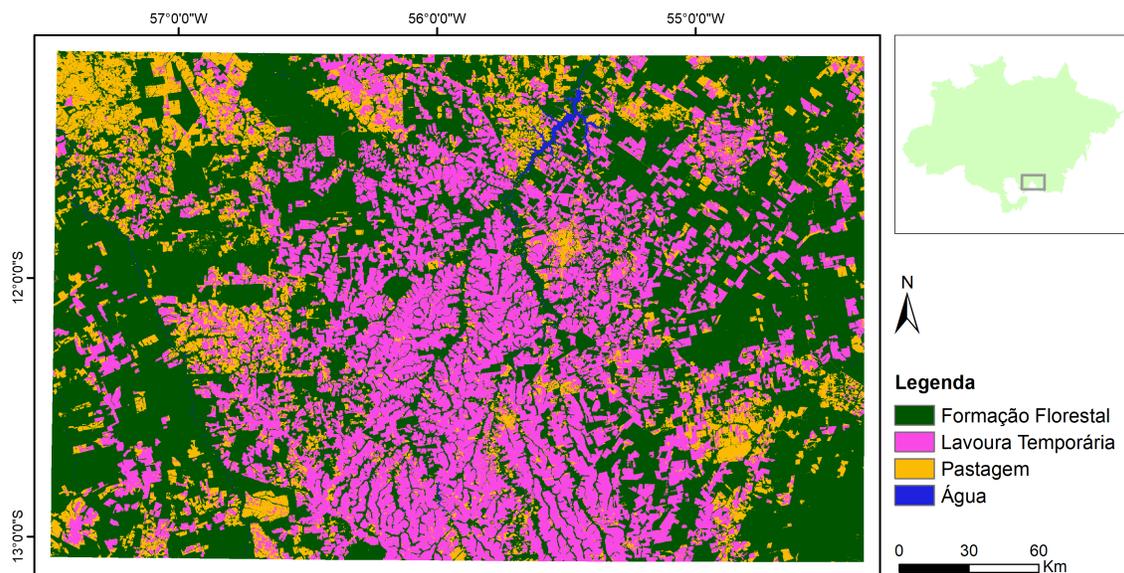
A classificação resultante do tile BDC-MD-041048 para o ano de 2018 com 4 classes analisadas (Formação Florestal, Lavoura Temporária, Pastagem e Água) pode ser vista na figura 35.

Os resultados da classificação estão disponíveis no portal do projeto Brazil Data Cube: https://brazildatacube.dpi.inpe.br/portal/explore?collection=LCC_

²⁰<https://global-surface-water.appspot.com/download>

L8_30_16D_STK_Amazonia-TC-1

Figura 35: Resultado de classificação da Amazônia (tile BDC-MD-041048), focada em uso e cobertura da terra.



4.5.5 Acurácia

Ao comparar a classificação resultante com os dados de desflorestamento produzidos pelo projeto PRODES Amazônia (Tabela 14), obtivemos uma acurácia global de 85.19% para o tile BDC-MD-041048. Parte considerável dessa discordância de menos de 5% ocorreu devido à presença de uma região de Não-Floresta que foi convertida para Lavoura Temporária. O PRODES considera a classe de Não-Floresta como natural e não a altera ao longo dos anos. Na classificação do BDC, como não temos essa restrição, foi possível identificar a conversão dessa região para Lavoura Temporária.

Tabela 14: Acurácia do produtor (PA) e do usuário (UA) da classificação (cubo 077095 Rondônia) de uso e cobertura da terra no bioma Amazônia quando comparada com os dados do PRODES Amazônia. A acurácia global é de 85,19%

		PRODES			UA
		Natural	Antrópico	Total	
BDC	Natural	37744676	5787287	43531963	86,71
	Antrópico	6138273	30846860	36985133	83,40
	Total	43882949	36634147		
	PA	86,01	84,20		

4.6 Bioma Pantanal

4.6.1 Área classificada

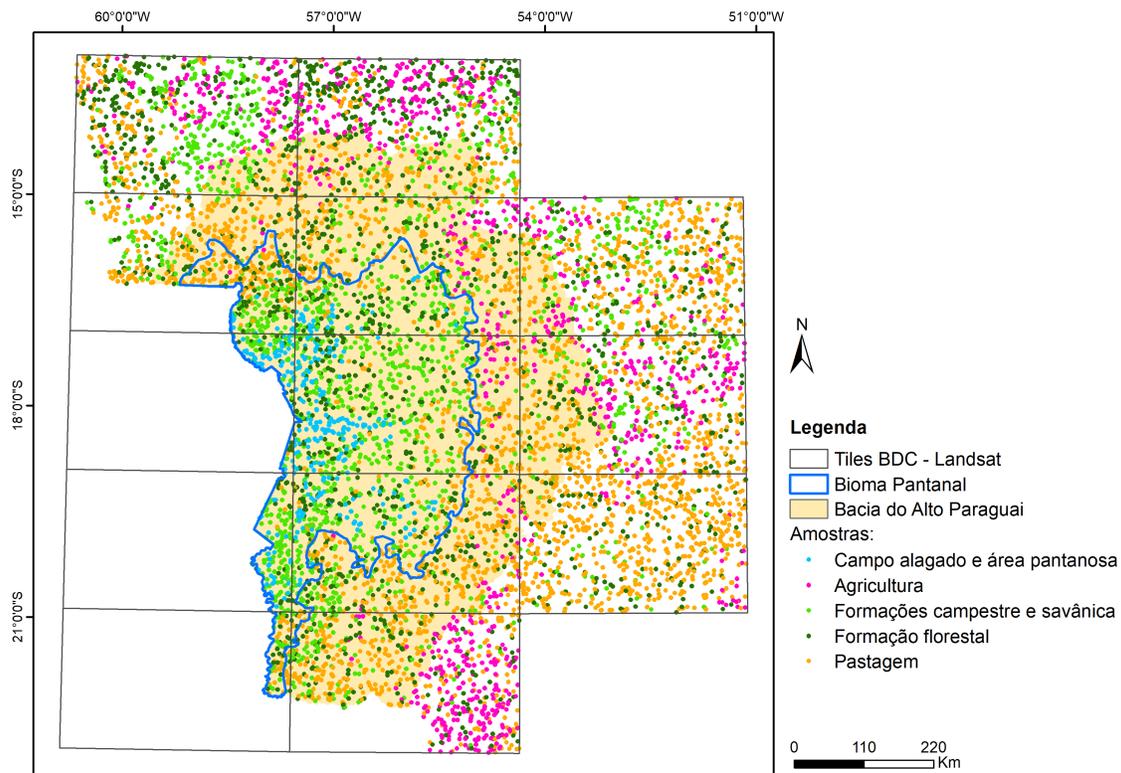
No Brasil, o bioma Pantanal ocorre na região de planície da Bacia do Alto Paraguai (BAP), ocupando parcialmente os estados do Mato Grosso e Mato Grosso do Sul. Além dos corpos d'água permanentes, o Pantanal é composto por uma vasta área de planície periodicamente inundável, onde são formadas lagoas, pântanos e brejos [41].

Considerado como uma das áreas úmidas com maior biodiversidade do planeta, este bioma vem sofrendo intensas modificações na sua paisagem nos últimos anos [42], demonstrando a necessidade do constante monitoramento e análise da dinâmica de uso e cobertura da terra na região. Considerando a legislação vigente e o ritmo de mudança que ocorre atualmente, estima-se que até 2050 este bioma poderá ter cerca de $7960km^2$ de vegetação nativa convertida para usos antrópicos [43].

4.6.2 Classes e amostras

Como amostras para a classificação, foram utilizados pontos interpretados visualmente com base em imagens Landsat do período seco e chuvoso do ano de 2018. Esse conjunto de pontos foi interpretado e cedido pela equipe do LAPIG/UFG. Como a quantidade de amostras coletadas nos dois tiles escolhidos era reduzida, foram consideradas todas as amostras contidas nos tiles que intersectam a Bacia do Alto Paraguai, bacia onde está inserido o bioma Pantanal (Figura 36).

Figura 36: Amostras usadas na classificação do Pantanal.



Do conjunto amostral disponível para essa região, foram usadas as classes que

possuíam mais de 100 amostras coletadas (Tabela 4.6.2). A classe *Formações campestre e savânica* consiste na união das amostras de Formações Campestres (832) e de Formações Savânicas (925). A classe *Agricultura* consiste na união das amostras de Lavoura Temporária (882), Lavoura Perene (5) e Cana-de-Açúcar (92).

Tabela 15: Classes e respectivas quantidades de amostras usadas no treinamento da classificação do Pantanal.

Classes	Amostras
Formação florestal	1731
Formações campestre e savânica	1757
Campo alagado e área pantanosa	347
Pastagem	3094
Agricultura	979
Total	7908

A decisão de agrupar as Formações Campestres e as Formações Savânicas baseou-se na análise das amostras utilizando o algoritmo SOM, de acordo com a metodologia descrita em [13]. As amostras usadas durante a classificação estão disponíveis no arquivo RDS *samples_pantanal.rds* na pasta *bdc_pantanal/data/samples* do repositório *BDC Classifications* (ver secção 4.1.4).

4.6.3 Cubos

A classificação foi feita usando o cubo de dados do sensor OLI/Landsat 8, na composição "stack" [1], com 16 dias de resolução temporal e 30m de resolução espa-

cial (LC8_30_16D_STK-1). Dois tiles do projeto BDC foram classificados: 041051 e 041052. Utilizou-se o algoritmo Random Forest, com 2000 árvores, e as seguintes bandas espectrais: "BAND1", "BAND2", "BAND3", "BAND4", "BAND5", "BAND6", "BAND7", "EVI" e "NDVI".

4.6.4 Pós-classificação

Para representar a classe de água na classificação, também se utilizou o Maximum Water Extent, disponível no Global Surface Water [40]. A alta similaridade entre os padrões espectrais de formações campestres (incluídas na classe Formações campestre e savânica) e pastagem resulta em uma significativa confusão entre essas classes. Como a primeira é de origem natural e a segunda é uma classe antrópica, optou-se pela inclusão do mapeamento do PRODES como máscara na classificação. Nesse sentido, foram consideradas as regras presentes na tabela 16.

As áreas classificadas como Agricultura ou Pastagem pelo BDC e consideradas naturais pelo PRODES foram reclassificadas como Formações campestre e savânica. As áreas classificadas como naturais (Campo alagado e área pantanosa, Formações campestre e savânica ou Formação florestal) pelo BDC e consideradas antrópicas pelo PRODES foram reclassificadas como Pastagem.

■	Reclassificadas como Formação campestre e savânica
■	Reclassificadas como Pastagem
■	Não se altera

Tabela 16: Regras adotadas para a inclusão da máscara do PRODES.

		PRODES	
		Natural	Antrópico
BDC	Agricultura	Amarelo	Ciano
	Campo alagado e área pantanosa	Ciano	Vermelho
	Formações campestre e savânica	Ciano	Vermelho
	Formação florestal	Ciano	Vermelho
	Pastagem	Amarelo	Ciano
	Água	Ciano	Ciano

4.6.5 Resultados

O resultado da classificação dos tiles BDC 041051 e 041052 para o ano de 2018, após a aplicação das máscaras de água e do PRODES, pode ser visto na Figura 37.

Os resultados da classificação estão disponíveis no portal do projeto Brazil Data Cube: https://brazildatacube.dpi.inpe.br/portal/explore?collection=LCC_L8_30_16D_STK_Pantanal-1

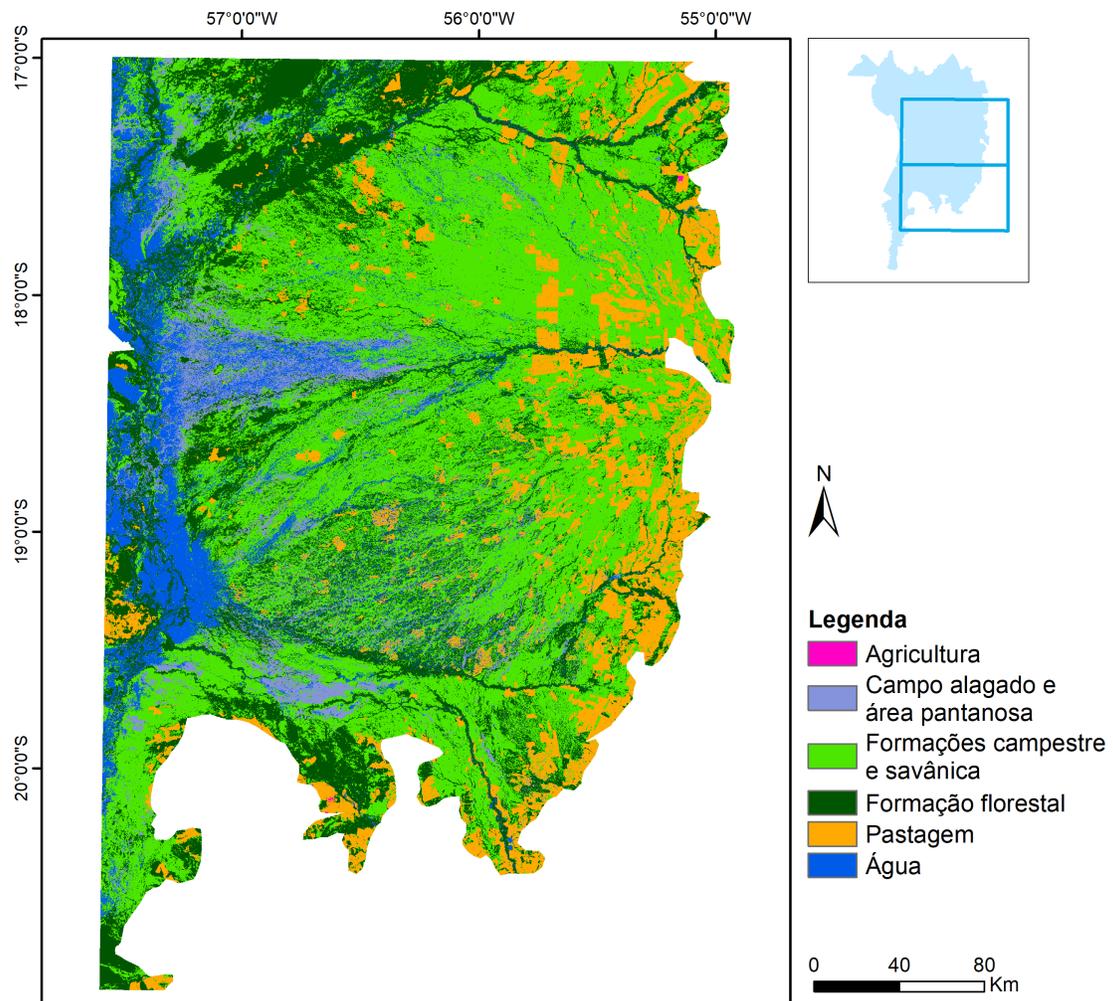
4.6.6 Acurácia

As acurácias da classificação dos cubos 041051 e 041052 do bioma Pantanal são apresentadas na Tabela 17. As acurácias foram computadas usando boas práticas reportadas na literatura científica [34].

Tabela 17: Acurácias do produtor (PA) e do usuário (UA) da classificação do bioma Pantanal (cubos 041051 e 041052) . A acurácia global é 0,7. Alagado quer dizer *Campo alagado e área pantanosa*; Savânica quer dizer *Formação campestre e savânica*; Florestal quer dizer *Formação florestal*.

		Referência					
		Agric.	Alagado	Savânica	Florest.	Pastagem	UA
	Agricultura	25	0	0	0	11	0,69
	Alagado	0	91	0	1	0	0,99
BDC	Savânica	0	48	31	2	3	0,37
	Florestal	0	12	11	52	0	0,69
	Pastagem	1	2	13	5	60	0,74
	PA	0,96	0,59	0,56	0,87	0,81	

Figura 37: Resultado de classificação do Pantanal BDC tiles 041051 e 041052 (Mato Grosso e Mato Grosso do Sul).



5 Divulgação e capacitação

O projeto BDC tem um site bem completo (www.brazildata.cube.org) que contém informações sobre todos os produtos de dados e software gerados; a participação da equipe em eventos, workshops e capacitações; a relação de artigos publicados em revistas e conferências; lançamentos de produtos e serviços; e divulgação de oportunidades. Foram criadas mídias sociais para atingir um maior e mais diversificado público e também estamos no Facebook <https://www.facebook.com/brazildatacube>, Twitter <https://twitter.com/BrazilDataCube> e Youtube <https://www.youtube.com/channel/UCIpWmw1AZ6xlXnnE7DDwSMw>. O alcance e engajamento destas redes sociais podem ser vistas na tabela 18. A seguir, listamos com mais detalhes algumas informações.

5.1 Artigos publicados em revistas e conferências

O projeto BDC tem uma componente forte de pesquisa e inovação. Assim, a equipe tem publicado os resultados obtidos em diversas revistas e conferências nas áreas de Geoinformática e Sensoriamento Remoto. Desde o início do projeto foram publicados 29 artigos em revistas e congressos.

A equipe publicou em Dezembro de 2020 o artigo *Eath Observation Data Cubes for Brazil: Requirements, methodology and Products* na revista internacional *Remote Sensing* que descreve as metodologias e produtos de dados e software desenvolvidos no projeto BDC [1]. Esse artigo foi destacado pela revista como um dos mais baixados entre Novembro de 2020 e Janeiro de 2021 - *TOP Downloaded (Nove 2020 - Jan*

2021). Isso demonstra o interesse internacional nas atividades de desenvolvimentos científicos e tecnológicos do projeto BDC.

Visando maior alcance e ampla divulgação tanto para pesquisadores quanto para a sociedade em geral, optamos por revistas que tenham política de *Open Access* (Acesso Aberto). As publicações são importantes para que outros grupos possam reproduzir nossos experimentos e até a contribuir com o trabalho, validando seu potencial de tecnologia e inovação.

Na seção de Referências estão listadas todas as publicações realizadas no âmbito do projeto, por pesquisadores, contratados e colaboradores associados. Em ordem de data de publicação: [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [7], [56], [57], [58], [59], [60], [61], [62], [63], [20], [1], [13], [17], [10], [21], [64], [22].

5.2 Divulgação

A equipe do BDC participa de diferentes grupos e iniciativas nacionais e internacionais relacionados ao tema de cubos de dados de observação da Terra. Desde o início do projeto, a equipe do BDC participou de 34 eventos, sendo 16 nacionais e 18 internacionais. A relação de eventos com participação e organizados pela equipe BDC pode ser vista na Tabela 19.

Tabela 18: Alcance e engajamento das Redes sociais do BDC.

Pagina Web / Portal

URL: brazildatacube.org

Período: 8 de novembro de 2019 a 27 junho de 2021

61.587 visualizações de Páginas

15.133 Usuários únicos

Fonte: Google Analytics

Perfil no Twitter

URL: <https://twitter.com/BrazilDataCube>

Período: Julho 2020 a 27 de junho de 2021

116.387 impressões

11.516 visitantes

Fonte: Twitter Analytics

Pagina do Facebook

URL: <https://www.facebook.com/brazildatacube>

Período: 15 de Janeiro de 2021 a 27 de junho de 2021

Alcance: 9.330

Fonte: Facebook business

Canal do Youtube

URL: <https://www.youtube.com/channel/UCIpWmw1AZ6xlXnnE7DDwSMw>

Período: 10 de fevereiro de 2021 a 28 de junho de 2021

6.300 impressões

866 visualizações

115 Horas de exibição

Fonte: Studio Youtube

Data	Evento	Tipo	Âmbito
set-19	Minicurso SITS	Curso Realizado	Nacional
set-19	ECML PKDD/MACLEAN	Simpósio	Internacional
set-19	WorCap 2019	Workshop	Nacional
out-19	Pecora	Simpósio	Internacional
nov-19	Geo Week 2019	Simpósio	Internacional
nov-19	MundoGEO Connect	Simpósio	Nacional
nov-19	GEOINFO 2019	Simpósio	Nacional
dez-19	Visita CSIRO	Palestra	Internacional
dez-19	Workshop BDC	Workshop Realizado	Nacional
jan-20	Curso de Verão	Curso Realizado	Nacional
abr-20	WGISS-49	Reunião Técnica	Internacional
jun-20	GEO VIRTUAL SYMPOSIUM 2020	Simpósio	Internacional
ago-20	Curso virtual e hands-on - SITS	Curso Realizado	Nacional
ago-20	Curso virtual e hands-on - SITS	Curso Realizado	Nacional
ago-20	Workshop Digital Earth Americas	Workshop	Internacional
ago-20	LAGIRS 2020	Simpósio	Internacional

Data	Evento	Tipo	Âmbito
set-20	BDC Tech Talks "Open Data Cube: Introdução e Integração com o projeto BDC"	Palestra Realizada	Nacional
set-20	WorCap 2020	Workshop	Nacional
set-20	BDC Tech Talks "SpatioTemporal Asset Catalog - STAC"	Palestra Realizada	Nacional
set-20	WGISS 50	Reunião Técnica	Internacional
out-20	BDC Tech Talks "Web Land Trajectory Service - WLTS"	Palestra Realizada	Nacional
out-20	SBIDE	Simpósio	Nacional
out-20	Palestra MCTI "Cubos de dados e análise de séries temporais de imagens de sensoriamento remoto"	Palestra	Nacional
nov-20	GEOINFO 2020	Simpósio	Nacional
fev-21	GEO-INPE Webinars on Brazil's forest monitoring system	Workshop	Internacional
fev-21	Webinar GEO Knowledge Hub (GKH)	Workshop	Internacional
mar-21	Workshop INPE e University of Twente	Workshop	Internacional

Data	Evento	Tipo	Âmbito
mar-21	Webinar da GEO-AWS EO Cloud Credits	Workshop	Internacional
mar-21	II Workshop do projeto Brazil Data Cube	Workshop Realizado	Nacional
abr-21	II Webinar GEO Knowledge Hub (GKH)	Workshop	Internacional
abr-21	Minicurso "Cubos de Dados de Observação da Terra e Análise de Séries Temporais" no SBSR Interim.	Curso Realizado	Nacional
abr-21	CSIRO event "Exploring the value and challenges of using Earth Observation Data"	Workshop	Internacional
abr-21	AWS Public Sector Summit Online 2021	Workshop	Internacional
abr-21	GEO-OGC Data Cubes Workshop	Workshop	Internacional
jun-21	Workshop GEO-GFOI	Workshop	Internacional

Tabela 19: Eventos com participação e realizados pela equipe BDC

5.3 Documentação

Toda a documentação associada ao projeto BDC está disponível de forma pública. Todos os códigos são abertos e estão disponíveis Github do projeto (<https://github.com/brazil-data-cube>), onde qualquer pessoa pode acessar, reproduzir experimentos, modificar ou contribuir. A documentação dos produtos e serviços desenvolvidos, assim como os tutorias de cursos e treinamentos realizados pela equipe, podem ser encontrados nas páginas do Github.io (<https://brazil-data-cube.github.io/index.html>).

5.4 Eventos de capacitação

Com o objetivo de transmitir conhecimento e capacitar profissionais, a equipe realizou nove eventos de capacitação e treinamento, sendo estes divididos em três cursos, quatro palestras e dois workshops. A seguir são destacados os eventos executados em 2020 e 2021.

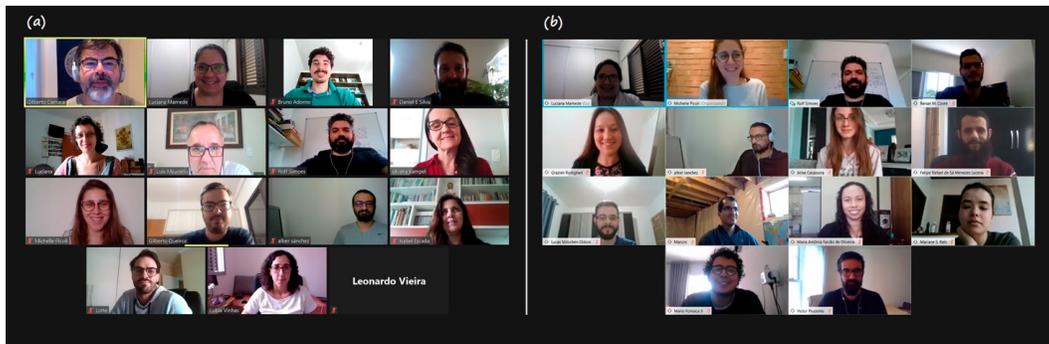
5.4.1 Curso SITS

Em agosto de 2020, a equipe realizou um treinamento virtual e *hands-on* sobre classificação de séries temporais de imagens de sensoriamento remoto usando o pacote de software SITS – *R package for Satellite Image Time Series*.

O treinamento foi de uma semana para duas turmas distintas e foi coordenado pelo Dr. Gilberto Camara com apoio dos membros da equipe, Alber Sanchez, Gilberto Queiroz, Felipe Carlos, Felipe Souza, Michelle Picolli e Rolf Simões. O objetivo

do curso foi demonstrar de maneira prática o uso do pacote SITS usando como exemplo imagens e cubos de dados MODIS e Sentinel-2. Contamos com a participação de cerca de 25 pessoas, como membros da equipe do projeto PRODES, alunos da pós-graduação em Sensoriamento Remoto e servidores do INPE. (Figura 38)

Figura 38: Participantes do curso SITS nos dias (a) 14/08/2020 e (b) 19/08/2021.



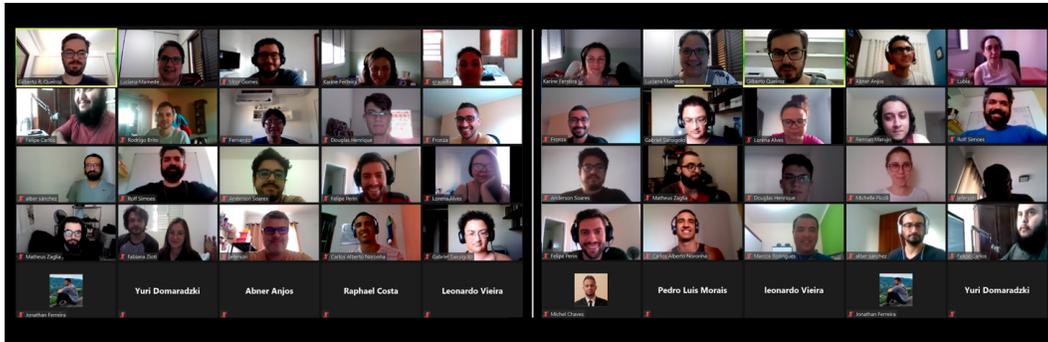
5.4.2 Tech Talks

A equipe do projeto também realizou três BDC-Tech Talks, que são palestras internas coordenadas pela equipe de desenvolvimento do BDC, com *hands-on* sobre tecnologias utilizadas ou desenvolvidas no projeto com o objetivo de integrar e nivelar toda a equipe.

Em 04 de setembro de 2020, o 1º BDC-Tech Talks foi sobre o tema *Open Data Cube: Introdução e Integração com o projeto BDC*, apresentado pelos colaboradores associados Felipe Carlos e Vitor Gomes. No dia 18 de setembro de 2020, tivemos o tema *Usando o STAC com Python e R*, realizado pelos colaboradores Matheus Zaglia e Felipe Souza. E, por fim, no dia 02 de Outubro de 2020 foi apresentado o serviço *Web Land Trajectory Service - WLTS*, coordenado pela Fabiana Zioti. As fotos na

Figura 39 foram tiradas em dois desses eventos.

Figura 39: BDC Tech Talks

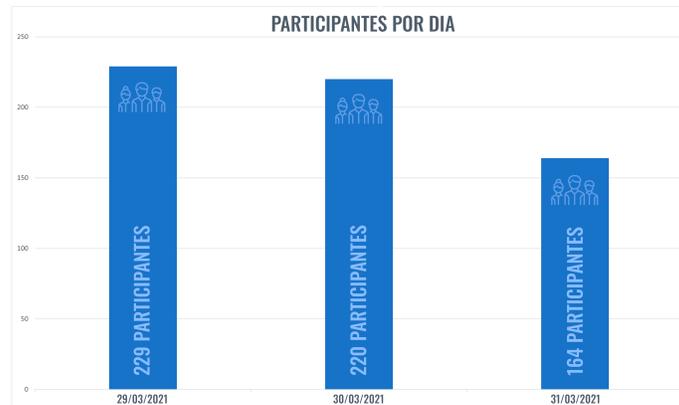


5.4.3 II Workshop do Brazil Data Cube

Nos dias 29, 30 e 31 de março de 2021 foi realizado o II Workshop do Brazil Data Cube online. O workshop contou com a participação de 200 pessoas por dia (Figura 40), em média, de cerca de 100 diferentes instituições governamentais, universidades e empresas privadas (Figura 41) das áreas de Geoinformática, Sensoriamento Remoto e Tecnologia da Informação. Dentre as instituições estiveram presentes BNDES, EMBRAPA, IBGE, CENSIPAM, entre outras.

A primeira sessão do workshop foi dedicada à uma visão geral do projeto, seu andamento e seu contexto no INPE/Funcate/BNDES. Nas demais sessões, a equipe realizou demonstrações práticas ensinando os participantes a usarem os dados e sistemas desenvolvidos no projeto. As fotos da Figura 42 são da equipe do projeto que organizou e realizou o evento.

Figura 40: Participantes do II Workshop BDC por dia.



5.4.4 SBSR Interim

Outro evento que destacamos é a realização do minicurso "Cubos de Dados de Observação da Terra e Análise de Séries Temporais" realizado pela equipe do BDC no evento SBSR Interim. Esse evento foi organizado pela Coordenação do Simpósio Brasileiro de Sensoriamento Remoto (SBSR) e realizado online de 05 a 16 de abril de 2021.

Esse minicurso foi de 6 horas e teve como objetivo apresentar uma introdução sobre conceitos de cubos de dados de observação da Terra e análise de séries temporais de imagens de sensoriamento remoto. O curso teve uma parte teórica e uma parte prática *hands-on* utilizando os cubos de dados e os pacotes de software criados pelo projeto BDC nas linguagens R e Python. O minicurso contou com cerca de 400 participantes online e mais de 80 perguntas respondidas ao vivo durante o evento.

6 Recursos humanos e infraestrutura

6.1 Recursos humanos

A equipe do BDC é formada, principalmente, por desenvolvedores, analistas de sistemas e especialistas em geoinformática e sensoriamento remoto, muitos com pós-graduação e com experiência na área. Atualmente são catorze contratados entre bolsistas e celetistas.

As contratações são realizadas após processos de seleção que são amplamente divulgados e abertos ao público, onde os Termos de Referência (TR) são escritos de acordo com cada perfil de vaga necessitada. Até o momento foram publicados vinte TRs para seleção de candidatos.

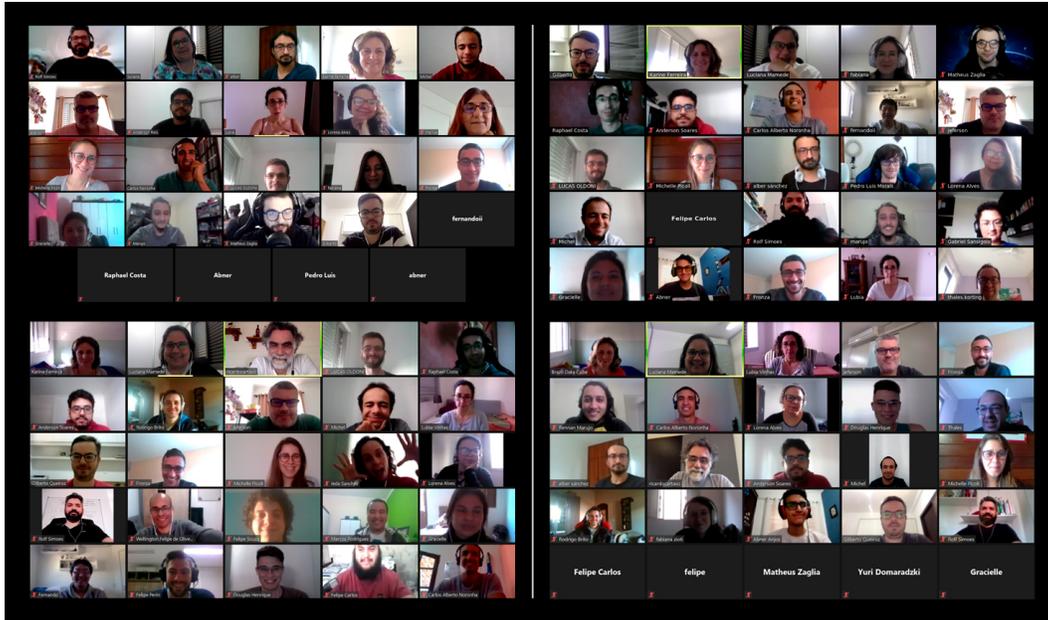
6.2 Trabalho remoto

Desde março de 2020 a equipe do BDC tem trabalhado em sistema *home office* devido à pandemia do Corona Vírus. A equipe se adaptou bem e manteve o desenvolvimento das atividades do projeto com muita produtividade.

Desde então, reuniões periódicas entre membros da equipe e com os coordenadores passaram a ser realizadas de forma virtual utilizando a ferramenta de reuniões online Zoom, cuja licença foi adquirida pelo projeto. Além disso, outras plataformas e ferramentas que auxiliam a integração da equipe durante o trabalho remoto foram adotadas, como por exemplo o Discord para comunicação instantânea entre a equipe com possibilidade de compartilhamento de tela. A figura 43 mostra algumas fotos

das reuniões online realizadas em 2020 e 2021.

Figura 43: Fotos de reunião online realizadas pela equipe do BDC.

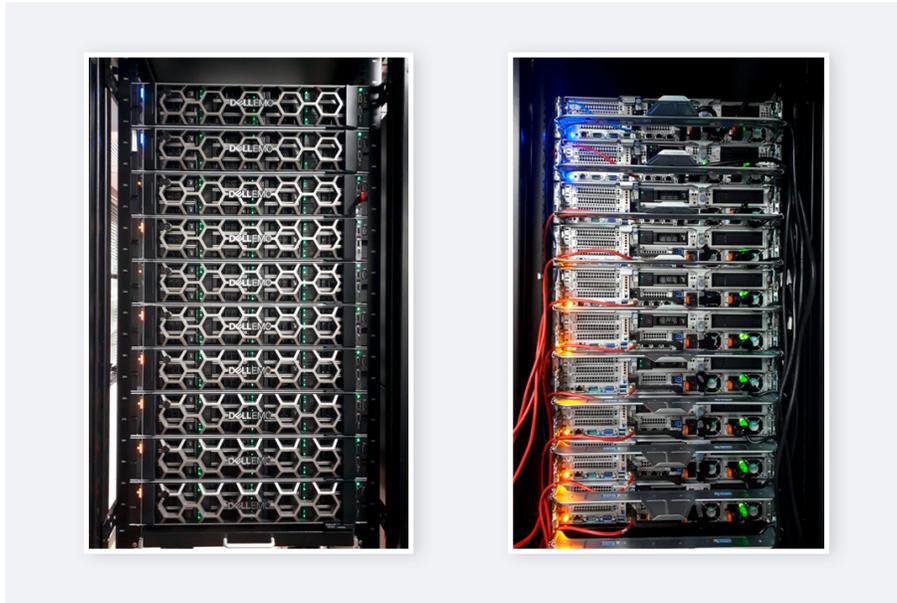


6.3 Aquisições

Para apoiar o desenvolvimento e criar um infraestrutura adequada de processamento, foram adquiridos softwares e equipamentos de informática. Para manter a plataforma BDC processando de modo contínuo e armazenar as coleções, os cubos de dados e os dados auxiliares, foram adquiridos 11 servidores e 2 *arrays* de discos com alto poder de processamento. A figura 44 ilustra o *rack* com os servidores do projeto. Além disso, foram adquiridos também dois *switchs gigabit*, material para instalação e manutenção da energia elétrica, como baterias de *nobreaks*, *pdus* e cabos.

Para uso individual da equipe do projeto, foram compradas quatro *workstations*

Figura 44: Rack com servidores do projeto



e dez *notebooks* com alto poder de processamento. Além dos equipamentos, com o intuito de subsidiar a realização do *Workshop*, foram adquiridos itens personalizados para serem distribuídos para os participantes e para equipe.

6.4 Infraestrutura

Atualmente o projeto BDC conta com quatorze servidores de *rack* e três *arrays* de discos, todos interconectados por interface de rede Ethernet a 10 Gbits/s. Desses servidores, oito foram adquiridos recentemente e estão em processo final de configuração para serem utilizados. Somando todos os recursos disponíveis temos um ambiente com aproximadamente 1.5 PB (Petabyte) de disco para armazenamento; Mais de 3.1 TB de memória RAM e 624 CPUs.

Os recursos desses servidores são compartilhados entre as diversas aplicações e

sistemas da plataforma BDC. Tem-se instalado nos servidores o Sistema Operacional Linux Ubuntu versões 16 ou 18, dependendo do servidor. As aplicações do BDC rodam em *containers*, implementados com Docker.

6.4.1 GlusterFS

O sistema *Gluster File System* (GlusterFS) foi implementado nos servidores do projeto BDC. O GlusterFS é um sistema de arquivos distribuído e escalável que agrega recursos de discos de múltiplos servidores em um único sistema de arquivos. Para a aplicação do projeto BDC, decidiu-se distribuir os arquivos entre os *bricks* que compõem o volume, o que é chamado de volume distribuído. Esse tipo de configuração pode ser usado quando o requisito é escalar o sistema de *storage* e a redundância nos dados não é importante ou um mecanismo de tolerância a falhas é fornecido por outras camadas de hardware ou software. No caso das máquinas que compõem o *cluster* do BDC, a tolerância a falhas de disco é feita pelo RAID5, implementado via hardware nos arrays de discos de cada servidor.

6.4.2 Manutenção da infraestrutura computacional

Outro trabalho importante realizado pela equipe do BDC consiste em manter o ambiente computacional em perfeito funcionamento para que todas as atividades de desenvolvimento, processamento e análise de dados descritos acima possam ser executadas sem adversidades e também para garantir que os dados e serviços fornecidos pelo projeto estejam sempre disponíveis ao público. Pode-se elencar nesse contexto as seguintes atividades:

- Monitoramento do ambiente de ar condicionado onde os servidores são instalados.
- Monitoramento de rede elétrica e status dos *nobreaks* e baterias que sustentam os servidores BDC.
- Verificação das condições dos servidores, discos e *arrays* de discos que compõem o ambiente operacional.
- Instalação, configuração e manutenção do ambiente computacional que contempla Sistema Operacional, Sistema de Arquivos Distribuídos, APIs do BDC, Catálogos de imagens, dentre outros.
- Especificações de equipamentos para novas compras e, conseqüentemente, escrita dos Termos de Referência.
- Acompanhamento dos processos licitatórios relacionados a infraestrutura.
- Contato com os fornecedores e acompanhamento da instalação física dos equipamentos adquiridos.

7 Comentários Finais e Próximos Passos

Usando as ferramentas de software desenvolvidas e os produtos de dados adquiridos, a geração de coleções e de cubos de dados continuará a ser feita no projeto a fim de cobrir todo o território nacional com imagens mais recentes, principalmente os produtos e cubos Sentinel-2. Além disso, há estudos e testes envolvendo a

geração de produtos multi-sensores, como por exemplo: (1) uso conjunto de dados Landsat-8 e Sentinel-2; (2) de dados históricos do programa Landsat, utilizando em conjunto dados Landsat-5, Landsat-7 e Landsat-8; e (3) cubos do sensor WFI com alta revisita temporal, por exemplo semanal, contando com imagens dos satélites CBERS-4, CBERS-4A e Amazônia-1. O item (1) já encontra-se em desenvolvimento e apresenta resultados preliminares promissores. Em relação ao item (3), no momento estamos finalizando a calibração e equalização das imagens de Reflectância de Superfície do satélite CBERS-4A para que elas sejam equivalentes às imagens do CBERS-4. Em seguida, o mesmo trabalho será executado para os dados da câmera WFI do Amazônia-1 lançado recentemente.

Os sistemas de software que estão sendo desenvolvidos no projeto BDC podem ser utilizados por outros países que não dominam as tecnologias para criação de cubos de dados de observação da Terra a partir de grandes volumes de imagens de sensoriamento remoto. Atualmente, existe uma iniciativa liderada pelo INPE, GEO (Grupo de Observação da Terra, em inglês *Group of Earth Observation*) e CfRN (Coalizão de Países com Florestas Tropicais, em inglês *Coalition for Rainforest Nations*) para utilizar as tecnologias desenvolvidas no projeto BDC para criar cubos de dados de observação da Terra para Moçambique. Esse projeto piloto está em andamento e tem como principais objetivos: (1) Criar dados ARD e cubos de dados de observação da Terra de imagens Landsat de 1990 a 2019 para Moçambique; e (2) Produzir mapas de uso e cobertura da Terra a partir desses cubos, utilizando métodos de aprendizado de máquina e análise de séries temporais. Essa é a primeira iniciativa de aplicar as ferramentas do projeto BDC para outro país.

Em relação às classificações de uso e cobertura da Terra realizadas pelo projeto BDC para atestar a qualidade dos cubos de dados e métodos produzidos, os próximos passos abrangem novos experimentos nos três biomas brasileiros que não foram contemplados neste relatório: Mata Atlântica, Caatinga e Pampa. Além disso, a equipe pretende investir no estabelecimento de parcerias com outros projetos e instituições para testar e difundir o uso dos cubos de dados e das técnicas de aprendizagem de máquinas desenvolvidas.

Uma destas parcerias será com o projeto TerraClass, que realiza o mapeamento do uso e cobertura da Terra para o bioma Cerrado e Amazônia. Este projeto é executado pelo INPE e Embrapa e, atualmente, realiza a maior parte de sua interpretação de imagens de satélite de maneira visual. Dessa forma, a inclusão de tecnologias criadas pelo projeto BDC permitirá a criação de classificações de maneira mais rápida e reprodutível.

Dentro das necessidades para continuar o andamento do projeto estão a aquisição de mais servidores para processar e armazenar as coleções e cubos de dados até 2022 e a aquisição de créditos da AWS para gerar cubos de dados de imagens Sentinel-2. A AWS, através de seu programa *Earth on AWS* (<https://aws.amazon.com/earth/>), mantém um repositório aberto e sem custos de imagens de satélites de observação o da Terra. O projeto BDC está usando o bucket Sentinel-2 da AWS para gerar seus cubos de dados, sem a necessidade de fazer download desses dados para o INPE.

Referências

- [1] K. R. Ferreira, G. R. Queiroz, L. Vinhas, R. F. B. Marujo, R. E. O. Simoes, M. C. A. Picoli, G. Camara, R. Cartaxo, V. C. F. Gomes, L. A. Santos, A. H. Sanchez, J. S. Arcanjo, J. G. Fronza, C. A. Noronha, R. W. Costa, M. C. Zaglia, F. Zioti, T. S. Korting, A. R. Soares, M. E. D. Chaves, and L. M. G. Fonseca, “Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products,” *Remote Sensing*, vol. 12, p. 4033, dec 2020.
- [2] A. Siqueira, T. Tadono, A. Rosenqvist, J. Lacey, A. Lewis, M. Thankappan, Z. Szantoi, P. Goryl, S. Labahn, J. Ross, S. Hosford, and S. Mecklenburg, “CEOS Analysis Ready Data For Land – An Overview on the Current and Future Work,” in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, (Yokohama, Japan), pp. 5536–5537, IEEE, July 2019.
- [3] M. Appel and E. Pebesma, “On-Demand Processing of Data Cubes from Satellite Image Collections with the gdalcubes Library,” *Data*, vol. 4, p. 92, June 2019.
- [4] E. Vermote, C. Justice, M. Claverie, and B. Franch, “Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product,” *Remote Sensing of Environment*, vol. 185, pp. 46–56, 2016.
- [5] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, “Sen2cor for sentinel-2,” in *Image and Signal Processing for Remote Sensing*

- XXIII*, vol. 10427, p. 1042704, International Society for Optics and Photonics, 2017.
- [6] S. Qiu, Z. Zhu, and B. He, “Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery,” *Remote Sensing of Environment*, vol. 231, p. 111205, 2019.
- [7] A. H. Sanchez, M. C. A. Picoli, G. Camara, P. R. Andrade, M. E. D. Chaves, S. Lechler, A. R. Soares, R. F. B. Marujo, R. E. O. Simões, K. R. Ferreira, and G. R. Queiroz, “Comparison of Cloud Cover Detection Algorithms on Sentinel-2 Images of the Amazon Tropical Forest,” *Remote Sensing*, vol. 12, no. 8, p. 1284, 2020.
- [8] D. P. Roy, H. K. Zhang, J. Ju, J. L. Gomez-Dans, P. E. Lewis, C. B. Schaaf, Q. Sun, J. Li, H. Huang, and V. Kovalsky, “A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance,” *Remote Sensing of Environment*, vol. 176, pp. 255–271, 2016.
- [9] Brazil Data Cube Team, *rstac: Client Library for SpatioTemporal Asset Catalog*. National Institute for Space Research (INPE), Brazil, 2021.
- [10] R. Simoes, G. Camara, G. Queiroz, F. Souza, P. R. Andrade, L. Santos, A. Carvalho, and K. Ferreira, “Satellite image time series analysis for big earth observation data,” *Remote Sensing*, vol. 13, no. 13, 2021.

- [11] C. Pelletier, G. I. Webb, and F. Petitjean, “Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series,” *Remote Sensing*, vol. 11, no. 5, 2019.
- [12] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, and F. Petitjean, “InceptionTime: Finding AlexNet for time series classification,” *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936–1962, 2020.
- [13] L. A. Santos, K. Ferreira, M. Picoli, G. Camara, R. Zurita-Milla, and E.-W. Augustijn, “Identifying spatiotemporal patterns in land use and cover samples from satellite image time series,” *Remote Sensing*, vol. 13, p. 974, 3 2021.
- [14] A. E. Maxwell, T. A. Warner, and F. Fang, “Implementation of machine-learning classification in remote sensing: An applied review,” *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784–2817, 2018.
- [15] P. Thanh Noi and M. Kappas, “Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery,” *Sensors*, vol. 18, no. 1, p. 18, 2018.
- [16] B. Frenay and M. Verleysen, “Classification in the Presence of Label Noise: A Survey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 845–869, 2014.

- [17] L. A. Santos, K. R. Ferreira, G. Camara, M. C. Picoli, and R. E. Simoes, “Quality control and class noise reduction of satellite image time series,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 75–88, 7 2021.
- [18] T. Kohonen, “The self-organizing map,” *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.
- [19] Open Data Cube, “ODC: Architecture and Ecosystem - A High-Level Overview,” tech. rep., Open Data Cube, 2019.
- [20] F. M. Carlos, V. C. F. Gomes, G. R. D. Queiroz, K. R. Ferreira, and R. Santos, “Integração dos ambientes Brazil Data Cube e Open Data Cube,” in *Proceedings XXI GEOINFO*, pp. 168–173, 2020.
- [21] V. C. F. Gomes, F. M. Carlos, G. R. Queiroz, K. R. Ferreira, and R. Santos, “Accessing and processing brazilian earth observation data cubes with the open data cube platform,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-4-2021, pp. 153–159, 2021.
- [22] F. M. Carlos, V. C. F. Gomes, G. R. D. Queiroz, F. Carvalho, K. R. Ferreira, and R. Santos, “Integrating Open Data Cube and Brazil Data Cube platforms for land use and cover classifications,” in *Artigo aceito na Revista Brasileira de Cartografia*, 2021.
- [23] T. Garnett, “Where are the best opportunities for reducing greenhouse gas emissions in the food system (including the food chain)?,” *Food Policy*, vol. 36, no. SUPPL. 1, pp. S23–S32, 2011.

- [24] D. Nepstad, D. McGrath, C. Stickler, A. Alencar, A. Azevedo, B. Swette, T. Bezerra, M. DiGiano, J. Shimada, R. Seroa da Motta, E. Armijo, L. Castello, P. Brando, M. C. Hansen, M. McGrath-Horn, O. Carvalho, and L. Hess, “Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains,” *Science*, vol. 344, pp. 1118–1123, jun 2014.
- [25] M. C. A. Picoli, G. Camara, I. Sanches, R. Simões, A. Carvalho, A. Maciel, A. Coutinho, J. Esquerdo, J. Antunes, R. A. Begotti, *et al.*, “Big earth observation time series analysis for monitoring brazilian agriculture,” *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 328–339, 2018.
- [26] J. R. Irons, J. L. Dwyer, and J. A. Barsi, “The next Landsat satellite: The Landsat Data Continuity Mission,” *Remote Sensing of Environment*, vol. 122, pp. 11–21, jul 2012.
- [27] D. Roy, M. Wulder, T. Loveland, W. C.E., R. Allen, M. Anderson, D. Helder, J. Irons, D. Johnson, R. Kennedy, T. Scambos, C. Schaaf, J. Schott, Y. Sheng, E. Vermote, A. Belward, R. Bindschadler, W. Cohen, F. Gao, J. Hipple, P. Hostert, J. Huntington, C. Justice, A. Kilic, V. Kovalsky, Z. Lee, L. Lymburner, J. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. Wynne, and Z. Zhu, “Landsat-8: Science and product vision for terrestrial global change research,” *Remote Sensing of Environment*, vol. 145, pp. 154–172, apr 2014.
- [28] European Space Agency, “S2 MPC L2A Product Definition Document,” tech. rep., European Space Agency, 2018.

- [29] R. Ihaka and R. Gentleman, “R: A Language for Data Analysis and Graphics,” *Journal of Computational and Graphical Statistics*, vol. 5, p. 299, sep 1996.
- [30] L. V. Oldoni, I. D. Sanches, M. C. A. Picoli, R. M. Covre, and J. G. Fronza, “LEM+ dataset: For agricultural remote sensing applications,” *Data in Brief*, vol. 33, p. 106553, dec 2020.
- [31] Y. E. Shimabukuro, J. R. Santos, A. R. Formaggio, V. Duarte, and B. F. Rudorff, “The Brazilian Amazon Monitoring Program: PRODES and DETER Projects,” in *Global Forest Monitoring from Earth Observation* (F. Achard and M. C. Hansen, eds.), ch. 9, pp. 166–183, CRC Press, 2012.
- [32] C. M. Souza, J. Z. Shimbo, M. R. Rosa, L. L. Parente, A. A. Alencar, B. F. T. Rudorff, H. Hasenack, M. Matsumoto, L. G. Ferreira, P. W. M. Souza-Filho, S. W. de Oliveira, W. F. Rocha, A. V. Fonseca, C. B. Marques, C. G. Diniz, D. Costa, D. Monteiro, E. R. Rosa, E. Vélez-Martin, E. J. Weber, F. E. B. Lenti, F. F. Paternost, F. G. C. Pareyn, J. V. Siqueira, J. L. Viera, L. C. F. Neto, M. M. Saraiva, M. H. Sales, M. P. G. Salgado, R. Vasconcelos, S. Galano, V. V. Mesquita, and T. Azevedo, “Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine,” *Remote Sensing*, vol. 12, p. 2735, aug 2020.
- [33] Y. Finegold, A. Ortmann, E. Lindquist, R. d’Annunzio, and M. Sandker, “Map accuracy assessment and area estimation: a practical guide,” tech. rep., FAO, 2016.

- [34] P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, “Good practices for estimating area and assessing accuracy of land change,” *Remote Sensing of Environment*, vol. 148, pp. 42–57, may 2014.
- [35] JARDIM BOTÂNICO DO RIO DE JANEIRO - JBRJ, “Refflora: lista de espécies da flora do Brasil,” 2015.
- [36] Gibbs, A. S. Ruesch, F. Achard, M. K. Clayton, P. Holmgren, N. Ramankutty, and J. A. Foley, “Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 38, pp. 16732–16737, 2010.
- [37] Z. Szantoi, G. N. Geller, N.-E. Tsendbazar, L. See, P. Griffiths, S. Fritz, P. Gong, M. Herold, B. Mora, and A. Obregón, “Addressing the need for improved land cover map products for policy support,” *Environmental Science & Policy*, vol. 112, pp. 28–35, oct 2020.
- [38] Y. Malhi, J. T. Roberts, R. A. Betts, T. J. Killeen, W. Li, and C. A. Nobre, “Climate Change, Deforestation, and the Fate of the Amazon,” *Science*, vol. 319, pp. 169–172, jan 2008.
- [39] INPE - Instituto Nacional de Pesquisas Espaciais, “Programa de monitoramento da Amazônia e demais biomas. Desmatamento – Amazônia Legal.” <http://terrabrasilis.dpi.inpe.br/downloads/>, 2021.

- [40] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, “High-resolution mapping of global surface water and its long-term changes,” *Nature*, vol. 540, no. 7633, pp. 418–422, 2016.
- [41] W. J. Junk, M. T. F. Piedade, R. Lourival, F. Wittmann, P. Kandus, L. D. Lacerda, R. L. Bozelli, F. A. Esteves, C. N. da Cunha, L. Maltchik, J. Schöngart, Y. Schaeffer-Novelli, and A. A. Agostinho, “Brazilian wetlands: their definition, delineation, and classification for research, sustainable management, and protection,” *Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 24, no. 1, pp. 5–22, 2013.
- [42] W. M. Tomas *et al.*, “Sustainability agenda for the pantanal wetland: Perspectives on a collaborative interface for science, policy, and decision-making,” *Tropical Conservation Science*, vol. 12, p. 194008291987263, 2019.
- [43] A. Guerra, F. de Oliveira Roque, L. C. Garcia, J. M. Ochoa-Quintero, P. T. S. de Oliveira, R. D. Guariento, and I. M. Rosa, “Drivers and projections of vegetation loss in the pantanal and surrounding ecosystems,” *Land Use Policy*, vol. 91, p. 104388, 2020.
- [44] L. Santos, K. R. Ferreira, M. Picoli, and G. Camara, “Self-organizing maps in earth observation data cubes analysis,” *Advances in Intelligent Systems and Computing*, vol. 976, pp. 70–79, 2019.
- [45] K. R. Ferreira, L. Santos, and M. C. A. Picoli, “Evaluating distance measures for image time series clustering in land use and cover monitoring,” in *2019 Euro-*

pean Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)., 9 2019.

- [46] K. Ferreira, G. Queiroz, G. Camara, R. Souza, L. Vinhas, R. Simões, C. Noronha, R. Costa, J. Arcanjo, R. Marujo, V. Gomes, and M. Zaglia, “Brazil data cube project using aws-initial results,” tech. rep., INPE, 2019.
- [47] R. S. S. Adeu, K. R. Ferreira, P. R. Andrade, and L. Santos, “Evaluating growing self-organizing maps for satellite image time series clustering,” in *Proceedings XX GEOINFO*, 2019.
- [48] M. C. Zaglia, L. Vinhas, G. R. D. Queiroz, and R. Simoes, “Catalogação de metadados do cubo de dados do brasil com o spatiotemporal asset catalog,” in *Proceedings XX GEOINFO*, pp. 280–285, 2019.
- [49] G. Sansigolo, G. R. D. Queiroz, and K. R. Ferreira, “Projetando uma Plataforma para Compartilhamento de Dados Científicos de Observação da Terra,” in *Proceedings XX GEOINFO*, pp. 249–254, 2019.
- [50] F. C. de Souza, R. D. C. dos Santos, and K. R. Ferreira, “Ggsom: Ferramenta de visualização baseada em mapas auto-organizáveis,” in *Proceedings XX GEOINFO*, pp. 310–312, 2019.
- [51] F. Zioti, G. R. D. Queiroz, and K. R. Ferreira, “Análise de ferramentas para processamento de grandes volumes de dados espaço-temporais,” in *Proceedings XX GEOINFO*, pp. 255–260, 2019.

- [52] A. Sanchez, M. Picoli, P. R. Andrade, R. Simões, L. Santos, M. Chaves, R. Begotti, and G. Camara, “Land cover classifications of clear-cut deforestation using deep learning,” in *Proceedings XX GEOINFO*, pp. 48–56, 2019.
- [53] M. C. A. Picoli, A. Rorato, P. Leitão, G. Camara, A. Maciel, P. Hostert, and I. D. Sanches, “Impacts of public and private sector policies on soybean and pasture expansion in mato grosso—brazil from 2001 to 2017,” *Land*, vol. 9, p. 20, 1 2020.
- [54] R. Simoes, M. C. A. Picoli, G. Camara, A. Maciel, L. Santos, P. R. Andrade, A. Sánchez, K. Ferreira, and A. Carvalho, “Land use and cover maps for mato grosso state in brazil from 2001 to 2017,” *Scientific Data*, vol. 7, p. 34, 12 2020.
- [55] V. Gomes, G. Queiroz, and K. Ferreira, “An overview of platforms for big earth observation data management and analysis,” *Remote Sensing*, vol. 12, p. 1253, 4 2020.
- [56] A. R. Soares, T. S. Korting, L. M. G. Fonseca, and A. K. Neves, “An unsupervised segmentation method for remote sensing imagery based on conditional random fields,” in *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, pp. 1–5, IEEE, 3 2020.
- [57] K. R. Ferreira, G. R. Queiroz, G. Camara, R. C. M. Souza, L. Vinhas, R. F. B. Marujo, R. E. O. Simoes, C. A. F. Noronha, R. W. Costa, J. S. Arcanjo, V. C. F. Gomes, and M. C. Zaglia, “Using remote sensing images and cloud services on aws to improve land use and cover monitoring,” in *2020 IEEE Latin American*

- GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, pp. 558–562, IEEE, 3 2020.
- [58] A. R.S., F. K.R., A. P.R., and S. L., “Assessing satellite image time series clustering using growing som.,” in *Computational Science and Its Applications ICCSA 2020.*, Springer, Cham, 2020.
- [59] A. R. Soares, H. N. Bendini, D. V. Vaz, T. D. T. Uehara, A. K. Neves, S. Lechler, T. S. Korting, and L. M. G. Fonseca, “Stmetrics: A python package for satellite image time-series feature extraction,” in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium.*, pp. 2061–2064, IEEE, 9 2020.
- [60] M. C. A. Picoli, R. Simoes, M. Chaves, L. A. Santos, A. Sanchez, A. Soares, I. D. Sanches, K. R. Ferreira, and G. R. Queiroz, “Cbbers data cube: A powerful technology for mapping and monitoring brazilian biomes,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2020, pp. 533–539, 8 2020.
- [61] M. E. D. Chaves, M. C. A. Picoli, and I. D. Sanches, “Recent applications of landsat 8/oli and sentinel-2/msi for land use and land cover mapping: A systematic review,” *Remote Sensing*, vol. 12, p. 3062, 9 2020.
- [62] A. Soares, T. Korting, L. Fonseca, and H. Bendini, “Simple nonlinear iterative temporal clustering,” *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–11, 11 2020.

- [63] F. R. D. Sá, M. Lucena, E. V. Escobar-Silva, R. D. Freitas, B. Marujo, C. Zaglia, L. Vinhas, K. R. Ferreira, and G. R. D. Queiroz, “Brazil data cube cloud coverage (bdc3) viewer,” in *Proceedings XXI GEOINFO*, pp. 222–227, 2020.
- [64] R. F. B. Marujo, J. G. Fronza, A. R. Soares, G. R. Queiroz, and K. R. Ferreira, “Evaluating the impact of lasrc and sen2cor atmospheric correction algorithms on landsat-8/oli and sentinel-2/msi data over aeronet stations in brazilian territory,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2021, pp. 271–277, 6 2021.