Sen4CAP Final User Workshop- Technical Session

Chat log - Day 2 - 5 March 2021

Guido Lemoine (EC-JRC): 09:52: what is the pixel spacing of the S1 level 2 outputs?

Laurențiu Nicola (CSRO): 09:54: @Guido we currently use 20 m/px

Thanassis Drivas: 09:54: I have already a couple of questions: 1. if an AOI has many acquisitions due to overlapping do you calculate a mean based on time windows? 2. Do u consider sen2cor better on identifying clouds than MAJA or FMASK?

Guido Lemoine (EC-JRC): 09:54: Is level 3 (LAI etc.) processing done at image level or at extracted parcel level?

Pieter Roggemans: 09:55: Is it already possible to calculate markers/raster statistics with overlapping input parcels?

Laurențiu Nicola (CSRO): 09:55: @Thanassis 1. that depends on the processor, L4A composites on a weekly basis, the others I think not; 2. MAJA and FMASK are generally better at detecting clouds

Diane Heymans: 09:56: @Guido: the L3B (LAI, etc.) are done by pixel then extract by parcel in the DB

Cosmin Udroiu (CS ROMANIA): 09:56: @Guido L3B are processed at image level. After that, markers (std dev, mean) are extracted for each BI at parcel level.

Laurențiu Nicola (CSRO): 09:57: @Pieter the L4A computes the statistics for all parcels, but only classifies non-overlapping ones; the others parcels are still available in the markers file. I'm not sure about the other processors

Guido Lemoine (EC-JRC): 10:00: How does a weekly NDVI practice marker work if you have a lengthy cloud period that is longer than a week?
Guido Lemoine (EC-JRC): 10:01: does it come from smoothing?

Sylwester Zaprzala (Tragsatec): 10:01: From Tragsatec (Spain) we are interested in Sentinel data extraction, not in ML process because we have our own models. thus, we use mainly NDVI, VH (descendent orbit), VH/VV Ratio etc. We are also interested in other indexes that would not enter in our ML but we will use as markers: for example, MNDWI for water detection in parcels, other indexes for buildings detection etc... It is possible extract only those indexes that we are looking for?

PAVEL MILENOV: 10:02: Do you make any topology checks on the input shape file?

Maurizio Laterza: 10:02: It's not really clear to me: markers have to be provided from the user or are they automatically extracted from parcels?

Falah Fakhri: 10:02: Thanks a lot Phillipe, did you try up to add other features resulted from S2, such as EVI, SAVI, and check-up the results?

Laurențiu Nicola (CSRO): 10:02: @Maurizio they are extracted automatically

Laurențiu Nicola (CSRO): 10:03: @Sylwester you can't define new markers easily, but you can access the ones that we compute. That includes a form of NDWI (I'd have to check which one).

Falah Fakhri: 10:03: Phillipe, could you share the link to the model, or you already shared it within your slides, and I didn't have a look.

Thanassis Drivas: 10:04: so there is no interpolation or data fusion?

Laurențiu Nicola (CSRO): 10:04: @Sylwester you can also obtain a list of

products from the database and compute the markers yourself

Thanassis Drivas: 10:04: when you have cloud data?

Laurențiu Nicola (CSRO): 10:04: @Thanassis the L4A does gap-filling by interpolation

Falah Fakhri: 10:04: My last question Phillipe, do you think other features could be extracted and added up from S1?

Sylwester Zaprzala (Tragsatec): 10:06: Thank you very much @Laurentiu ;), it is not easy because it can only be done by modifying the extraction script?

BERIAUX Emilie: 10:07: Do the user has to define some treesholds (ex: for the high NDVI value for M1 marker)?

Laurențiu Nicola (CSRO): 10:08: @Sylwester that kind of depends on where you want to stick it (in the L4A processor which does some gap-filling), or only compute them per-product. The L4A is a bit more complicated because it reads the entire time series at once, keeping only small regions in memory, so it's written in C++ and needs to know exactly the number of extracted markers and pass it through one computation to another

Harijs Ijabs: 10:08: Thank you for the presentation. When LPIS data is added/modified, does the system recalculate the markers for all fields or just the modified ones?

Laurențiu Nicola (CSRO): 10:08: @Harijs they are recomputed
Benjamin Kroll: 10:09: What are the differences between EFA practices in L4C practice ("Catch Crop", "NFC", "Fallow" or "NA")? What should I use if I want to detect mulching on fallow land and in which collum is the result that there was a mulching?

Laurențiu Nicola (CSRO): 10:10: @Sylwester if you want to compute the markers from a single product at once, I guess that's easier

Harijs Ijabs: 10:10: Thank you!

Cosmin Udroiu (CS ROMANIA): 10:10: @Emilie: The thresholds are defined in the configuration file for the L4C processor that needs to be uploaded by the user. Nevertheless, a sample configuration file is provided with some default values.

Sylwester Zaprzala (Tragsatec): 10:10: @Laurentiu I am only interested in extracting data and storing it in the database (processes L2 and L3). The rest of the processes (L4) are implemented by the company where I work.

BERIAUX Emilie: 10:10: thank you Cosmin

David Nafría (ITACyL-Spain): 10:10: The update of just newly added GSAA is essential in operational environment

David Nafría (ITACyL-Spain): 10:15: I miss the actual value of the bands in MDB1.

Sylwester Zaprzala (Tragsatec): 10:15: thank you Laurentiu ;)

David Nafría (ITACyL-Spain): 10:16: because in MDB4 you have the bands but are interpolated

Guido Lemoine (EC-JRC): 10:17: Your marker concept is different from what we use in cbm. Your markers are simply aggregated time series. They are definitely NOT practice markers

Laurențiu Nicola (CSRO): 10:17: @David that's true, we only extract the ones we use

Laurențiu Nicola (CSRO): 10:17: @Guido that depends on the processor; some of the markers are reflectances or indices, some are actual markers (like the one Philippe is showing)
Laurențiu Nicola (CSRO): 10:17: The naming is indeed a bit unfortunate

Thanassis Drivas: 10:18: Is coherence generated by snap (snappy) or do you use subtract VV and VH?

Laurențiu Nicola (CSRO): 10:19: @Thanassis it's generated by SNAP

Thanassis Drivas: 10:19: @Laurentiu thanks! It's too slow although....

Laurențiu Nicola (CSRO): 10:20: @Thanassis yes!

Harijs Ijabs: 10:21: Just to clarify. So, the L4A processor retrieves data from the L4A mdb and does not recompute entire time series when it is run? (provided parcels don't change)

Anders Munck (FI): 10:22: Will the "technical lessons learned" from Sen4CAP feed to the various components that it incorporates, e.g. the above-mentioned SNAP issue?

Guido Lemoine (EC-JRC): 10:22: do you store both the full temporal time series AND the MDB L4A "markers", or only the latter?

Laurențiu Nicola (CSRO): 10:23: @Guido L4A does gap-filling on the S2 data due to the clouds, but S1 has its own acquisition density inconsistencies, which is why we still do a gap-filling of the aggregated parcel values

Laurențiu Nicola (CSRO): 10:23: @Harijs L4A produces the MBD when it runs

Laurențiu Nicola (CSRO): 10:24: @Guido we store the time series as raster

products and produce e.g. the L4A "markers" as per-parcel aggregated values

Cosmin Udroiu (CS ROMANIA): 10:24: @Guido: both

Laurențiu Nicola (CSRO): 10:25: L4A is a multi-date classification over the entire AOI, so we need to keep a consistent set of features regardless of any missing acquisitions. Averaging is ONLY done for L4A, not for practice monitoring

Laurențiu Nicola (CSRO): 10:26: Re Arrow IPC, they can be memory-mapped so you can also access them without loading the entire file in RAM

DE SN - Okke Gerhard: 10:27: Will there be new Technical Documents regarding the marker Database and all the other new features? Especially for the REST-API?

Laurențiu Nicola (CSRO): 10:27: @Okke they are documented AFAIK

DAA - Justina Vitkute: 10:29: For how many parcels max I can request the marker values at once?

Laurențiu Nicola (CSRO): 10:29: @Anders that's on a case-by-case basis, I suppose. There were some optimizations in recent versions of SNAP which slightly sped up the processing (IIRC); we've also filed bug reports and/or fixed issues in other libraries like OTB and GDAL during the course of the project

Laurențiu Nicola (CSRO): 10:29: @Justina I'm not sure, but if you want to process a lot of them at once (as opposed to an interactive viewer) you can download the marker file and read it

Laurențiu Nicola (CSRO): 10:30: (that's likely better than doing e.g. 1000 API requests for 1000 parcels each)

Kemal Moetz (Ibykus AG): 10:31: When I look at the L4A product and I open the Data_validation_final file to see the different values of the Bands, NDWI, NDVI... I can see NDVI Mean Values in the NDVI columns like XX_2020_03_02_s2_mean_ndvi: 573264. Does this mean that the value of the NDVI is 0,573 at this date or is there another relation to this number?

DE SN - Okke Gerhard: 10:31: @Laurentiu. Thank you. I see there are already parts in the SUM.

DAA - Justina Vitkute: 10:33: Thanks, I was just wondering how easy is it to request values for the set of parcels of my choice :)

Maurizio Laterza: 10:33: Is the system expandable, i.e. is it possible to calculate further Biophysical params?

Laurențiu Nicola (CSRO): 10:35: @Kemal I'm not sure you're supposed to look at that file, but yes, those should be the NDVI values. The markers file would be better-suited for this

BERIAUX Emilie: 10:36: When you look at the increase/decrease of the sigma0 in order to detect agricultural practices (no smoothing), how do you deal with the uneven sigma0 (general increase/decrease of the curve but from one date to another, it can present little jumps)?

Laurențiu Nicola (CSRO): 10:36: @Kemal I'm not sure why you got a value like 573264, they shouldn't be scaled

Laurențiu Nicola (CSRO): 10:36: Also, as a heads-up: that file is written in the Arrow format in 2.0, but the extension is still .csv (that's a bug I've just noticed)

Laurențiu Nicola (CSRO): 10:37: @Justina it depends on how many parcels

you have :-). You can request ask for a couple but for anything non-interactive I would read the file directly instead

Manuel Cañizares (Tragsatec): 10:37: Nowadays, from Tragsatec we are carrying out signal extraction using Sen4CAP and CREODIAS in small part of Madrid region that includes 3 tiles of Sentinel2 for whole year period (S2, S1 and L8). The L3 process has been running for 5 days ago and we don't know how long it will take approximately... Is there any way to predict execution time?

Harijs Ijabs: 10:39: So the Marker Database is dedicated for user interaction? Meaning - other processors do not use this db as a data source?

Maurizio Laterza: 10:40: What happens if some older Sentinel products are in "archive" mode in Open Access Hub? Does the system wait until every needed product has been downloaded?

Laurențiu Nicola (CSRO): 10:40: @Harijs the processors also use them (L4A at least), but they're there mostly for the users who want to access the data

Laurențiu Nicola (CSRO): 10:41: @Maurizio we support some alternate data sources if you can't use Sci-Hub. I'm not sure how retrieving archived products works, even when manually downloading a product

David Nafría (ITACyL-Spain): 10:42: A very important thing that I miss is the cloud mask data for S2 in the markers database.

Maurizio Laterza: 10:42: Thanks @Laurentiu

DE SN - Okke Gerhard: 10:43: @Philippe. Could you share the R-script for the extraction?

Cosmin Udroiu (CS ROMANIA): 10:43: @Manuel: Usually, an L2A tile is

processed into an L3B product between 10 to 20 minutes. Nevertheless, you can increase the parallelism to make them processing faster. If you want, you can write me an email (cosmin@c-s.ro) or post a question on forum on how you can increase the processing parallelism

Laurențiu Nicola (CSRO): 10:43: @David how would that work? The markers are aggregated by parcel

Maurizio Laterza: 10:43: Are sample scripts like this already available?

David Nafría (ITACyL-Spain): 10:44: You need a mark with the presence of clouds on that specific date and parcel

Laurențiu Nicola (CSRO): 10:44: @Maurizio, Okke, I guess they will be made available. Anyway, if you download the markers file you can read it directly (import the arrow library and use read_feather), and you get an R data frame

David Nafría (ITACyL-Spain): 10:44: It could be binary or the count of pixels affected

Laurențiu Nicola (CSRO): 10:45: @David yeah, that could be useful; no, we don't have it :-(

Maurizio Laterza: 10:45: Thank you @Laurentiu

Laurențiu Nicola (CSRO): 10:46: @Maurizio, Okke there's also a Python library which has a conversion function to Pandas

Maurizio Laterza: 10:46: @Laurentiu Very good news!

Laurențiu Nicola (CSRO): 10:47: (and Java, C++ and Rust libraries, and probably others)
Manuel Cañizares (Tragsatec): 10:47: Thank you, I have other question, it is

possible access using external PostgreSQL to CREODIAS VM?

DE SN - Okke Gerhard: 10:47: @ Laurentiu, Thank you!

Laurențiu Nicola (CSRO): 10:47: @Manuel you can set up SSH port forwarding to access Postgres

Kemal Moetz (Ibykus AG): 10:49: Is there a topic/discussion planed about safety issues and measurements (Data Security) using Sen4CAP in Cloud Services like DIAS?

David Nafría (ITACyL-Spain): 10:52: I reply to what I said, to develop more markers based on the API it is desirable to have the actual bands data in MDB1

David Nafría (ITACyL-Spain): 10:53: NO bands in MDB1

Maurizio Laterza: 11:14: Maybe it's my fault, but I can't remember what gray, blue and red strips on the graph stands for. Please, can you describe them? Thank you

Guido Lemoine (EC-JRC): 11:15: check the weather as well...
Guido Lemoine (EC-JRC): 11:18: The real significant coherence drop is just before the tillage data
Guido Lemoine (EC-JRC): 11:19: you found one that is more likely due to rainfall

David Nafría (ITACyL-Spain): 11:23: H2020-Sensagri proposed a methodology to overcome the rain issue in the backscattering

Anders Munck (FI): 11:23: Regarding Guidos' remark: it would be very interesting to see how the incorporation of weather data would affect decisions...
Lubos Kucera: 11:25: there are four coherence values forming consistent

drop, I do not expect this is the effect of rainfall, also the backscatter does not indicate the rainfall, generally coherence is much less affected by rainfall comparing to backscatter

Guido Lemoine (EC-JRC): 11:27: backscatter goes up during your little coherence dip -> rainfall. You overlooked the far more significant drop shortly before the tillage
Guido Lemoine (EC-JRC): 11:28: an optical bare soil indicator would have helped as well

David Nafría (ITACyL-Spain): 11:31: Sensagri did not use coherence, just backscattering and optical bare soil indicator

Guido Lemoine (EC-JRC): 11:31: alternative explanation would be that they break up the straw or something
Guido Lemoine (EC-JRC): 11:32: but the more significant one is the last
Guido Lemoine (EC-JRC): 11:33: splitting asc/desc could also help

Diane Heymans: 11:33: @Guido Did you publish on how the backscatter respond to the weather and how to deal with it?
Guido Lemoine (EC-JRC): 11:34: long time ago. But soil moisture and frost/thaw effects in SAR are well known.
Guido Lemoine (EC-JRC): 11:35: the real issue is always to separate them from vegetation and tillage effects

Jan Musial (IGiK): 11:42: How to speed-up Sen4CAP processing using more workers/parallel threads? This is particularly important for CREODIAS, where we could rent a powerful VM but the Sen4CAP was not able to use all of its resources.

SE Bastian Berlin: 11:42: A lot of the data should fall under Inspire. Farmer ID could be changed based on internal code

Adam Knaze (PwC): 11:43: The shapes are usually public anyway, aren't

they?

Guido Lemoine (EC-JRC): 11:43: All DIAS platforms provide a full private environment. Putting your data there does not mean that the rest of the cloud sees it
Guido Lemoine (EC-JRC): 11:44: If that was not the case, there would be no cloud market
Guido Lemoine (EC-JRC): 11:46: Clouds are FULLY private, unless YOU decide to open up

Laurențiu Nicola (CSRO): 11:47: @Jan the amount of parallelization we can do depends on what we are currently processing. For example, S1 pre-processing can be run in parallel, but uses a very large amount of memory and you need to manually configure that. S2 pre-processing runs sequentially inside a tile, so if you use MAJA and have a single tile, you won't have much luck. Most other processors parallelize well, with the exception of L4A where the S2 feature extraction works less than great.

Laurențiu Nicola (CSRO): 11:48: @Guido that's arguable from an IT perspective since without disk and in-memory encryption the cloud provider will probably be able to access the data

Guido Lemoine (EC-JRC): 11:50: Yes, maybe a suicidal cloud provider would do that. Read the SLA

Jan Musial (IGiK): 11:50: Thank you @Nicola. So, in this respect are there are Sen4CAP setting to tune? Regarding RAM if you have a VM with >128 GB RAM you could run few S-1 processing in parallel.

Laurențiu Nicola (CSRO): 11:50: With the recently-published CPU bugs (Spectre, Meltdown and friends), you can even exfiltrate a small amount of private data from one VM to another. These have been mostly patched, but it's harder to check.
Laurențiu Nicola (CSRO): 11:51: @Guido agreed, but that's a legal, not

IT/security matter.

Laurențiu Nicola (CSRO): 11:51: @Jan I think it's around 10-20 GB per product pair

Cosmin Udroiu (CS ROMANIA): 11:51: @Jan For the MDB and L3B extraction you can increase also the parallelism by changing the values of qoslai or qoss4cmdb1 in Slurm

DE SN - Okke Gerhard: 11:52: I think the problem is kind of a German problem, as privacy and data security is a hot topic. Therefore, we always have to be aware of that and I think a lot of times, people in charge to not understand cloud infrastructure and geodata, therefore they are really conservative with that.

Grega Milcinski (SIN): 11:53: You might want to make use of CODE-DE then, which is a German cloud very much alike DIAS. It should have all the benefits of the DIAS and the security requirement of Germany.

Grega Milcinski (SIN): 11:53: https://code-de.org/

Harijs Ijabs: 11:54: Thank you very much, Philippe!

DE SN - Okke Gerhard: 11:55: @ Grega Thank you, we are aware of that and are already investigating it as it seems to match the German requirements.

Felix Lobert: 11:58: We already tried to install Sen4CAP on a CODE-DE VM but are facing issues configuring to use the local satellite data. I already posted a question in the forum on this, but didn't receive a reply on it yet unfortunately.

Harijs Ijabs: 11:58: A bit related question to my previous: If only some parcels have changed in my GSAA, do I have to reupload the entire dataset

or can I just replace the modified parcels?

Joeri B: 11:58: Is there a form of user control in the web interface? Can we have users who can only view the map/classification/graphs or can anyone edit the settings?

Laurențiu Nicola (CSRO): 11:58: @Harijs you can do both, there's an option in the UI IIRC

Laurențiu Nicola (CSRO): 11:59: @Joeri there is support for authentication and roles, but I'm not sure what exactly you can enable or disable

Marcin Bialecki (CREODIAS): 11:59: Greetings! :) Thank you for introduction. If anyone would like to ask anything on the Sen4CAP on CREODIAS feel free to ask.

Harijs Ijabs: 11:59: OK, thank you!

Cosmin Udroiu (CS ROMANIA): 12:00: @Joeri you can limit the users access for some sites for example or only for user operations (not admin)

Laurențiu Nicola (CSRO): 12:00: @Felix we have to special-case every DIAS because they don't usually have the same folder structure

Joeri B: 12:00: Ok, thank you

LUCAU-DANILA Cozmin: 12:04: Thank you very much !
Very nice and open system to be adapted to most of requirements for preparing the AMS! With a reactive « helpdesk »!
Waiting for V3!! What’s next?

Piotr WOJDA: 12:04: Thank you all, it was excellent and very interesting! Impressive work.
Guido Lemoine (EC-JRC): 12:04: By all!

Guido Lemoine (EC-JRC): 12:04: bye

SE Bastian Berlin: 12:05: Thank you everyone and well done with the database addition in version 2.

Cosmin Udroiu (CS ROMANIA): 12:16: @Felix I answered to your question on forum but I think we would need some additional information for your issue. Also, for the general conversation we can continue on forum but giving us access to the machine will be quicker to solve the issue