Project Owner	Guy Janssen
Science Contributors	Robert Waterworth Werner Kurz Malcolm Francis Robert de Ligt Max Fellows Stephen Kull Benjamin Hudson Jean-Francois Carle Jean Girard Guillaume Cyr Donna Lee
Science Editors	To be decided
Science Review Panel	To be decided
Science Approval	Werner Kurz
Code maintainers	To be decided
Code Coach	To be decided
Documentation Lead	To be decided
Release Approval	Werner Kurz

PLEASE READ THIS BEFORE YOU CONTINUE:
This is an open source document. Everybody is welcome to contribute. To allow us to be efficient we have only a few rules:

- All contributors use comments to ask questions or make substantive and structural suggestions. Always suggest solutions in your comments not problems.
- Only Owners, Maintainers and Editors can resolve comments.
- All contributors make edits in "suggesting mode" (even if you are an owner, maintainer or editor, so the other maintainers or editors can cross-check and accept your contribution.)
- Write suggested edits directly into the text in 'suggesting mode' do NOT use comments to suggest edits
- Only Owners or Editors can approve edits
- Spelling is UK English
- Referencing Style is Harvard (author-date) type. Tools such as Zotero may be useful.

More info about how the moja global community works on projects can be found here.

Projections of GHG emissions and removals using FLINT

v1.1.0

This design document was made possible thanks to the financial support from the **Secretariat of the United Nations Framework Convention on Climate Change**.

Abstract:

This document explores which approaches to projecting GHG emissions estimates serve FLINT users best. The aim is to foster agreement on the approaches to projections, how they can be used and any potential enhancements to FLINT required to implement them. The document was drafted with financial support from the Secretariat of the United Nations Framework Convention on Climate Change.

Table of Contents:

Executive Summary:

One of the uses of the Full Lands Integration Tool (FLINT) is to estimate future greenhouse gas (GHG) emissions and removals based on complex policy scenarios. FLINT can already project emission estimates. This document identifies how FLINT's projection capability can be further improved.

The scope for improvement is limited by four key technical design features: First, FLINT is a model-based approach, so projections are possible when model inputs can be projected. Second, as a minimum, initial land cover conditions and either a time series of activity data or a land cover change time series with attribution of changes from which activity data can be inferred, are required inputs for any FLINT configuration. Third, FLINT models carbon stocks and fluxes for every timestep of a spatial unit (e.g. pixel) before moving to the next spatial unit. This results in higher accuracy at constant computer resources, but it makes it harder to use changes in the landscape from the previous time step to inform events in the next time step. Fourth, FLINT generates spatially-explicit results which are crucial for some projections but are of marginal importance in other situations. For the latter situation, aggregating data and using a compatible spatially-referenced model might be more effective than attempting to create spatially-explicit projections.

Due to its design features, FLINT has strong capabilities for projections but there are exceptions: mainly, when stocks or fluxes estimated at the previous time step are used in the

calculations for the activity data in next time step or when spatially-explicit outputs do not add sufficient value to justify the greater effort required for spatially explicit projections.

The following key lessons are learned from existing projection approaches: Consistency between estimates of past (reported) GHG emissions and removals and future estimations is important and can be achieved by using the same tool (e.g. FLINT) for both. Hence, FLINT is used for projections when the effort to generate spatially explicit input data remains smaller than the value of spatially explicit outputs. Often input data are already generated by a separate tool (e.g. harvest planners). Special tools to generate input, separate from FLINT, create flexibility and simplicity. Even if some tasks are duplicated between the input generation tool and FLINT, the reduced complexity will result in resource efficiency. In most cases where carbon stocks or fluxes from the previous time step are used as input data for the next time step, it is possible to use proxies like age or volume.

FLINT can be used for the four typical projection types but some additional tools are required: (1) Baseline projections require, in addition to FLINT, algorithms to predict key inputs and a map builder to translate the algorithm results into spatially explicit maps. (2) Land Management Policy Scenarios require the same input projection tool. (3) Emissions Optimizations require the same input projection tool and two or possibly three additional tools: a results comparator, an inputs instructor and possibly a spatial unit selector. (4) Co-benefit Optimizations require the same four tools but each one would need to be upgraded considerably.

The proposed design avoids major changes to FLINT and focuses on the development of the four additional tools: (1) The Inputs Generator is a raster calculator that creates layers of input data using algorithms. It only requires high level instructions to generate future input layers but it needs tools to allocate disturbances spatially when the high level instructions do not result in specific locations. (2) The Inputs Instructor varies high level instructions to the Inputs Generator within a predefined range. The variation is based on the feedback received from the third tool, i.e. the Results Comparator. The Input Instructor design can build on the inputs generator for the Monte Carlo Uncertainty Module. (3)The Results Comparator compares FLINT results and possibly other outputs with an optimization objective and produces a boolean or a measured output that then informs the selection of the inputs for the next iteration of simulations. The Results Comparator can build on the Reporting Tool. (4) The Spatial Unit Selector increases efficiency by selecting a representative sample of pixels to drive the optimization process. The Monte Carlo Uncertainty Module already has a tool with a similar function.

The conclusion is that the projections capability of FLINT can be upgraded without making major changes to the FLINT code. Instead, four separate tools need to be coded building on existing tools. Keeping the four tools separate will make it easier to code and define projections. While coding the tools is not too complex, developing algorithms to project land cover changes (e.g. the rate and spatial allocation of future deforestation events) requires advanced science inputs.

This design document was made possible thanks to the financial support from the Secretariat of the United Nations Framework Convention on Climate Change.

Background:

The Full Lands Integration Tool (FLINT) is a platform to support Measurement, Reporting and Verification (MRV) of greenhouse gas (GHG) emissions and removals estimates. Users (e.g. countries) can build on this platform to implement and operate their (national) systems to estimate emissions (and other metrics) for the land sector. FLINT integrates data from remote sensing, ground observations and other sources to estimate fluxes and stocks of greenhouse gasses in different pools consistent with the guidance from the Intergovernmental Panel on Climate Change (IPCC).

moja global is an open source collaboration under the Linux Foundation that allows users of the FLINT to work together to continuously improve the tool in line with the needs of the users. Moja global was set up and continues to be run by the users of FLINT. All FLINT users can request a seat on the Strategy Board of moja global.

Users have indicated that the ability to make projections is an essential feature for FLINT. While it is already possible to run certain types of projections with the FLINT and models built on this platform the development of such projections is constrained by the need for spatially-explicit input data about future events (disturbances, management, land-use changes). There is a desire to make it easier to apply projection methods that are based on rules and algorithms, or that can use spatially-explicit projections generated by other models, such as harvest scheduling tools .

This document evaluates different approaches and identifies the steps and tests necessary to improve FLINT's projections capability.

First, some features of FLINT's technical design are highlighted as these decisions might limit or enhance the scope for projection approaches. As a next step, existing systems to make projections are analysed and compatibility with FLINT's technical design is tested. This analysis will result in an overview of which approaches can or cannot work with FLINT. Next we evaluated which of the approaches that can work with FLINT are also capable of meeting the need for projections as communicated by potential users. Finally, the potential approaches are translated into a high level technical design.

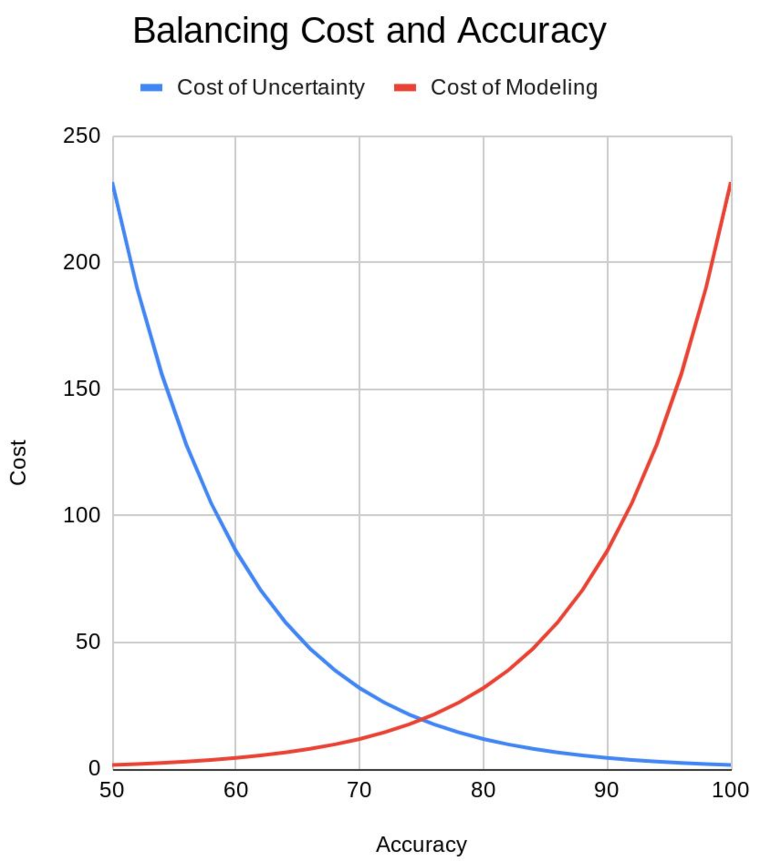
FLINT's Technical Design Features

Modelling Approach Versus an Empirical Approach

There are two main approaches to estimating GHG emissions and removals by the landscape: Empirical measurements on statistically relevant locations and modelling real life processes through mathematical relationships based on measurements. These are typically described as stock-difference and gain loss methods and are described in detail in the IPCC Guidelines. FLINT uses a modelling approach, which allows for projections as long as spatially-explicit inputs that define the type and location of future events are provided.

An observational or empirical approach to estimating GHG emissions that relies only on field observations has limited predictive power. The stock-difference method estimates the carbon stock at two points in time and the difference is the net emissions or removals over that time. Since measuring these shifts on every strip of vegetation in the reporting area would be cost prohibitive and physically impossible, a probabilistic sample-based method is applied where measurement plots are allocated across the landscape. Statistics theory enables one to use these plot measurements to estimate carbon stocks for the landscape with a known error margin. It is possible to use these measurement plots to find trends and correlations with factors such as ecological zones, species, management and climate. However, these observations of past stock changes do not provide a predictive capability for future changes. For example, while it may be possible to establish that average temperature has an influence on some carbon pools, the actual process through which temperature exercises its influence remains unclear. As a result it is not clear whether the correlations will hold when the correlation is applied to a place or a point in time where the circumstances have changed. The method also only provides for estimates of net change rather than being able to report emissions and removals separately. While the end results are the same, for projections is it often important to be able to determine the emissions and removals separately.

The gain-loss method, used by FLINT, relies on an inventory of the initial conditions and then applies gains (growth) and losses (e.g. from disturbances) to derive annual estimates of stock changes, emissions and removals. Extension of the time series of gains and losses allows for a seamless transition between past and projected future estimates. The approach can use numerical equations, models or yield tables to represent how carbon and other GHG components are shifting between the atmosphere and carbon pools based on factors such as management, disturbances, age, vegetation and climate. There are a wide range of models available of varying complexity. Typically, the more complex the model the more likely it will be able to represent these management and ecological processes, but



at a cost of increased need for high-quality scientific data for calibration and validation. Without these data the results can have high uncertainties. Simpler models may require less data but can't easily represent more complex processes.

The choice of model depends on the available data, the national circumstances and on the policy and reporting needs of the user. All models still require measurements to calibrate and validate them. However, rather than use these measurements directly, the models are developed to represent the required process. Models still need inputs to calculate their outputs. For example, as a minimum a model might need to know how much time has elapsed to calculate the fluxes of carbon. Additional variables might increase the accuracy of the model. For example, vegetation type or climate data might be used as a variable to better calculate fluxes of carbon. But these variables will need to be estimated as well. So, while adding variables may increase accuracy, it will also increase cost. Depending on the application, accuracy may be of greater or lesser importance. For example, for pools that are subject to large fluxes it may be desirable to be more accurate than for smaller pools. Further, some policies and programs, in particular those looking at results based payments, may discount the number of available units based on uncertainty. In such cases increasing accuracy may increase available income. Cost effective systems find the best compromise between the cost of uncertainty and the cost of increasing accuracy.

A model based approach provides a consistent basis for estimating past and future estimations. A model is a mathematical relationship between the input variables and the fluxes and pools generated as output. The mathematics link represents processes that can be observed on the ground. So a specific change in input values will result in a predictable change of observable outputs. As long as the input variables can be observed or estimated, it is possible to generate values for outputs whether the input values are from past observations or planned future events. So generating projections follows precisely the same process as calculating past estimates, one only needs to provide the input values for future situations. The mathematical relationship between input and output continues to represent the underlying processes even in the future, so the model continues to generate outputs with the same accuracy when input values for future events are fed to the model.

Availability of projected data about future events determines the models that can be used and thus the accuracy of projected emissions estimates. FLINT is a platform that uses models to calculate emission and removal estimates. This is an advantage as it allows for the estimates of future emissions and removals as long as the input variables are known. The models used to generate future estimates can be chosen on the basis of the available input data. The accuracy of the results can be increased by replacing the currently used model with a more accurate model. As with estimations for past events, projections will have to find the optimum balance between accuracy and cost. The most appropriate compromise between accuracy and cost might change between locations and over time.

FLINT Projection Method is Determined by the Ability to Generate Key Input Data

Selecting the best way to run projections using a FLINT-based system, boils down to designing the best way to generate input data for the future. FLINT is an integration tool, i.e. a tool that combines data from different sources and in different formats to calculate its outputs. Configuring FLINT starts with a review of the user needs, assessing the available data and models and moves on to select (or develop) a module that can combine the available data in the FLINT. Accuracy can be improved over time as more or better data and models become available. Using these data and models, the FLINT can generate projections if the cause (e.g., harvest, disturbances, afforestation) and locations of land cover changes are known such that future land cover can be calculated. While this type of data is available from past observations, it will have to be generated for future events. The design of a projection method for a FLINT-based system will be primarily determined by the ease with which the input data can be generated for the future.

FLINT is an integration tool, which means that it combines various types of data to estimate land sector greenhouse gas emissions and removals. A wide range of data can be integrated including data collected using remote sensing, forest inventories, research sites, soils maps, vegetation maps and climate information, to name just a few. Few emission estimation systems have been designed to allow flexible data inputs. Most carbon estimation methods use rigid formulas or methods that require specific data inputs. This is particularly unfortunate given both increasing policy and reporting needs and the proliferation of large amounts of data collected through earth observation systems, internet of things technologies, and ground data combined with exponentially increasing data storage capacity and data exchange speeds. As a result, even if a user does not have any data available yet, it is possible to generate emission estimations (albeit with lower accuracy) solely by using currently publicly available global datasets. But a framework like FLINT is needed to flexibly combine the data from global, national and regional sources in support of MRV systems.

The available data inform the configuration options for a FLINT-based system. The different types of data are combined in a way that maximises the utility and accuracy of the outputs. While it seems logical to develop a model to represent reality and then collect data to feed the model, the configuration of a FLINT-based system starts from a review of the policy and reporting needs and the available data and then moves to select or build a model that combines the available data in a way that yields the most useful and accurate results possible. By starting from the available data, a FLINT-based system can produce initial results very quickly. The initial results may not be perceived as accurate enough due to the quality of the input data, but having a functioning system quickly has two important advantages: First, even inaccurate results are useful for policy makers where little or no other information is available and these initial results allow policy makers to understand that their emissions estimation system can generate policy information. This early feedback saves time and resources as corrections can be applied before large investments have been made. Second, running the system identifies where the opportunities for improvements are. Moreover, the possible improvements can be prioritized by their required effort and improvement effect on the emissions estimates. Since the system is operational i.e. embedded within their institutions, the potential improvements are not only related to the science behind the system but can also include capacity building (training), administration, budgets, logistics, data, and infrastructure.

The initial land cover and the causes of land cover change are the minimum data inputs for a FLINT-based system to account for forest carbon stocks and carbon stock changes. Initial land cover (e.g. from a forest inventory or from remote sensing) is an essential input to understand what is growing on the land at the start of the simulation so that the correct models and parameters can be applied to estimate emissions and removals for each land cover type (forest, grassland, cropland, etc.) and carbon pool. Publicly available datasets are sufficient to run the first version of a FLINT configuration, but most have significant limitations in coverage, accuracy and usefulness (e.g. often they do not track forest types, regrowth or other measures). Over time, accuracy can be improved by developing user specific input maps, for example dividing the landscape into additional vegetation types, or productivity classes of the same vegetation type, or by better tracking different types of change. The causes of land cover change are the second essential input. The causes are necessary to select the disturbance model. Each type of disturbance has a very different effect on various carbon pools and fluxes. Publicly available information is sufficient to build basic disturbance models using broad assumptions but accuracy and usefulness can be vastly improved by differentiating between ever more specific disturbances and by using additional variables to make the calculations of carbon shifts more precise.

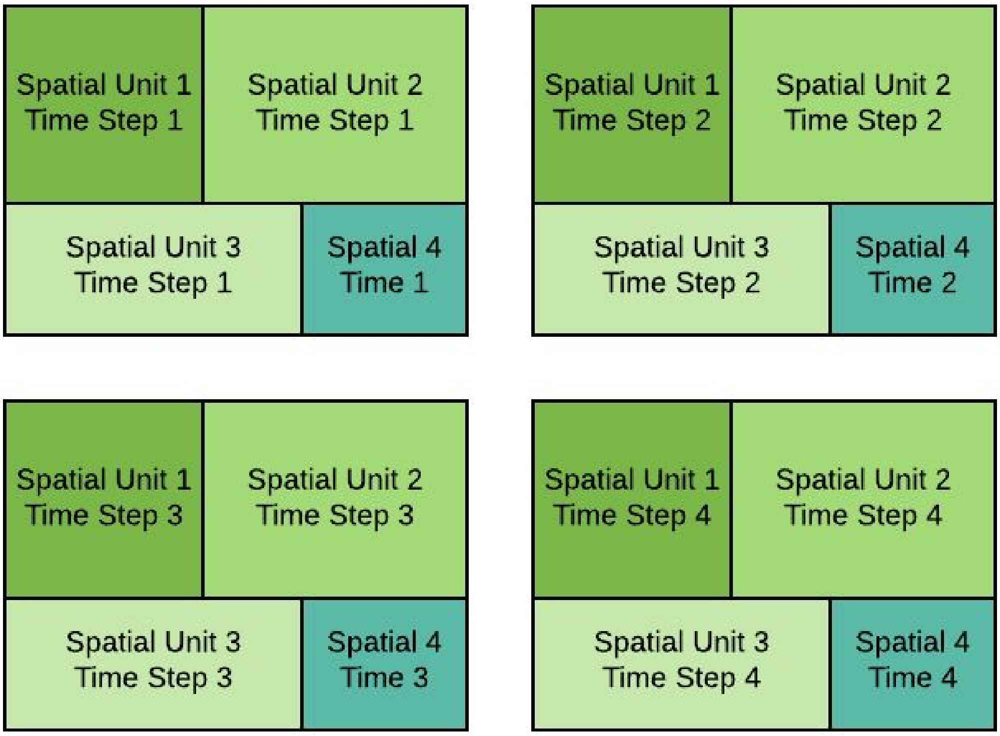
The best approach to run projections with a FLINT-based system is determined by the way key data inputs can be generated. Projections can be generated if the same basic inputs are available for future as for past estimations. So at a minimum, initial land cover and the cause of land cover change are required. As indicated above, the growth and disturbance models remain valid whether they are applied to past or future data. These models only need key data inputs to generate carbon estimates. However, while there are many datasets available for past observations, projections of these data are scarce. More importantly, assessing the impacts of a policy or management plan will require the generation of land cover change expected as a result of such a plan. These datasets cannot be readily available. So systems need to be designed to generate these datasets. There are various approaches possible as will be discussed below, at this point the key conclusion is that FLINT-based systems can produce projections without change to the system as long as the input data are available. Therefore the discussion about approaches to projections, should focus on how the input data can best be generated.

Time and Space: Loops in a Loop When Calculating Carbon Fluxes

The FLINT's sequencing of calculations allows for increased accuracy with the same computer resources as it first calculates all time steps (inner loop) for one spatial unit (outer loop) and then moves on to calculate all time steps for the next spatial unit. The advantages of this approach are that it allows for very fast calculation of numerous spatial units, that these calculations can be distributed (i.e. used in computing clusters or cloud computing) and that the approach can be scaled to very large landscapes with hundreds of millions of pixels. The disadvantage of this approach is that landscape-wide calculations of the previous time step are not available for calculations in the next time step. This prevents the use of rule-based algorithms or models that use the condition of the landscape in one year to project the conditions in the next year. This document suggests some solutions to overcome this disadvantage.

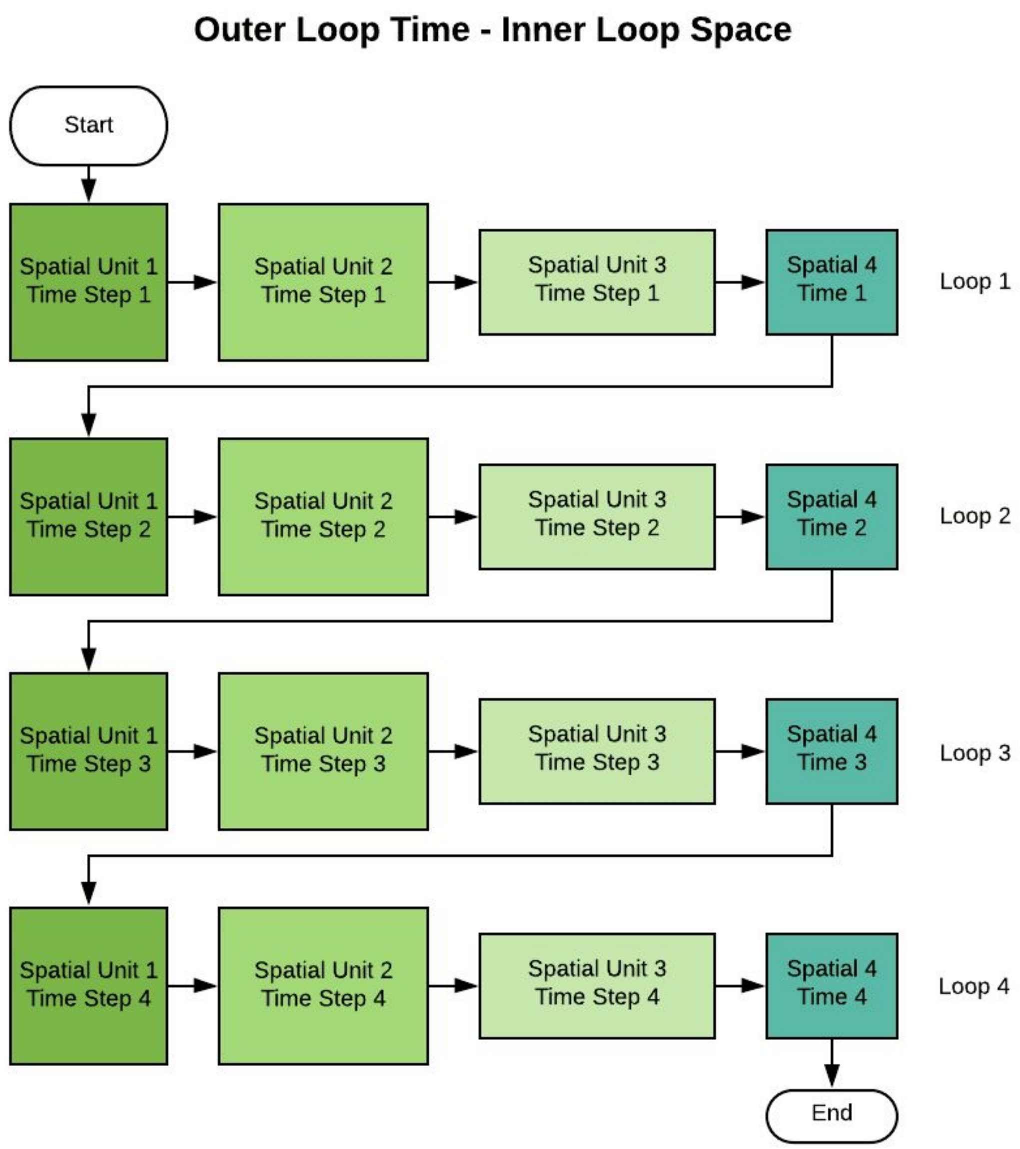
Carbon estimation tools will calculate all time steps for all spatial units but the order in which calculations are made can differ between tools. Carbon stocks and fluxes are typically calculated for a spatial unit (e.g. pixel, stand, inventory record, etc.) with similar features (e.g. vegetation, soil, eco-climatological zone, etc.) for several time steps (e.g. every month, every year, or for several points in time for which observations are available.)
There are two main approaches to the sequencing of the calculations: Either space is the inner loop and time is the outer loop or the other way around. In other words, either the fluxes and stocks are calculated first for all the spatial units for one time step and then for the next time step or the fluxes and stocks are calculated for all the time steps of one spatial unit and then for the next spatial unit.

Below the following simplified landscape will be used as an example. Imagine that a forest is divided into 4 spatial units for which change has been observed for 4 time steps. To simplify the visualisation the representation is approximate in two ways: First, spatial units have different sizes. While this provides an impression of the combined spatial units amounting to a varied landscape, the actual spatial units in a FLINT-based system are of equal size (e.g. 1ha, or 30x30m). Second, time <u>steps</u> are used as opposed to observations in time. A time step refers to the changes that take place between two observations in time. So for 4 time steps, 5 observations in time are required: Step 1 is the change between observation time 0 (the initial condition) and observation time 1. Step 2 is the change between observation time 1 and observation time 2. etc.



Outer Loop Time - Inner Loop Space

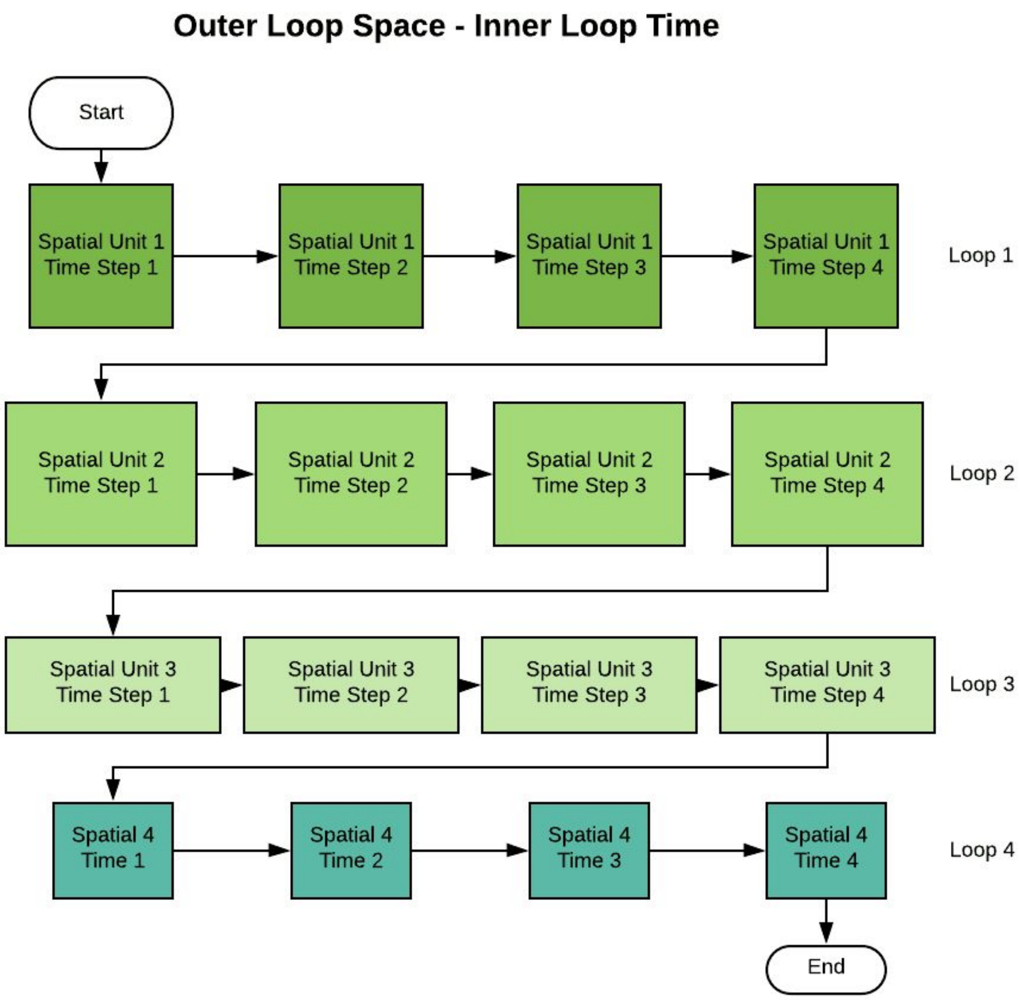
The sequence of calculations when time is the outer loop and space is the inner loop, would be that for each land unit time step 1 is calculated first, followed by time step 2 for each unit, etc. This can be visualised as follows:



The advantage of this approach is that at the end of each loop, the state of the whole landscape has been calculated and potentially this output can be used as input for the next time step. Considering that the essential inputs are land cover over time and the cause of land cover change (see above) the changes to the landscape in each time step are important inputs for the next time step. For example, the user wants to apply a harvest scenario on the inventory above with 4 spatial units (SPU) and 4 time steps (TS) based on the rule that the spatial unit with the highest carbon stock will be harvested first. Imagine that each SPU has a different sequestration and decomposition speed. Imagine that based on the initial values, SPU 4 has the highest carbon content and will be harvested during the first loop1. At the end of the first loop, the carbon stocks for each spatial unit have been updated. SPU 4 probably no longer has the highest carbon stock after the harvest. Imagine SPU1 now has the highest carbon stock, so SPU1 will be harvested in loop 2 and all the carbon values are updated. During each loop, the values generated in the previous loop are used to decide on the next set of disturbances applied to the landscape. This is only possible if the inner loop is space and the outer loop is time.

Outer Loop Space - Inner Loop Time

The sequence of calculations would follow the steps below when space is the outer loop and time is the inner loop. In this case, all the changes for all time steps would be calculated for a spatial unit and only then would the process move on to calculate all the time steps for the next spatial unit. This can be visualised as follows:



Time as the inner loop is possible as current carbon stocks have an influence on fluxes and stocks over time but not (so much) on fluxes and stocks in other spatial units. Vegetation growth and disturbances can be modeled using variables having an effect on the spatial unit (SPU) being modeled. Rarely are carbon stocks and fluxes influenced by what happens in adjacent SPUs, i.e. if temperature or soil conditions in the adjacent SPU change dramatically, the vegetation in the observed SPU continues to grow at the same rate. An exception could be that adjacent carbon stocks might have an influence on the type of fire that develops in the area and subsequently engulfs the SPU observed. But even in this exceptional situation, the changes in stocks and fluxes on the observed SPU are not influenced by the stocks and fluxes in the adjacent SPUs. Rather the type of disturbance (contagion) and thus the type of disturbance matrix is influenced by the stocks in other SPUs in the area. Since, spatial units do not have local influence (but do have an influence over time) it is possible to calculate the stocks and fluxes for 1 SPU over the complete observed time period before moving to the next SPU without loss of accuracy.

The advantage of this approach is that increased spatial and temporal detail can be achieved with the same computer resources. This can also help improve accuracy by capturing temporal and spatial changes more precisely. When using high quality data and more advanced models, dividing the landscape into smaller spatial units and adding time steps will typically make estimates more precise because the boundaries and time of the changes observed in the real world are better represented by the input data. But more observations will increase the computing power required to process the calculations. Speed and memory are the two most important determinants of the cost of computer resources. Memory can be reduced by reducing the number of variables necessary to execute calculations. Speed can be increased by increasing the number of processing units, typically running numerous processes in parallel. Since stocks and fluxes have an influence over time and not space (see para above) and since the number of time steps are many times smaller than the number of SPUs, using time as the inner loop reduces the computer resources required. The influence of current stocks over time and not space, has an influence on computer memory because only the pools and variables for the current SPU have to be kept in memory to process all the future stocks and fluxes. Even if all variables of that SPU for a complete time series are loaded into memory, memory size can be many times smaller than loading the variables for all SPUs for one time step. This allows for very small spatial units to be used (theoretically down to meters), corresponding to the size of the pixels of satellite images. So for an average country, the number of spatial units can easily reach into the billions. The time steps will always be several factors smaller as even a daily observation for twenty years would only result in about 7300 observations. Moreover, since stocks and fluxes only have an influence over time, two SPUs can be calculated independently. So theoretically, each SPU can be calculated by a different computer processor. The potential to process many SPUs in parallel reduces the required computer resources considerably.

FLINT Generates Spatially Explicit Results

FLINT should only be used for projections if spatially explicit projections have added value. FLINT-based systems generate spatially explicit carbon stocks and fluxes for projections as well as for past emissions estimates. The added value of spatially explicit projections will determine whether FLINT should be used for projections. The added value should outweigh the added effort: i.e. generating spatially explicit input. This can vary from application to application. For applications that profit from spatially explicit emissions estimations, tools can be developed to reduce the effort to develop spatially explicit inputs. If spatially explicit results do not weigh up against the generation of spatially explicit inputs, FLINT should probably not be used for projections. For these situations, aggregating the spatially-explicit data and using an existing non-spatially explicit system (like CBM-CFS3) is more suitable for projections.

Spatially explicit projections maintain all information but aggregated data simplify projections. FLINT-based systems generate spatially explicit carbon stocks and fluxes for projections exactly in the same way as they do for past emissions estimates. FLINT is running each SPU separately through time. As indicated above, FLINT-based projections require that the initial land cover and future disturbances (including management activities) are provided as spatially explicit data. Various other variables might be available for each SPU. All the information available for each SPU at the point of input into a FLINT-based system, will be maintained into the output generated. No information is lost. Systems that are not spatially explicit aggregate spatial information into groups of spatial units with the same relevant features (e.g. species, age, climatological zone, administrative area, etc.). A feature will be relevant if it will be used by the emissions estimation system (e.g. to decide whether future disturbances occur or to feed a calculation module). The more features are relevant, the more categories will be required. But by grouping spatial units some specificity is lost (at a minimum the exact location).

Spatially-explicit approaches increase transparency and verifiability. This is of particular importance where investments are made into forest-based climate change mitigation, such as afforestation, rehabilitation or avoided deforestation. By tracking the location of the investment, the outcomes, e.g. the magnitude of the carbon stocks after afforestation or confirmation that forests remain in place, can be verified by third parties using readily available data. This can serve as the basis for results-based payments.

The added value of spatially explicit projections has to outweigh the extra effort to develop spatially explicit inputs. There are several advantages of spatially explicit systems: First, all (soft) information remains available in the output data. For example, predicting deforestation will likely be based on a combination of geographic specific (i.e. hard) observations: e.g. roads, hamlets, agriculture, past deforestation. This type of information is lost when information is aggregated. Locations can also create links to information not yet available when projections were calculated, e.g. when spatially explicit results are presented in public consultations, community knowledge might be made explicit through the maps. The second advantage of spatially explicit projections is consistency of the estimates over time. Because the same system is used for past and future estimates of stocks and fluxes, the systemic accuracy remains the same. This will make the outputs more comparable. When inputs are aggregated, the additional processing can result in different estimates. At the least, the additional complexity will require expert review and careful quality control to ensure the highest possible accuracy is achieved. Thirdly, by generating outputs spatially it is possible to use these in other analysis systems more easily, to relate the projected outputs to other spatial data. Finally, maps are a powerful communication tool that can more easily engage policy makers and communities because they have a powerful visual attraction. Moreover, maps are a very accessible knowledge format. There are other advantages of spatially explicit systems but there are also additional efforts related to generating maps of projected land-cover time series and disturbances. The extra effort is related to determining the location not just the amount of future events. Amounts need to be generated for any type of projections: These can be planned disturbances (e.g. forest management plans) or events with a probability like fires, insect attacks, diseases, and weather. The quantities for both can be predicted (albeit within a range of probability based on scientific research for the random events). The extra effort is required to turn this quantitative knowledge into a map. This requires additional processing and additional data collection that need to occur outside the FLINT.

The balance between added value and extra effort of spatially explicit projections can differ considerably between applications. The example above, i.e. predicting deforestation on the basis of features in the landscape (e.g. roads, hamlets, agriculture, past deforestation) is only possible if executed in a spatially explicit manner. Community consultations on planned forest management are likely to profit from spatially explicit projections as their use as an effective communication tool, will most likely outweigh the additional effort to generate maps. Projecting the carbon impact of forest management might profit from spatially explicit projections, but not in all cases. Planned events, e.g. harvest of 6000 ha of 50 year old spruce in British Columbia, cannot be predicted and mapped out without additional information like maximum size of harvest plot, minimum distance between harvest plots, maximum slope and accessibility of the harvest plot, etc. Even with all this additional information, it might not be possible to uniquely determine the location where the planned event will take place. If this information is readily available, the advantages of spatially explicit projections might outweigh the additional effort. If this information needs to be collected specifically to generate the spatially explicit projections, the extra effort is likely to outweigh the advantages. Finally, mapping out probabilistic events might have limited added value as the location might be selected with some randomness and multiple maps (i.e. spatial configuration of outcomes) can have equal probabilities. This is hard to visualise on a map and might thus create confusion rather than added information (e.g. probability of forest fires represented on a map does not mean that the marked area is really going to burn down in the future.)

Tools can be developed to reduce the additional processing effort required to make inputs spatially explicit. As indicated above, generating the spatially explicit inputs requires additional data and additional processing. Tools can be developed to reduce the effort related to determining where future disturbances will take place and structure the information in a FLINT-readable format. The design of such tools will be further explored in this paper.

If spatially explicit results do not weigh up against the generation of spatially explicit inputs, FLINT-compatible aggregation systems might provide a solution. FLINT should probably not be used for these projections as FLINT cannot be used for aggregation. Theoretically it would be possible to recode FLINT to enable emissions estimations of aggregated data but redesigning FLINT would probably undo FLINT's key advantages. This is caused by the nature of aggregation: SPUs are grouped based on features that are relevant for the estimation system to decide whether a disturbance (harvest, thinning, etc.) should be applied or not. Whether the disturbance will actually be applied and at what point in time, will depend on the amount of land that will be disturbed at a particular point in time. This, by definition, means that an aggregated group of SPUs will be divided at a later stage into two or more different aggregated groups, i.e. a group to which disturbances have already been applied and a group to which the disturbance has not yet been applied. This would require larger amounts of memory and additional computational steps reducing the computer resource advantage FLINT has. Completing a full time series of changes before moving to the next SPU, requires the whole SPU to experience the same changes at each time step. It is possible to apply a change to a section of the SPU at each time step but it is not possible to divide the SPU into two parts that will experience different disturbances and growth. When an SPU needs to be split into two or more new SPUs, the inner loop of the FLINT (i.e. time) turns into two or more loops potentially at every time step. Each open loop has to be completed before FLINT can move to the next SPU. So larger amounts of memory and more complex computation are unavoidable, except if SPUs are aggregated in such a way that all SPUs experience the same changes at the same time. While it is possible to develop tools that can generate such aggregated groups of SPUs, these tools would have to run through the landscape for each time step, i.e. run space as the inner loop before moving to the next time step. So if data for projections are going to be aggregated, then switching to a compatible system that runs space as the inner loop might be advisable. This transition would only require a much simpler tool to aggregate the last historical time step generated by FLINT into groups of SPUs with relevant features on the condition that a complementary system running space as the inner loop already exists. For a FLINT-based system like GCBM, the CBM-CFS3 would be the logical complementary system as it is using the same modules for growth and disturbances. If no complementary system is available, generating spatially explicit inputs might be the easiest option even if the spatially explicit outputs do not add much value. Hence, this paper will not explore the design of tools for data aggregation except for a transition from GCBM to CBM-CFS3.

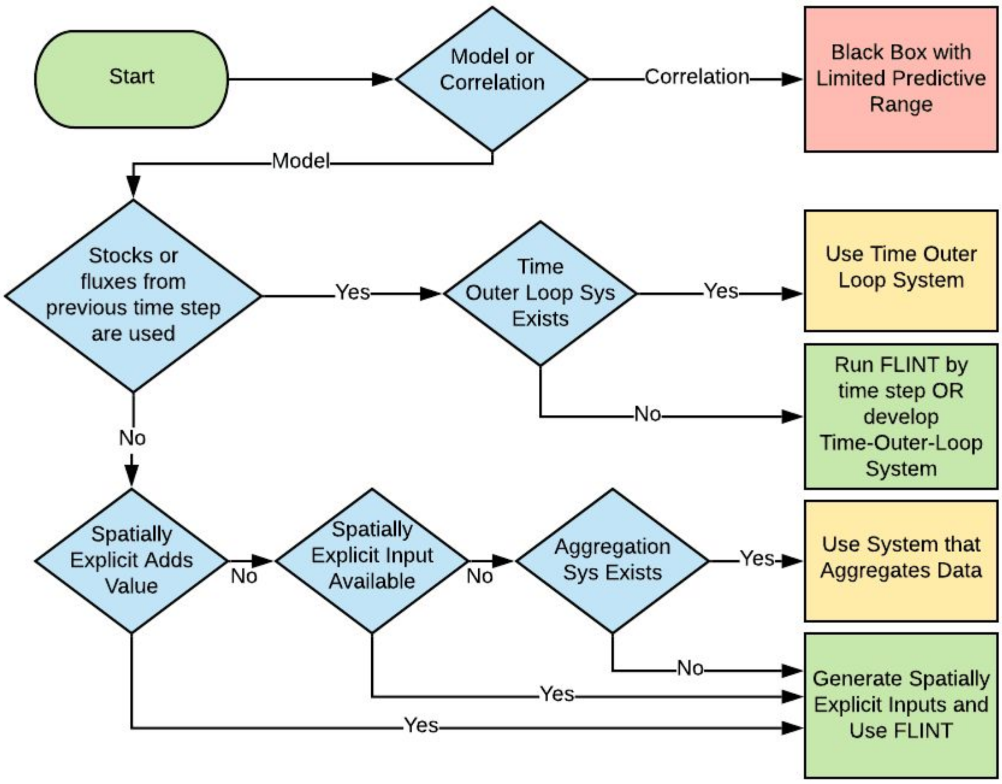
Conclusion

Due to design decisions FLINT has strong capabilities for projections, but to take advantage of these capabilities, specific input data are required. Below the above design decisions are reworked into a diagram that indicates what type of projections tool can best be used.

If projections are important, using a consistent modeling approach for both historic and future emissions definitely is preferred. If the system used for historic projections is based on field observations of differences between stocks, the scope for projections is limited. Some correlations between input variables and emissions estimates might provide limited predictive scope, but it will be hard to judge under which circumstances the correlations remain valid.

FLINT cannot currently be used if disturbances are planned on the basis of carbon stocks or emissions/removals of GHG in the previous year(s) across the landscape (or time steps). A compatible system using time as its outer loop should be used for projections that use the results of the emissions estimation tool. A key design feature of FLINT is that it calculates the changes for one spatial unit for every time step of the whole time series before calculating the next spatial unit. So FLINT uses time as the inner loop and space as the outer loop to iterate through its calculations. This makes it very hard to use FLINT outputs across the landscape, as inputs for model decisions, i.e. sorting emissions estimates for all spatial units and then applying disturbances to a range of estimates in next time step is not possible as it would require the output of all spatial units for one particular time step before moving to the next time step. For example, it is not possible to plan a harvest every year of the forest with the lowest carbon sequestration during the year before. By definition, this can only be achieved with a system that first calculates the whole landscape for a particular time step and then moves on to the next time step, i.e. a system with time as the outer loop.

Decision Tree for the Type of Projections System to Use



Using different systems for historic and projected emissions estimates requires these systems are compatible to ensure consistency. If a FLINT-based system has been used for the historic emissions calculations but this system cannot be used for projections, it is advisable to calculate projections with a system that is compatible with the FLINT-based system but that uses time as its outer loop. The best example is CBM-CFS3 for the FLINT-based GCBM system. If such a system does not exist (or is too hard or costly to develop), only few alternatives remain for projections that use carbon stock or fluxes from previous time steps as inputs into the next time step: For example, one could approximate the expected carbon stocks or fluxes by running FLINT for five time steps and then register the values across the landscape before running the next five time steps. Another compromise could be that one could approximate stocks with another variable (e.g. stand age) and plan the disturbances on the basis of stand age.

If spatially explicit results do not add value, spatially explicit inputs are hard to generate and a compatible system that aggregates data exists, it is advisable to use the aggregation system as it reduces required computer resources and facilitates decision making. If the input data can be generated in a spatially explicit manner with little effort (relative to generating the same data in an aggregated format), continuing to use the FLINT-based systems is preferred for consistency reasons even though the spatially explicit results may not add (much) value. If a FLINT-based system has been used for the historic emissions calculations and it is hard to generate spatially explicit input data for projections, consistency can only be maintained if the aggregation system is compatible with the FLINT-based system. If such a compatible system is not available, it might still be easier to generate spatially explicit inputs and continue to use the FLINT-based system.

If spatially explicit results have added value, it is advisable to use the same FLINT-based system for both historic and projected emission estimations. This requires the generation of spatially explicit inputs, at a minimum the initial land cover and attribution of land cover change.

Analysis of Some Existing Projection Systems

In this section, some existing projection approaches are analysed and their designs are compared with those of FLINT. So for each system the following features are reviewed:

1. Model or Correlation Based: Is the system using models or correlations to project future emissions. For the systems that are using a modeling approach, have they opted for a lower cost of projections or a lower cost of uncertainty: i.e. are the models using many variables to increase accuracy or are the models simple to reduce computing and data cost.
2. How are inputs generated: Model based systems can produce projections without change to the system as long as the input data are available. So the method to generate the input data is the key process.
3. Is time the inner loop and space the outer loop: Not only is it important to determine what the inner and outer loop of the system is, it is also important to establish whether it is a design feature, i.e. a conscious decision. If so, it is useful to understand why it was decided to use that particular inner and outer loop.
4. Spatially explicit or aggregation: Finally, it is useful to find out whether the system is spatially explicit and why.

The examples used below are only a few examples of projection systems. It is useful to continue to analyse additional systems in the future to continue to learn about reasons that drive design decisions.

FLINT-projections without an input generating tool

FLINTpro has been used to generate projections using manually prepared spatial inputs. A project manager used FLINTpro for a small project. The site had been degraded over many years. The past degradation was used to establish a baseline. FLINTpro was used to project the maximum sequestration potential for the site over 20 years, i.e. if the whole project area would be completely reforested in one go, how much carbon would be sequestered over the next 20 years.

The four design features are the same as those discussed for FLINT (and FLINT-based systems like GCBM) above since FLINTpro is a FLINT-based system:

1. FLINTpro is a model based system. The project used Tier 1 default emissions and growth factors as total cost had to remain low and the projection only needed to provide a ballpark value.
2. Generating the inputs was easy as the whole project area was reforested in one go. So the development of spatially explicit maps was easy, i.e. turning every pixel that was deforested back into forest.
3. FLINTpro uses time as the inner loop. This was a conscious design decision to reduce computer resources and allow parallel computing. There were no future disturbances so there was no need to sort landscape features to decide where the next disturbances would need to be applied.
4. The system is spatially explicit. There was no added value to using spatially explicit information because the whole project area was turned into forest. Still a spatially explicit system was used as the consistency between past and projected emissions could be maintained. This also allows for the generation of maps as a way to permit third party verification of the reforestation outcomes.

Consistency between past and future estimations is an important consideration. Spatially explicit projections were the obvious choice because it was easy to generate spatially explicit inputs but also because there was no compatible aggregation system available.

The future inputs were generated manually but this method quickly reaches its limits. The required inputs were particularly easy to generate. Generating spatially explicit inputs can be useful for more complex projections. An overview of what can easily be done without additional tools can be found here. Events like reforestation, harvest and even wildfires can be mapped manually if the project area is not too complex but the limits of manual generation of spatially explicit inputs are reached fairly quickly when issues like beetle attacks, drought, or combinations of disturbances have to be mapped out.

Carbon Budget Model of the Canadian Forest Sector Version 3 (CBM-CFS3)

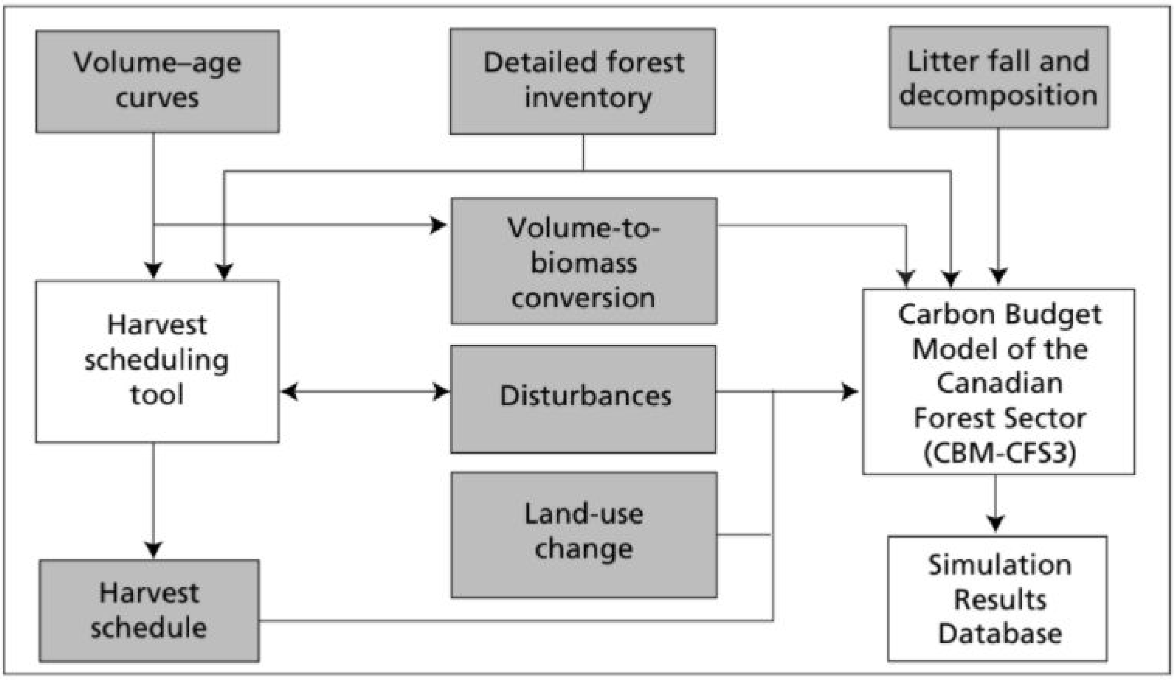
The Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3) is well established and used by many countries, organisations and companies. The tool is a stand- and landscape-level model of forest ecosystem carbon dynamics that forest managers and analysts can use to assess carbon stocks and stock changes on their forest land base.

Monitoring and projection are supported by the CBM-CFS3. The model can be used to assess past carbon stocks and stock changes using information on forest inventory, and past management actions and natural disturbances, or to evaluate future carbon stock and stock changes that would result under different scenarios of forest management and natural disturbances.

The CBM-CFS3 is an integration framework that combines data from multiple sources to estimate GHG emissions and removals. To create a project, the user defines groups of spatial units based on administrative and ecological boundaries, where forest areas may be subject to similar biomass, decay, climate, and disturbance parameters. Within each spatial unit are forest stands, each defined by an age, a leading species, up to nine other classifiers (e.g. administrative area, climate zone, etc.) and an area (ha). Much of this information is typically readily available from existing forest inventories. For smaller projects (ideally less than 10 stands) a manual data entry process can be used in a graphical user interface, and for larger-scale projects (e.g. 100's to 1000's of stands), properly formatted data files can be imported. When a project is created in the model, default assumptions are automatically established in all of the Assumption Composer windows based on the user's imported data, and any Archive Index Database (AIDB) data assigned to the project based on the user's selection of administrative and ecological boundary(ies) for the project (e.g. climate data, disturbance matrices, decay parameters, species specific volume-to-biomass conversion coefficients, etc.).

The key inputs are growth curves and disturbances (i.e. the types, amounts and intensities of disturbances, such as fires, insects, management and land-use changes). Carbon stocks in the CBM-CFS3 are mainly driven by the empirical or model-derived volume (m^3/ha) over age growth curves submitted by the user to represent the growth of their different stand or forest types (defined by classifier sets) over time. The volume (converted to biomass in the model) on the curve at the age of a stand is multiplied by the area of the stand to obtain the aboveground biomass, and then equations are applied to derive the associated belowground biomass components (i.e. fine and coarse roots). Biomass and dead organic matter turnover rates are used to estimate changes to carbon pools resulting from annual processes of turnover. A wide range of disturbance types (each with their own matrix of associated carbon transfers) are available in the model, and users can easily add their own. Disturbance events can target area (ha), merchantable carbon (tC), or a proportion of eligible records.

Data Input Diagram for CBM-CFS3 (from the User's Guide v1.2, figure 1-6)



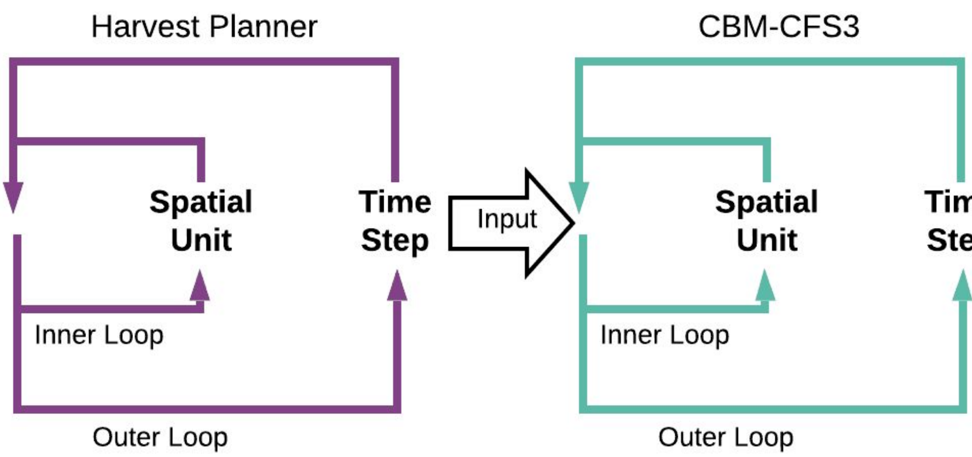
The model is used extensively for projections. Future disturbances, whether natural or anthropogenic , can be based on historic or predicted frequencies and amounts of disturbance (e.g. increases, decreases, or business as usual), and typically, harvesting is a little more predictable (at least in the short-term). The user can create several CBM projects based on the same inventory and growth curves, but applying different future disturbance predictions/plans to see what impact on carbon stocks and flows each scenario will have, and more adept users can do the same by setting up multiple scenarios within the same project so that their results can be compared and contrasted . Each scenario can employ a different change to the default CBM-CFS3 parameters assigned to a project, in order to determine any sensitivities of carbon stocks to changes for each. The CBM-CFS3 implements simulations and reports on carbon indicators on an annual basis for each inventory record:

1. Increment the time step and age all stands by one year
2. Apply disturbances in order of ascending default disturbance type ID:
    a. Determine stand eligibility by classifier set, age, etc.
    b. Apply to eligible stands based on sort type (highest merchantable, oldest first, etc.) until target (area, carbon, or proportion of eligible records) is met
    c. Apply transition rule, if present, to disturbed stands
    d. If no explicit transition rule and live biomass drops to 0, reset age to 0
3. Apply growth and turnover
4. Apply decomposition
5. Update all carbon pool values in real-time as steps 1-4 occur
6. Increment the time step and age all stands by one year

The CBM-CFS3 has similarities with FLINT, but it uses space as the inner loop and the input is generated by a harvest tool that also uses space as the inner loop. FLINT is modeled in part after the CBM-CFS3 and hence there are many similarities in particular with the GCBM that uses the same science as CBM-CFS3 (see below). The four design features of CBM-CFS3 are:

1. The CBM-CFS3 is a model based platform. Moreover, the use of forest inventory and growth curves allows for very precise modeling with very few variables. By developing growth curves for ever more specific circumstances (i.e. type of vegetation, soil, climate, etc.) the system gets more accurate without increasing complexity and thus without increasing computer and data costs. The disadvantage is that this approach requires empirical observations to develop the growth curves, and ongoing updates to the inventory.
2. Inputs are generated manually (for very small projects up to 10 stands) or they are imported from input data files, which, in the case of disturbance time series can be generated from harvest scheduling tools. Harvest planners might not include disturbances like wildfires and insect infestations, so the user may need to import additional disturbance events and transitions using rules and criteria. Rules-based application of disturbances is very powerful as not every disturbance needs to be defined. Instead the system will find instances where disturbances need to be applied. These instances can be defined on the basis of criteria including age, merchantable biomass, and time since last harvest. The option to use rule-based disturbance definitions, such as biomass carbon as an annual harvest target, a set of rules that define the harvest eligibility (e.g. species, age, ownership. etc.) and merchantable biomass as a sort-option for disturbances, is a key difference with FLINT-based systems.
3. Space is the inner loop. For each time step, the whole landscape is processed. This provides a state of the whole landscape before the next time step starts, which in turn allows for rules-based disturbances. The landscape classifiers can be sorted before disturbances are applied. But, there are actually two inner and outer loops!

Double Inner and Outer Loop



In those cases where an (optional) harvest planning tool is used in combination with the CBM-CFS3, the harvest planning tool that generates the input for larger landscapes, also runs through every spatial unit of the landscape and for every time step. So, the harvest tool applies space as an inner loop too.

4. CBM-CFS3 is aggregating the forest types by classifiers which reduces the number of spatial units and thus computing time.

Using tools for discrete tasks, increases efficiency even if it leads to some duplication of calculations. The (optional) Harvest Planner and the CBM-CFS3 are repeating some calculations: e.g. timber volumes are calculated over time by both systems. This seems like a waste of resources, but in fact, the separation of the tools reduces complexity considerably which compensates for some duplication. Operational harvest planners are routinely used in forest planning, such as the determination of long-term sustainable cutting levels. Future harvest scenarios can be assessed for their impacts on carbon emissions and removals by running these through the CBM-CFS3.

Biomass carbon can be used, but could be approximated by age in many cases. The CBM-CFS3 can use a range of carbon related values to trigger disturbances. The indicators related to merchantable carbon are approximated by the harvest planner tool using age or wood volumes derived from the volume over age curves. Actually, it would be more correct to state that the CBM-CFS3 uses biomass carbon as a proxy for wood volume in most calculations. Other important indicators like snag volume are not age dependent and are not calculated by most harvest planner tools. A later section of this report evaluates whether it is possible to build a FLINT input generator for spatially explicit information. If disturbances are triggered by variables like snag volume that cannot easily be approximated (e.g. by age), the input generation tool will have to be more complex and more calculations might have to be duplicated.

The CBM-CFS3 meets all the criteria of an advanced MRV system with the exception of spatially explicit output. The model achieves high accuracy efficiently: i.e. requiring few input variables and only limited computer resources. Consistency is achieved between historic and projected emissions estimations as the same system with the same models generates both. Space is used as the inner loop, so results from the previous time step can be used to sort and select areas where disturbances should be applied.

Australia's Full Carbon Accounting Model

The spatially explicit Full Carbon Accounting Model (FullCAM) uses aggregation to generate projections. FullCAM can in theory project expected emissions based on the standard modeling approach it applies to historic emissions. In practice however, the projections are typically run using aggregated information and averaged growth models. FullCAM is spatially explicit for historic emissions as it derives land use changes from satellite imagery. Similar to CBM-CFS3, FullCAM models the exchange of carbon between different pools of the atmosphere, vegetation, litter and soil. So it is a closed system that maintains mass balance. For projections however, spatial data are aggregated:

- For projections of net emissions from forest lands, log harvest forecasts were adopted from the 'business as usual' scenario published in the Outlook Scenarios for Australia's Forestry Sector: Key Drivers and Opportunities (ABARES 2015). The projections for forest lands are not run using the spatially explicit capability of FullCAM, they are modelled using the estate function which is spatially referenced.

- Projection rates of primary clearing of forest lands and their conversion to croplands and grasslands are derived by using correlations to other measures, such as farmers' terms of trade.
- Projected rates of clearing of regrowth on previously cleared land - for example, management of bush encroachment on grazing land - are based on historical averages.

These approaches to future disturbances yield data that are no longer spatially explicit. Hence the FullCAM modeling approach is not applied to these projections. Rather the projected rates are combined with external models to calculate the greenhouse gas emissions and removals. However, the growth rates and carbon stocks applied are derived from the same data types as used for the historic reporting, ensuring a degree of time-series consistency.

Spatially explicit information has marginal added value for national scale projections, so a compatible aggregation system was developed. FullCAM was designed to inform policies and generate reports for the national government. Spatially explicit projections are not required for these objectives and generating the spatially explicit inputs would require more effort than developing a tool that can generate estimates on the basis of aggregated information even if it reduces consistency:

1. Full-CAM is a model based system. While the models used for historic emission estimations are tier 3 by the UNFCCC definition, for projections tier 2 emissions factors are used. So the designers have clearly chosen for a lower accuracy to reduce the cost of projections.
2. The inputs are generated on the basis of past averages, regressions or business-as-usual scenarios. These methods by definition are not spatially explicit. Redistributing the aggregated information across the landscape is theoretically possible to continue to run the spatially explicit Full-CAM but would not yield any useful information.
3. Space is the inner loop and time is the outer loop for both historic and projected estimates. It is not clear whether this is a conscious design decision or just the result from other system features. The inputs are generated for the whole landscape and for the whole time series before the system is run. The system has no internal feedback loops so the inner and outer loops have no influence on the emissions estimates. It is therefore not possible to use carbon stocks from previous time steps in the distribution decision of disturbances.
4. Projections are made with a complementary tool that aggregates the spatially explicit information used by Full-CAM. Full-CAM uses spatially explicit inputs for historic emission estimations because nesting allowed for emissions trading (see next para). Projections are not driven by these same reasons. So while FullCAM has the capacity to run the same modeling framework for past and future emissions estimates, the return-on-investment of spatially explicit projections is too small: On the one hand the investment is high as the lack of spatially explicit information about future disturbances makes it hard to generate the spatially explicit input necessary to run FullCAM in regular mode. On the other hand, the return is small as the projections are only used for national policy decisions without planning or consultations, in addition to aggregated reporting (e.g. to UNFCCC).

Spatially explicit information has considerable added value for sub-national and project-level accounting, so the system design allows for this. A core design aspect of FullCAM was to enable nesting of projects within the national account. This was done to support the development of emissions trading systems by reducing costs and increasing mitigation actions. To support this, the system allows for areas of new projects (such as deforestation or reforestation) to be added to the system and modelled. As such new project proposals could be evaluated using data and systems consistent with the national account. Presently this is completed using a simple 'point-based' approach but the fundamental design aspects remain intact. As such FullCAM provides an excellent example of differing needs for projections at different scales.

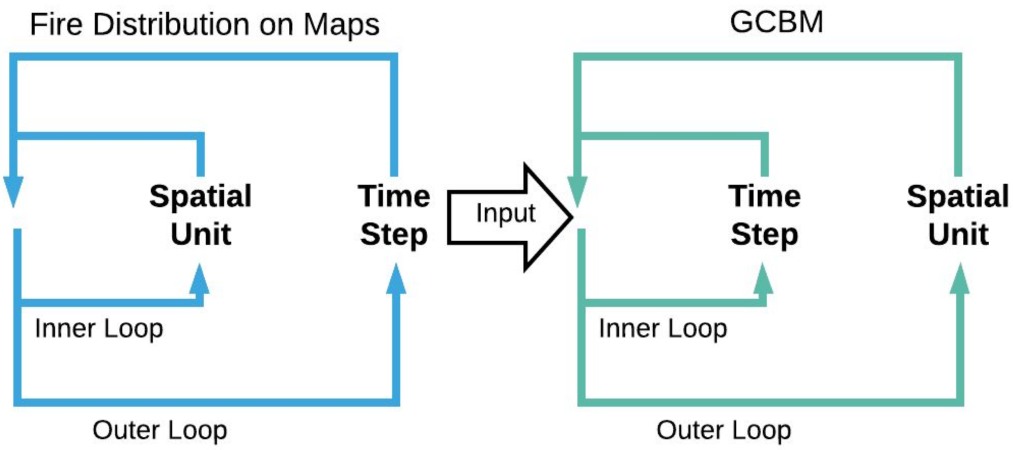
CFS use of GCBM for Fire Projections

The spatially explicit GCBM is used to estimate the emissions impact of fire regimes. The Canadian Forest Service is reviewing the risk to mitigation outcomes in the forest sector from wildfires in the province of British Columbia. Based on analysis of spatial records of historic fire, a regionally-differentiated probability distribution for annual area burned is applied to each year for the period 2020 to 2070. The selected area is further broken down into a number of fires by selecting fire sizes from a probability distribution for fire size, and adding up the fire sizes until the total annual area burned is reached. The latter fire sizes are applied to the landscape as ellipses with random location, size, shape and orientation. This process is repeated 100 times to obtain different scenarios for each of two fire regimes: One that emulates current area annually burned, a second that simulates a doubling of average area annually burned. So 200 maps of British Columbia are generated for each of the 50 years, i.e. 10'000 maps. The 10'000 maps (i.e. 200 runs of 50 time steps) are run using GCBM at 1 ha resolution for the entire forest area of BC (about 60 million ha) to estimate the carbon losses in various pools.

GCBM is a FLINT-based system with the same four design features:

1. GCBM is a model based system. Since GCBM uses the same science logic as the CBM-CFS3 recoded into modules that operate on the FLINT platform, it has the same approach to high accuracy and low computer and data cost. And the system can be made gradually more accurate by developing growth curves for ever more specific circumstances.
2. The inputs have to be generated on the basis of maps as they are based on regionally-differentiated probability distributions for annual area burned, fire sizes, shape and direction. This is only possible in a spatially explicit manner for each year.
3. GCBM is a FLINT-based system so time is the inner loop and space is the outer loop. However, the tool used to allocate the fires across the landscape can also be considered a double loop with the opposite sequence: i.e. space is the inner loop and time is the outer loop.

Double Inner and Outer Loop With Opposite Sequence



This is possible because the carbon impact of the fires allocated across the landscape is not a necessary input into the probabilistic distribution of the fires in the future. Only when fire spread models are used, that use biomass on the landscape as a factor that determines the size and type of fire which again has an impact on carbon fluxes and stocks, will carbon pools become on input into the emissions estimations and will FLINT no longer be useful as its inner loop is time.

4. Projections are made in a spatially explicit way. The additional effort is justified because the projections are combined with road network, mill location and other spatial information to estimate potential future biomass availability under different fire regimes. The fires are distributed across the landscape in a probabilistic manner. This means that the mapped results are only one of many possibilities that have an equal chance of occuring and therefore Monte-Carlo type analyses are conducted to obtain probability distributions of burn riks. This is research in progress.

GCBM is used because inputs and outputs have to be generated spatially. GCBM systems have spatially explicit inputs and outputs. In case of the probabilistic fire regimes, the spatial outputs allow for the estimation of biomass availability relative to road networks and locations of biomass utilization and the selection of the location of potential future facilities that can use biomass from salvage logging after wildfires.

Forest carbon calculation at the Bureau du Forestier en Chef in Québec, Canada :
The Bureau du Forestier en Chef (BFEC) is trying to optimize not only yields and biodiversity but also carbon sequestration in the complete carbon cycle. Québec's 76.1 million hectares of forests account for 20% of Canada's forests and 2% of the world's forests. In terms of area, 92% of Québec's forests are under public ownership. Québec's forests extend over seven degrees of latitude and three major bioclimatic zones, each with its own highly specific characteristics. These three major zones, with their principal forest species, are the boreal forest (black spruce, balsam fir and white birch), the mixed forest (yellow birch and balsam fir), and the hardwood forest (sugar maple and yellow birch). (copied from SFM Canada.) Québec's Ministère des Forêts, de la Faune et des Parcs (MFFP) prepares integrated forest management plans for roughly 60 management units, in collaboration with local integrated land and resource management panels. The BFEC has the mandate to ensure the constant renewal of Quebec's forests, all with a view to

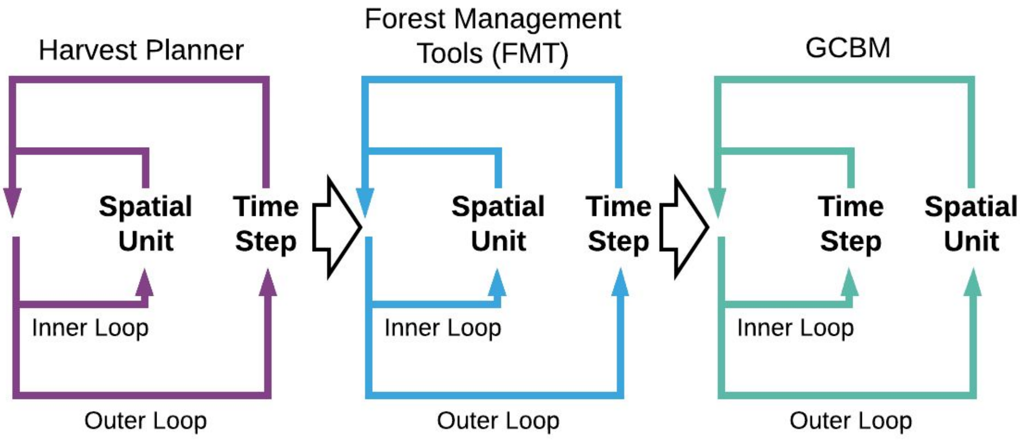
sustainable development. The BFEC calculates the allowable cuts for each forest management unit. The integrated forest management plans are then prepared with a view to ecosystem-based management, which seeks to maintain ecosystem biodiversity and viability. (SFM Canada) BFEC would like to add a variable to this optimization process: i.e. maximising the removal and minimising the emissions of GHG from forests.

Advanced tools are used to combine spatially referenced harvest planning with spatially explicit GCBM estimations. Woodstock, the harvest planning tool from Remsoft, is used to create a future land disturbance calendar based on various management strategies in the public forests of the province of Quebec. The most optimal scenarios are selected to calculate carbon stocks and fluxes. Woodstock produces coordinated results at different scales: from the forest down to the polygon level. But the output is not spatially explicit as the results are under-defined, i.e. additional information is necessary to determine where exactly disturbances will take place in space and time. To make disturbances spatially explicit an intermediary tool called “Forest Management Tools (FMT)” is used. FMT is an open-source project using the LiLiQ-R-1.1 license. FMT uses Woodstock model files to optimize or simulate forest planning in four different ways: spatially explicit simulation (FMTsesmodel), spatially referenced simulation (FMTnsmodel), spatially explicit optimization using Simulated Annealing (FMTsamodel) and finally, spatially referenced optimization using linear programming (FMTlpmodel). The main approach used in FMT is called the spatially explicit simulation model (FMTsesmodel). FMTsesmodel uses a spatially explicit simulation process combined to a Monte-Carlo approach to spatially allocate disturbance blocks using various user’s defined parameters such as maximum block size, green-up delay, adjacency delay, etc., and meeting harvest volume targets on the minimum area possible. This spatially explicit disturbance schedule is combined with other spatial inputs: inventory data, growth and yield curves, natural disturbances, expected land use change and ecological parameters (climatic zone, temperature, precipitation, etc.) This input for each scenario is used to calculate carbon stocks and fluxes with GCBM. The Simulated Annealing approach (FMTsamodel) is presently under development in FMT. This approach will allow spatially explicit optimization based on multiple non-linear objectives. This new approach will allow the usage of a carbon emission variable to the optimization process.

GCBM’s four design features are known but decisions were motivated slightly differently:

1. GCBM is a model based system that achieves high accuracy and low computer and data cost due to simplicity. Adding complexity will undo this strong performance.
2. The inputs are aggregations that need to be made spatially explicit using additional information. A separate tool has been developed to transform the spatially referenced management plan into spatially explicit rasters that can be read by GCBM.
3. GCBM is a FLINT-based system so time is the inner loop and space is the outer loop. The harvest planner (which is actually a forest management planner) is spatially referenced and uses space as the inner loop. FMT also uses space as the inner loop and time as the outer loop. So the landscape for each timestep is processed three times in the actual simulation approach used by the BFEC.

Triple Inner and Outer Loop

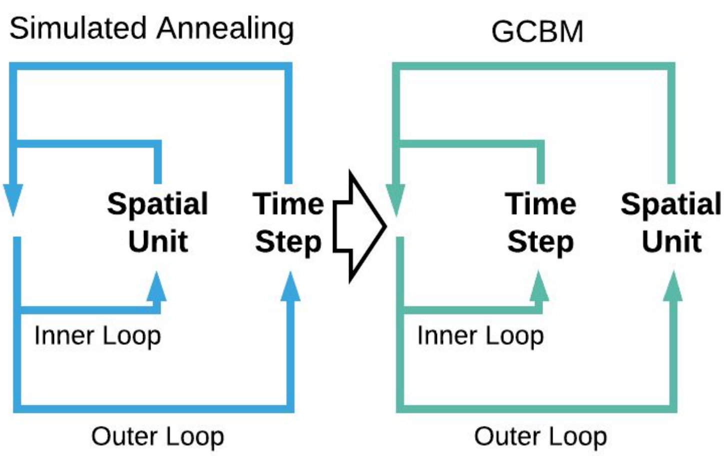


The three tools repeat some calculations, e.g. the harvest planner and GCBM generate volumes based on growth curves.

4. Projections are made in a spatially explicit way. While inputs are generated spatially referenced at first, a special tool (FMT) is developed to make them spatially explicit. This extra effort is compensated by the added value of spatially explicit emissions estimations. The most valuable advantages of spatially explicit results are the additional information available when information is in a map format. This additional (soft) information allows for more engaging public consultations and more realistic forest management plans.

The new Simulated Annealing approach (FMTsamodel) developed at the BFEC is planned to be based on double Inner and Outer Loops. The new approach will not use a spatially referenced harvest schedule generated by Woodstock like the spatially explicit simulation approach (FMTsesmodel). It will entirely generate its own spatially explicit harvest schedule based on multiple non-linear objectives. The new approach could also make great use of callbacks to GCBM and use the ability of GCBM to calculate carbon fluxes to direct its search to a global optimal solution.

Double Inner and Outer Loop Preparing for Optimization



Building an additional tool with a focused function can result in more efficient processing than upgrading an existing tool with additional features. The development of the spatially explicit simulation tool (FMTsesmodel) requires that the landscape for each timestep is processed one additional time. Since the tool has a focused (albeit very complex) function, the overall performance of the system is likely higher than a system where the same functionality would be added to the existing tools (e.g. allowing GCBM to ingest spatially referenced data and process the data in the same way as the current spatially explicit simulation tool). Even though there might be some overlap in the calculations executed by the different tools, the obtained simplicity will compensate for this loss in efficiency. There are additional gains from using a tool for each contained function, the design and maintenance of the tools becomes simpler. (There are limits to this rule: Too many tools will continue to increase simplicity but the efficiency gain is lost to the increased effort to maintain interfaces and interoperability.)

Projections Needs Analysis

In this section, the needs FLINT users face in terms of projections of emissions estimations are analysed with a focus on the required inputs and tools. Based on the analyses above, it is preferred for consistency reasons to use a FLINT-based system for projections if that system has already been used for historic emissions estimates. If spatially explicit inputs are easy to generate, FLINT-based systems tend to be used even if the spatially explicit output does not add substantial value. If carbon stocks or fluxes from previous time steps are needed as inputs however, FLINT-based systems can only be used if a proxy (e.g. age) can be used for these input variables. Therefore this section will identify the types of projections commonly used for the land sector and assess whether:

1. Carbon stocks or fluxes from previous time steps are required as inputs for projection calculations. If so, whether these inputs can be approximated by other variables
2. Identify the other inputs and assess whether they can easily be formatted in a spatially explicit way relative to a spatially referenced way.

Based on the information above, it is evaluated whether FLINT-based systems can be used for this type of projection. Finally, the tools are identified that one requires to run projections assuming the regular FLINT-based emissions estimation tool is used as the core of such a projection system.

Deforestation Baseline

A national baseline that is consistent with sub-national and project level baselines can be achieved with a FLINT-based system as projecting emissions requires an indication of future deforestation in a spatially explicit way. A flexible tool needs to be developed to generate spatially explicit projections of deforestation.

In many countries there are REDD+ performance-based efforts occurring at multiple scales, including national, subnational and project levels. Many of these are engaging in some kind of results-based finance, whether non-market instruments that pay for jurisdictional performance (such as the Green Climate Fund REDD+ RBP Pilot Programme, the FCPF Carbon Fund, or BioCF-ISFL) or through the implementation of forest carbon projects that engage in (growing) voluntary carbon markets.

One of the most challenging issues for countries is to promote alignment of data and systems for estimating emissions among these initiatives. In most cases different groups use different data, methodologies, standards and guidelines for measuring performance. One emerging solution - specifically for "avoided emissions from deforestation" (i.e. RED) - is for a country to set a national forest reference level (FRL) as a benchmark and then allocate that baseline down to subnational units or projects. However, even in these cases the data and systems used for calculating the FRL are then inconsistent with the data being collected for monitoring. In these cases, having spatially-explicit systems ensures that double-counting and within-country leakage (i.e. avoided deforestation in one location triggers increased deforestation elsewhere) can be detected.

Breaking up a national baseline can be challenging because the rate of deforestation will differ considerably across the landscape. For example, forests close to agricultural expansion areas will have high rates of deforestation and remote forests on a steep slope will have low rates of deforestation. A single national value does not represent this variability. For example, if the national deforestation rate is applied to forest areas across the country than the baseline in the former areas (e.g. agricultural expansion) will be too low, and in the latter areas (e.g. steep slope) will be too high, i.e. the estimated deforestation does not reflect well what would happen in case of "business as usual". This creates disincentives for project developers who will find it hard to even meet the baseline in the priority areas (e.g. agricultural expansion) and who will without any interventions stay below the national baseline in low priority areas (e.g. steep slope). So a baseline should ideally be linked to the local drivers of deforestation, but still be aggregated to a national baseline without losing consistency.

There are many ways to establish a local baseline and the best proxies to predict deforestation are the subject of extensive research. Most of these proxies are based on spatial information: e.g. location of hamlets, roads, types of agriculture, population density and growth rates, most recent deforestation, etc. The best proxies and algorithms to predict deforestation will differ by locality or country but most approaches will use maps with information about location of roads, hamlets, types of agriculture, time since deforestation, land-use change, etc. Some of these maps are combined using an algorithm to generate a map with the predicted deforestation in a spatially explicit manner.

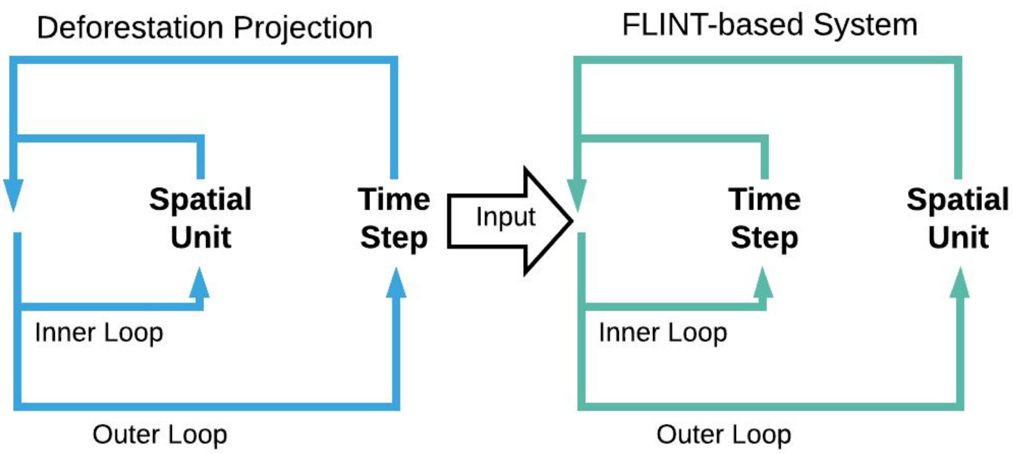
FLINT is ideal for these projections because the inputs have to be generated spatially explicitly:

1. Carbon stocks or fluxes over time are not required as inputs for baseline projections. The baseline calculation is the expected emissions estimates caused by deforestation if no interventions would influence business as usual. FLINT can calculate the emissions if the rates of deforestation and resulting land cover change are projected in a spatially explicit manner. So deforestation has to be projected on

the basis of an algorithm. Even though the exact algorithm is still being researched and will most likely differ between locations, it is unlikely that such an algorithm will require a time series of carbon stocks and fluxes as an input.

2. The inputs most likely used by a deforestation projection algorithm include location of roads, hamlets, types of agriculture, population density, growth rates, location of most recent deforestation, location of recent land use change by type and potentially a starting condition (i.e. is it forest?), etc. By definition these are spatially explicit inputs that can not be generated in a spatially referenced manner. Similar to the fire regimes above, the inputs have to be mapped and thus a spatially explicit system like FLINT can be used to translate input data of time series of maps of annual deforestation locations into estimates of emission projections. .

Deforestation Maps as Inputs for Spatially Differentiated Baseline



Projections of a baseline requires a flexible map builder projecting deforestation and its component parts with space as its inner loop. The key input for the baseline is future deforestation which is estimated on the basis of a number of precursors, i.e. visual indicators that feed into a formula (roads, hamlets, etc.) The formula can be country specific and could use different parameter values for different subregions. So the visual precursors need to be projected first. Algorithms or trends can be used to map projections of precursors. These are then in turn used to calculate an estimate for the expected deforestation at that time step. So both precursors and deforestation use the landscape conditions at the current time step to estimate the changes for the next time step. So the landscape conditions have to be calculated for the whole landscape before moving to the next time step. Probably a flexible tool is required to project the precursors individually as well as to combine the precursors to match the local type of deforestation as closely as possible.

Land Management Policy Scenarios

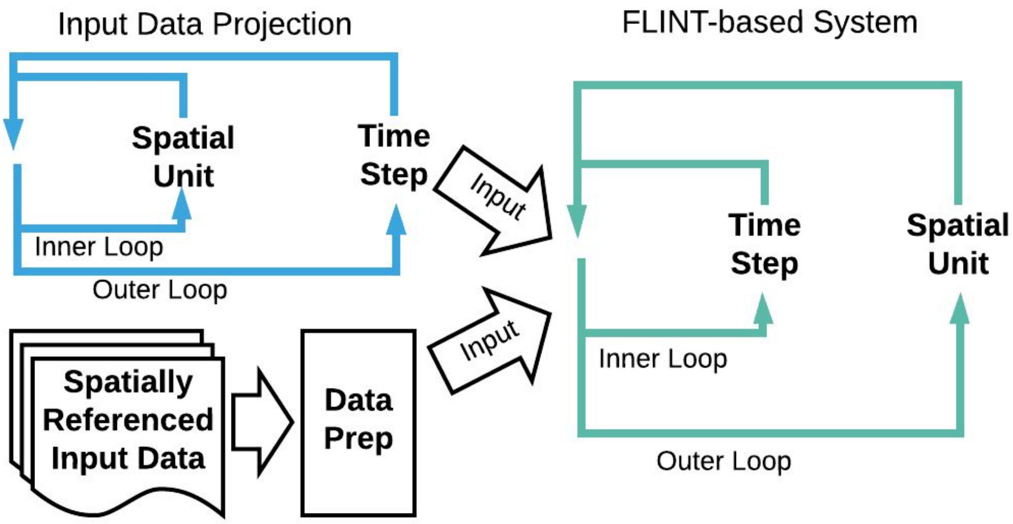
FLINT-based projections of policy impacts are viable if output maps and consistency with historic emissions outweigh the effort to generate spatially explicit inputs. If a compatible aggregation system is available, it could be used for instances when spatial outputs are not important. Even for the FLINT implementations that have a compatible aggregation system, it is useful to develop a tool that can project the inputs FLINT needs to calculate projected emission estimates.

Land management policy decisions are made on predictions of their impact which includes emissions and removals of greenhouse gases (GHGs). The policy decisions range from the approval of regular operations (e.g. forest harvesting plans), to the management of natural disturbances (e.g. fires), to the announcement of international commitments (e.g. Nationally Determined Contributions), and to the funding of interventions (e.g. promoting woodproducts to substitute high emission alternatives). Each of these decisions has important GHG emissions and sinks implications. To compare policies, it might be useful to project the impact of each policy separately. Some policies require that several projections are run so that impacts can be combined (e.g. forest harvest plans require projection of growth of the forest estate but also the projection of several scenarios of the losses caused by natural disturbances like fires and pests.)

FLINT-based systems can be used for these projections if inputs of future events (i.e. the activity data on harvest, disturbances and land-use change) are available in spatially explicit formats:

1. Carbon stocks or fluxes are most likely not required as inputs for policy impact projections. It might be important to compare the impact of two or more policy options over a period of time (e.g. 100 years). So the FLINT-based system would run for each policy option over the agreed time period. While the carbon stocks and fluxes over the time period are important information to inform the policy process, the results from previous time steps are not essential as inputs in the calculation of the next time steps. The policies tend to be set and will most likely not include a clause that changes interventions on the basis of GHG stocks and flows at an intermediate point in time. Thus the policy interventions are influencing carbon stocks and flows but not the other way around. This means that the FLINT with its time inner loop can be used for these calculations.
2. The inputs used to project policies impacts are likely a mixture between spatially explicit and spatially referenced. The inputs depend on the policy of course. A wide range of projected inputs might be necessary ranging from those that are only available in spatially explicit formats to those that are most likely only available as spatially referenced data.

Policy Impact Projections Might Require Spatial and Non-Spatial Inputs



Please note that the inputs are projections of what will happen to these variables over the projection period. Spatially explicit projected input data includes occurence of fires or pests, reforestation, afforestation, harvest, land-use change, etc. Other projected input data can be available spatially explicit or referenced depending on local circumstances including plans for forest management, peatland management, agroforestry, species management, etc. Finally there are projected input data that will most likely only be available in a spatially referenced format including application of fertilisers, tillage, grazing, manure management, residue management, firewood collection, etc. All spatially referenced information can be formatted so FLINT-based systems can ingest this data. But it is also possible to aggregate the spatially explicit data and run a system that is compatible with the FLINT-based system used to generate the historic emission estimations.

The importance of spatially explicit results will determine whether a FLINT-based system or an aggregation system is used for projected policy impacts. As discussed above, there are three factors determining whether spatially explicit projections are used: (1) the importance of spatially explicit results, (2) the importance of consistency with historic emission estimations, and (3) the availability of the projection input data in a spatially explicit format. The latter factor has limited influence as for most policy impacts the data can be transformed into a spatially applicable format (i.e. a format that FLINT can ingest). In some cases, this might reduce the value of the spatially explicit outputs however. Consistency is always an important factor but a compatible aggregation system might be available or can be developed to maintain consistency between historic and projected emissions and removals. So the most determining factor will be the importance of spatially explicit results. This is entirely dictated by factors external to the emissions estimates system, e.g. public consultations require maps.

Policy impact projections systems ideally have an input projections tool and a compatible aggregation tool. Since the importance of spatially explicit outputs is dictated by external factors, FLINT should ideally offer both options to its users. Fortunately, an aggregation system exists for GCBM, an important FLINT implementation. However, developing a compatible aggregation system for all FLINT-based systems might be more challenging than developing tools to generate projections spatially explicit inputs. Hence, the development of a tool that can make it easy to project inputs in a spatial way seems to be a necessary addition to the suite of moja global tools.

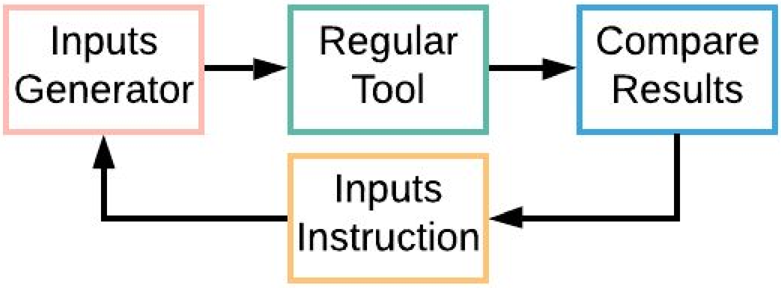
Emissions optimizations

Optimization processes add considerable complexity but in relation to emissions and removals from land are mostly iterations of regular projections. Optimization for spatially explicit systems will most likely be based on an iterative method. If a FLINT-based system is being used for regular projections (i.e. without iteration), it will most likely be used for optimizations too. Some additional tools are required to manage optimization processes. Automation of the optimization process makes it easy to increase the number of iterations and thus accuracy of the results. Automation can however quickly increase the complexity of the system.

Tools for optimization of forest yields exist but few of these tools include carbon in their optimization. Land use management aims to generate the maximum benefits for the smallest possible input. The direct benefits from management include yields and carbon sequestration and storage (so including forest products). There are co-benefits that will be discussed in the next section (e.g. biodiversity or the substitution effect of wood products). The inputs are costs related to planning, monitoring, silviculture, thinning, harvesting, transportation, etc. Forest management software normally includes functions that allow for the optimization of asset valuation, environment and policy planning, silviculture, harvest scheduling, road activity, resource allocation, risk mitigation, wood flow, etc. Most of these tools do not have advanced carbon projections that cover all pools including forest products and related substitution factors.

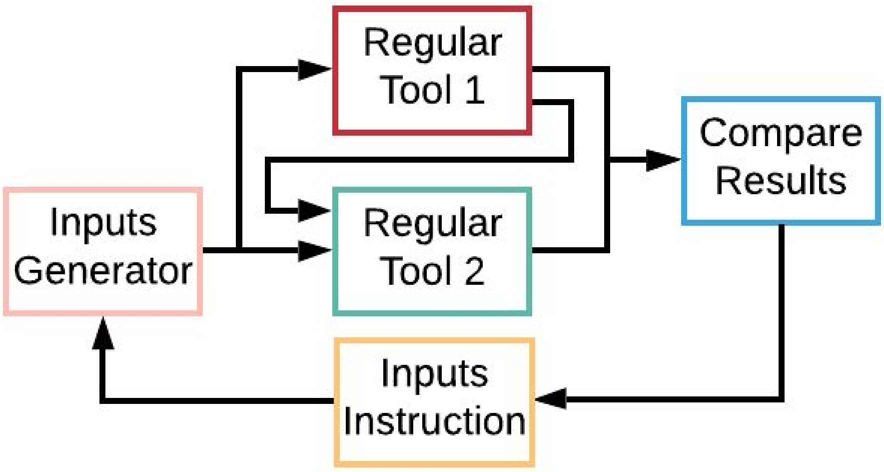
Optimization can be achieved by iteratively improving a candidate solution. There are various approaches to optimization but roughly one could group them as linear programming approaches or iterative approaches. While the linear programming approaches are more efficient, they cannot be used for large spatially explicit problems because the optimization solvers that are currently available cannot cope with the large number of options. So iterative approaches are likely being used for spatially explicit problems. This requires that the results of the previous calculation are used to improve the inputs of the candidate solution. The process goes through the following steps: Initial inputs are selected based on a priori knowledge, at random or through an initialisation formula. These inputs are used to generate the outputs using the regular tool (e.g. FLINT) to generate outputs. The results are compared with the previous results based on the underlying objective one wants to achieve with the optimization (most likely a formula that gives a weight to each cost, yield, and carbon). The best result is adopted. Next a formula is used to adjust the inputs. There are two types of algorithms to improve the inputs: Deterministic algorithms use formulas to sharpen the previous inputs based on the previous results. Stochastic algorithms generate new inputs using probability distribution around the most optimum input found to date. The new inputs are fed into the regular tool again and the cycle is repeated. With increasing numbers of iterations the expected size of the improvement goes down.

Regular Optimization Process



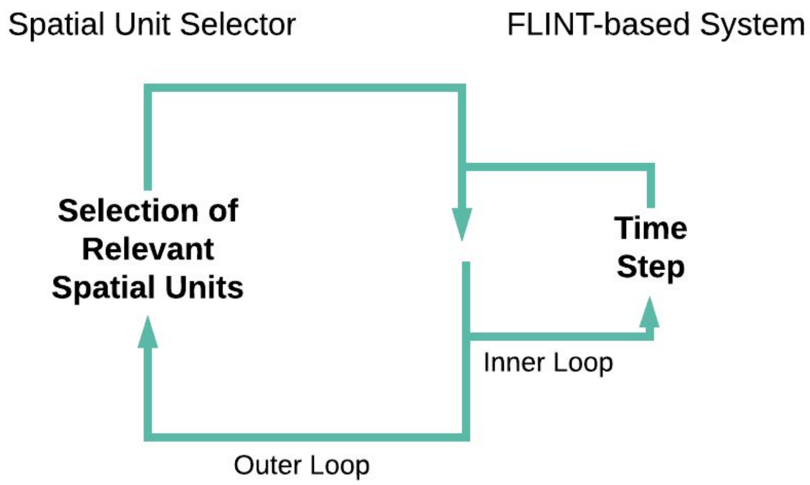
Using a combination of two or more "regular" tools exponentially complicates optimization. Ideally optimization processes are completely automated so the number of iterations can be made dependent on the available computer resources without human interference. When automation is not possible the process becomes very onerous and the number of iterations will have to be reduced to keep the system manageable. If the "regular" tool in the figure above, is replaced with a combination of two (or more) tools (e.g. a forest management planner and a FLINT-based emission estimation system) the complexity of the process will go up considerably and automating the optimization might become too arduous. The number of iterations will be reduced, the initial inputs will have to be based on a priori knowledge to make the few iterations meaningful and as consequently the candidate solution might be considered final at a point where size of improvement with each iteration can still be relatively large.

Optimization Process with Two "Regular" Tools



FLINT-based systems can reduce computer resources required for optimization by only calculating relevant spatial units. Optimization calculations might not require the calculation of the whole landscape to compare results and improve inputs. Since FLINT has time as its inner loop, the total computation load can be reduced if only a relevant selection of spatial units has to be processed. The selection can be based on a representative sample or on the selection of those spatial units that are affected by the input changes. The tool used to select the spatial units would probably be very similar to the tool used to sample spatial units for a Monte Carlo Uncertainty

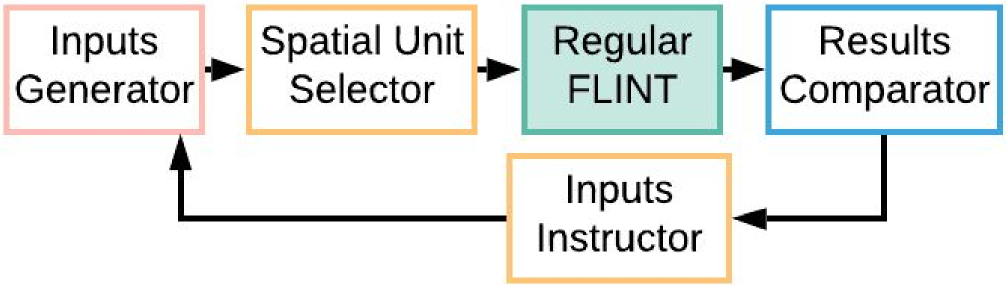
Reduce Computation Load by Selecting Spatial Units



FLINT-based optimization cycles are hard to automate but selection of spatial units can reduce computation load:

1. Carbon stocks or fluxes are most likely not required as inputs for optimization processes. As with the policy impact projections above, FLINT outputs are compared to improve inputs at each iteration but there is no need to use the carbon stocks and fluxes after each time step as input into the next time step. This means that the FLINT with its time inner loop can be used for these calculations.
2. The inputs used for optimization processes are the same as those used for a single run. The inputs are likely a mixture between spatially explicit and spatially referenced data depending on the optimization problem. The inputs are adjusted after each run and are used in the next iteration. A deterministic adjustment would be hard to implement as many variables are required as inputs and for each a sensitivity analysis would be necessary to inform how each input variable needs to be adjusted for each time step. A stochastic approach would most likely be more manageable but due to the many inputs, even probabilistic distributions for each input would be a complex endeavour. The reduction of spatial units to a relevant sub-set might reduce the complexity and the calculation cycles necessary to optimize inputs. FLINT based systems are particularly suited to this type of selection precisely because it has time as the inner loop.

Running a Representative Sample to Reduce Computational Load



FLINT-based systems are used for optimization when they are already used for other projections. Optimization processes are iterations of projections and each iteration uses slightly different inputs. As explained above, the optimization processes start with all initial inputs for a first projection that are run through the “regular” system (i.e. FLINT-based system). It is possible to reduce the variables that can change at every iteration to only those that have the biggest influence on the objective the optimization is trying to achieve. Starting the process however requires the projection of all inputs for the whole time series just like a regular policy projection discussed in the previous section. If regular projections were made using a FLINT-based system, the tools to generate these inputs are already in place. Continuing to use the FLINT-based system for optimization by iterating the same projection seems to be logical. As discussed, FLINT based system can reduce computation load which would give a FLINT-based system an advantage but optimization adds so much complexity to the system that computational load will probably not be the determining factor when choosing the tools for emissions estimations.

Optimization systems need additional formulae and tools compared to regular projections. As discussed, optimization processes have four steps: The generation of the initial input which can be done with the input projections tool used for regular projections. The estimation of emissions which is done with the FLINT-based system. The next step is the comparison of the result with the objective the optimization is trying to achieve (i.e. a combination of cost, yield, and carbon). The objective most likely differs between optimizations so the formula combining cost, yield, and carbon will differ too. A tool could be developed to read the outputs into the formula underlying the optimization process. The result of the comparison will be used by the input improvement tool, to generate a new set of inputs. This tool does not necessarily generate new values for all inputs that were used for the initial run. A tool will be necessary to generate the correction (in case of a deterministic approach) or the probabilistic range (in case of a stochastic approach) used to generate the new inputs. As explained, it could be possible to reduce the number of variables used in the optimization to those with the biggest effect on the objective of the optimization. The remaining variables can be kept constant. Most likely the same input projections tool used for the initial values can be used to improve the inputs but some add-ons will most likely be necessary to take into account the instructions from the input improvement tool. So optimization adds considerable complexity and requires several tools in addition to the input projection tool and the spatial unit selector.

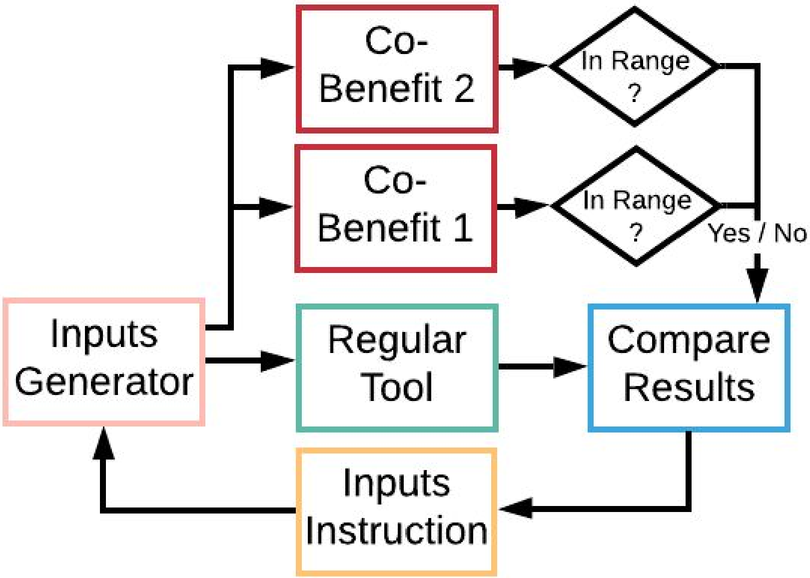
Co-benefit optimizations

Land sector management influences many policy priorities but optimizing more than one in a single system results in hard to manage complexity. While the one policy priority is optimization, the others (the co-benefits) are kept within a set range (e.g. at least at current levels.) This is achieved by using different tools to calculate each co-benefit separately. To ensure the results are consistent, it is important that the inputs provided to each system are the same. To generate the inputs for each of the tools, the inputs projection tool might have to be upgraded.

Land sector policy has to optimize investments in a wide range of crucial priorities. Land sector management influences a wide range of policy priorities that go far beyond yields and carbon storage discussed in the previous section. The co-benefits include food security, biodiversity, wildlife habitat, watershed management, etc. While these policy priorities are termed co-benefits in the context of FLINT-based systems, their policy weight can often be far greater than the importance the policy makers assign to carbon storage and fluxes. The optimization objective is to maximize all these indicators with the smallest possible investment (or at least within the available budget). Cost can also be treated as the dependent variable. This would require the system to sort all acceptable benefits by effort (i.e. inputs) and calculate the cost of each additional unit of any objective. In other words, it would be possible to calculate the cost of increasing the benefit function with one unit. A simple example would be the cost of storing one additional ton of carbon in biomass while keeping all co-benefits at least at the same level.

Complexity of optimizing several co-benefits can best be reduced by only optimizing one objective and keeping others within an agreed range. When policy makers require projections of several or all of the co-benefits complexity will explode. Projecting the impact of policy decisions on each of the co-benefits separately most likely requires several tools. The effort necessary to combine these tools into a single system that automates optimization iterations will probably not weigh up against the benefit of the optimized results. Even the formulation of an optimization function using several or all of the possible co-benefits would be highly contentious, complex and computationally demanding. In these types of situations, complexity is often reduced at the expense of accuracy by optimizing only the most important objective and keeping the other objectives (i.e. co-benefits) within an acceptable range.

Optimization Process Combined with Co-Benefits



Optimization of co-benefits uses the same inputs as emissions optimizations. As indicated above, optimizing across several co-benefits is most likely too complicated so one benefit will be selected for optimization while other co-benefits have to remain within a specific range. If emissions and sinks are selected as the primary benefit, the optimization is the same as the regular emissions optimization explained in the previous section. If emissions and sinks are only a co-benefit, the process is similar to the calculation of policy impacts explained in that section. For both situations the key questions are answered as follows:

1. Carbon stocks or fluxes are most likely not required as inputs for optimization processes. This means that the FLINT with its time inner loop can be used for these calculations.
2. The inputs used are likely a mixture between spatially explicit and spatially referenced data depending on the optimization objective. The set of inputs must be widened to include the inputs required by the co-benefit calculators too.

FLINT-based systems can create consistency if co-benefits require spatially explicit inputs. While the optimization iterations are limited to one objective and the co-benefits are kept within a range, it is still necessary to calculate each co-benefit and thus the inputs required need to be generated. If some of the calculations of co-benefits require spatially explicit inputs that overlap with the inputs required by the emissions estimation tool, a FLINT-based system would be the most appropriate tools to use. If the co-benefits are calculated on the basis of aggregated data, it might be more consistent to switch to a compatible aggregation tool to estimate the emissions.

Optimization of co-benefits mainly requires the upgrading of the inputs generator. The optimization processes for each benefit will be run separately but the inputs for each system will have to be consistent. Therefore the same inputs generator or input projections tool should be used for all projections. This might require that the tool be enabled to generate a wider range of projected inputs. It also confirms that the input projection tool can best be developed as a separate tool to reduce complexity and allow for the projected inputs to be used by other systems than FLINT.

Projections Design Options

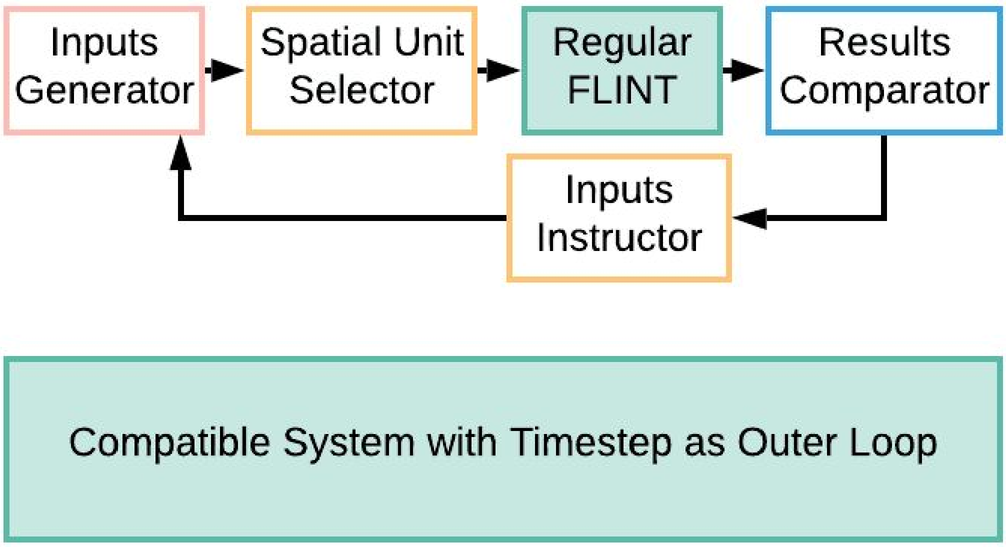
In this section, the design options to enhance a FLINT-based projections system are reviewed. Projections of future GHG emissions and removals using FLINT are easily generated, as long as spatially-explicit input data about future activities (management, disturbances, land-use change, etc.) are available. However, as explained in the analyses above, spatially-explicit input data are not always available or expensive to generate and additional tools required to run projections were identified. These are tools required in addition to facilitate the generation of spatially-explicit input data for use by the regular FLINT-based system. Together they should create a coherent but flexible system that allows the FLINT users to meet most of their projection needs.

A combination of separate tools provides for flexibility and reduces complexity at the cost of some duplication. Rather than integrating the capability to generate spatially-explicit forecasts of the location of future disturbances into the FLINT code, the approach proposed is to combine FLINT with one or more separate tools that each have a particular function. This approach creates flexibility for the user as smaller, more manageable tools might be required for some projections, but not for all. Reducing the functionality of each tool also reduces its internal complexity which in turn results in faster run times (or fewer required computer resources). The downside is that some calculations might be repeated in several tools. For example, the inputs generator might create a raster with merchantable timber volumes to decide when harvests could be applied. The FLINT might use a similar operation to track the carbon flux between the atmosphere and the vegetation. The duplication might result in additional computation requirements but these additional computer needs are compensated most likely by the reduction in computer requirements due to the simplicity of the code. Separate simple tools also facilitate code maintenance, debugging, and adjustments to particular needs.

The inputs generator is the most important of the four separate tools required to meet the range of most common projections. In the remainder of this section, each of the tools required for projections is discussed in more detail. There are four tools and based on the analysis above, the inputs generator would be required for every type of projection. The inputs generator projects the inputs the FLINT needs for every time step. The required inputs can vary considerably between projection types from simple land cover change for a baseline to inputs related to co-benefits for multi-benefit optimization. (Please review the previous section for details.) Therefore, the design of the inputs generator is discussed for each application separately below, but it is one and the same tool. (The increased functionality requirements could be used as sprints in the development of the tool but that is not the purpose of the sections below. The explanation by purpose is solely aiding to explain the function of the inputs tool clearly and simply.) Only optimization processes possibly require the three additional tools: The Results Comparator, tool to compare results with the objective of the optimisation; the Inputs Instructor, a tool to translate this comparison into an instruction for the inputs generator; and the Spatial Unit Selector, a tool that can select spatial units to reduce the computational load as discussed in the previous section. Optimizations also require some additional functionality for the inputs generator as it needs to adjust inputs in line with the commands from the Inputs Instructor.

A FLINT-compatible tool with time step as its outer loop, is a valuable projection alternative for specific applications. If a FLINT-based system is used to estimate historic emission estimations, using it for projections is advisable to increase consistency. But a FLINT-based system cannot be used if carbon stocks or fluxes from the previous time step are used in projection decisions for the next time step. So an alternative with the time step as its outer loop, is required. It might not be viable to develop a completely new system for a few projection cases, but if such a system already exists it is a valuable addition to the range of projection options. In the case of GCBM, the CBM-CFS3 would be a compatible system with time as its outer loop. In addition, the CBM-CFS3 is an aggregation system which can be useful for projections that do not require spatially explicit outputs.

Suite of Tools to Support Projections



Inputs Generator for Deforestation Baseline

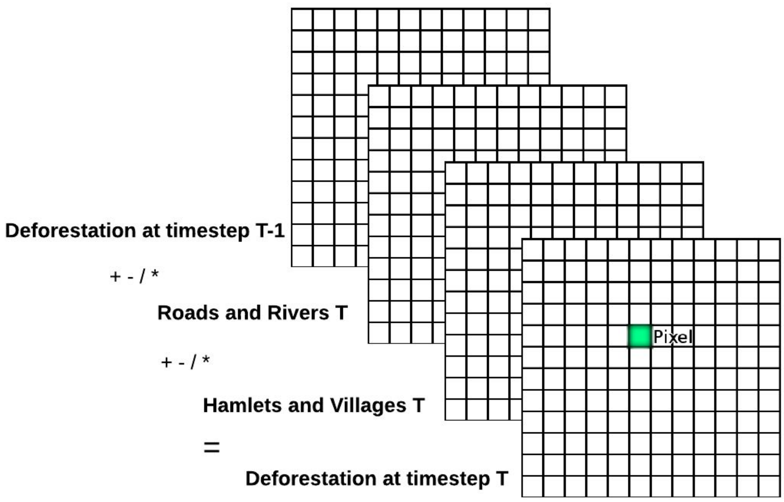
The Inputs Generator for the Deforestation Baseline combines several spatial layers with an algorithm to create a layer that projects deforestation. The Input Generator is used to generate all the inputs FLINT needs to run projections of emission estimations. When the Inputs Generator is used to project a deforestation baseline, spatial layers of landscape features are entered into an algorithm that predicts where the next deforestation will take place. This requires that the Input Generator can perform raster functions that can be divided into four types.

The Inputs Generator executes four types of operations in a specific order. The Inputs Generator is a raster calculator that uses other rasters as input to generate new rasters using an algorithm. This requires first that the layers required are loaded or created. The four operations necessary will be explained in more detail below. The definition provided here will become clear through the detailed examples that follow:

1. Triggered by: These are the layers necessary to determine the value of the raster. The actual value might be obtained through a complex algorithm that uses values from other layers as inputs.
2. Trigger for: The resulting layer might be used as input into other algorithms. The order of the calculations is determined by these dependencies.
3. Transition rules: Once a new layer has been calculated, it might have feedback effects on other layers.
4. Projection of the input layer: Once the layer has been created, it might change over time. An algorithm to project situation at time step T+1 is required.

Triggered by: Deforestation is dependent on a series of observable features in the landscape. The best proxies to predict deforestation are the subject of extensive research. Most of these proxies are based on spatial information: e.g. location of hamlets, roads, types of agriculture, most recent deforestation, etc. These spatial features are used by an algorithm to predict where the next deforestation is likely to happen. The algorithm might require pre-calculations like distance to nearest hamlet as opposed to hamlet alone. Alternatively, rather than developing an algorithm based on field research, it might be possible to develop the algorithm based on machine learning.

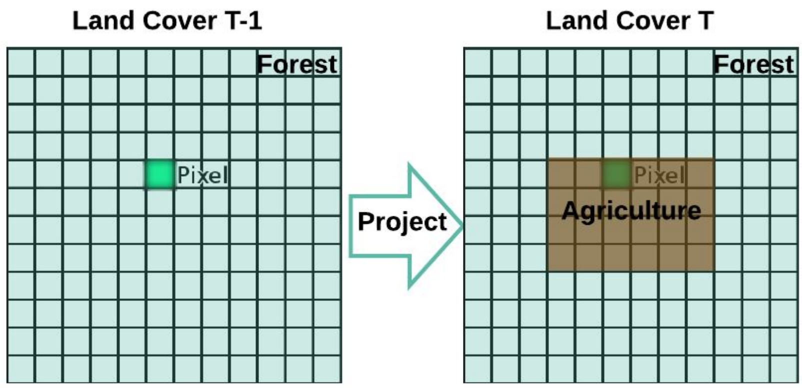
Algorithm to Predict Deforestation Based on Spatial Layers



Trigger for: Deforestation is itself a predictor for future deforestation. Some research indicates that deforestation can best be predicted on the basis of where previous deforestation happened. So the deforestation observed at time T-1 is a valuable input into the algorithm that predicts deforestation at time T. Deforestation might, but is unlikely to be a trigger for other disturbances or events. It is possible that under specific policy scenarios deforestation is used as a trigger for other interventions. This might fit in one of the classifications used in the next section “Inputs Generator for Land Management Policy Scenarios.”

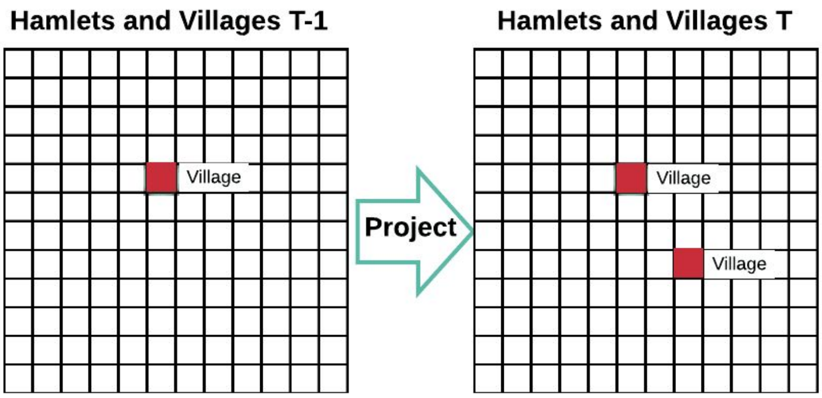
Transition Rule: Deforestation requires various changes including to the land cover or the tree age layer. Once the algorithm has identified the pixels that are projected to be deforested at time T, several follow up actions are required. These actions need to be defined as part of the deforestation intervention. Once an area has been deforested, the land cover has to change. It is possible that the area will be used for agriculture, so the land cover layer needs to be adjusted. If the area remains forest but different trees will regrow, the layer with species needs to be adjusted. Alternatively it is possible that the same species will regrow, in which case only the age layer needs to be adjusted. Other changes might be required too, at this point, it is only important that the design of the Inputs Generator should include a function that allows the user to define transition rules.

After a Deforestation Event, the Land Cover Layer is Adjusted



Projection of Input Layer: The input layers of the algorithm also need to be projected to serve as inputs in the next time step. To calculate the deforestation at time step T+1, the deforestation at time T is required, but also the roads layer and the hamlets layer at time T. Since the number of roads or villages might change over time, each of the input layers also requires a projection algorithm. Fortunately, many layers are constant or are going through marginal or simple changes that can easily be applied to rasters. This particular example is slightly more challenging as projecting roads and hamlets might not be complicated to capture on a raster layer but the development of a robust algorithm might be a challenge.

Projection of the Input Layers for Algorithm Calculating Deforestation



Inputs Generator for Land Management Policy Scenarios

To project policy scenarios, the Inputs Generator must be able to manipulate various layers in a systematic manner. The same four types of raster functions must be performed on layers that can be classified in ten major types. In addition, some functionality is required to deal with under defined operations.

Further experimentation should determine whether non-spatial data should be managed by the Inputs Generator. Policy scenarios might require spatial and non-spatial data. The latter can be ingested by a FLINT-based system and FLINT will integrate them so they can be applied in a spatially explicit manner. The projection of the non-spatial information might require that the information is rasterized, so it can be manipulated by the Inputs Generator or it might be easier to keep the data in a non-spatial format and project it in a different way. At this point the design is not taking the non-spatial data projection into account. This requires further testing.

The Inputs Generator might not add value when specific tools are already used to make policy projections. Some land management policies like forest planning have traditionally been supported by robust tools. When such tools exist, the outputs of these tools should be used instead of duplicating these functions in the Inputs Generator. Some of these tools produce spatially explicit outputs (e.g. Patchwork by Spatial Planning Systems) that can be directly used by FLINT. Other tools (e.g. Woodstock by RemSoft) generate spatially referenced information that might have to be made explicit using intermediate scripts or tools (as explained above).

The Inputs Generator must be able to execute the same four operations to a range of layers. As discussed in the previous section, the Inputs Generator must be able to execute the four functions. The same functions are required for policy projections. The definition for projection of inputs, has been slightly modified to “projection of this layer” because each layer can be an input layer for any other layer. So the four functions are:

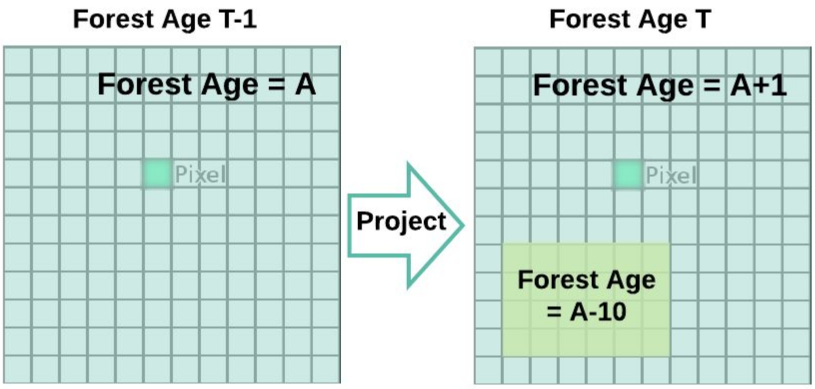
- Triggered by: The changes in this layer are triggered by other spatial layers. The input layers might be used by an algorithm and the influence is not necessarily at the same spatial location. (e.g. hamlet influences deforestation elsewhere.)
- Trigger for: The layer may be used by an algorithm that calculates another layer. Updates of dependent layers might better be done after the input layers have already been updated.
- Transitions: When changes to this layer occur, other layers might need updates too.
- Projection of this layer: The layer is updated at each time step using a specific algorithm or formula.

There are ten types of spatial layers that need to be defined in the Inputs Generator. Below the four operations for each of the ten types of input layers is discussed briefly.

1. Fixed: Soil type, ecological zone, etc.
    a. Triggered by: These are fixed features of the landscape
    b. Trigger for: These can be useful inputs into algorithms triggering disturbances but they are most often secondary
    c. Transitions: These layers are constant and do not change over time
    d. Projection of this layer: These layers do not change over time
2. Fixed Change: Average Temperature
    a. Triggered by: These features are not constant but have a defined change for every time step (derived from other sources such as climate projection models).
    b. Trigger for: While they can be an important input for emission estimations, they are not common as triggers for disturbances. There might be some algorithms for pests and diseases that use temperature as an input condition.
    c. Transition: This layer is not affected by disturbances, so there are no transitions
    d. Projection of this layer: The layer is updated at every time step based on a fixed formula (or based on climate layers provided externally).
3. Fixed Change Plus Reset: Age of vegetation, time since last disturbance
    a. Triggered by: These features have a fixed change at every time step but in addition can be reset as a result of disturbances, e.g. age of an undisturbed forest increases but is reset to zero when the forest is harvested
    b. Trigger for: These features are common inputs into other disturbance algorithms: e.g. age is used as input into a wide range of management decisions
    c. Transitions: There are no transitions after changes have been made to this layer
    d. Projection of this layer: Each value is increased by one at every time step
4. Fixed Change plus Variable Adjustment: Relative age of the vegetation
    a. Triggered by: These events only start after a disturbance and have a defined change for every time step
    b. Trigger for: These layers are often used to trigger disturbances. Relative age has a specific application as input into growth curves to determine the growth rate at every time step.

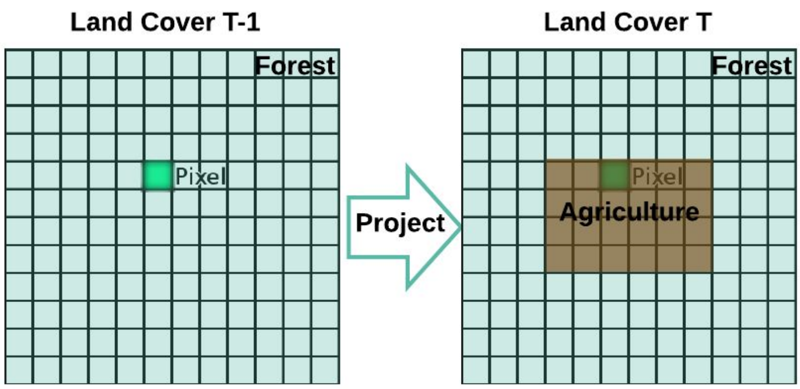
c. Transition: There are no transitions after changes have been made to this layer
d. Projection of this layer: as with age above, all cells are increased with 1 for each time step

Annual Increase of Relative Age Combined with Variable Adjustment



5. Disturbance Change: Land cover, type of vegetation, species, management regime, etc.
   a. Triggered by: These features only change as a result of a disturbance (natural or planned)
   b. Trigger for: These are important input layers as most disturbances are specific to a combination of these layers
   c. Transitions: Changes to this layer might require changes to a range of other layers, e.g. land cover changes might also require changes to the species layer and the age layer.

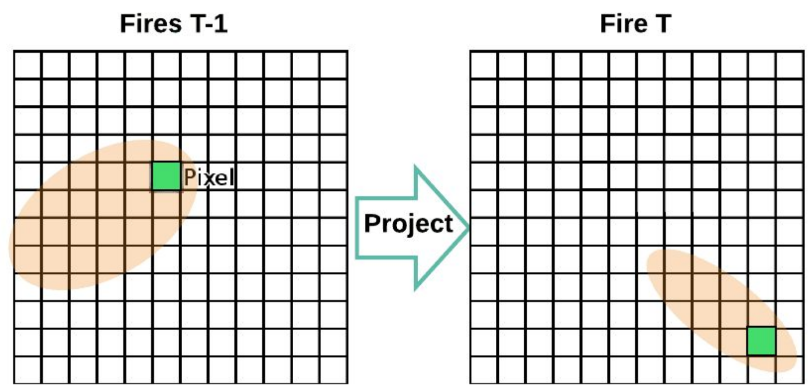
Changes Require Adjustments to Other Layers Too



   d. Projection of this layer: These input layers only change as a result of disturbances so once the transition after disturbances have been applied, no further update is required.
6. Probabilistic or Random Change: e.g. Some types of fires

a. Triggered by: Some disturbances are not triggered by local circumstances but are allocated to the landscape in a random or probabilistic manner (e.g. due to lightning strikes). The points are generated by providing a raster with “available” points and selecting points from that pool using the random or probabilistic function (e.g. lightning strike probability as a function of topography). Additional calculation might be necessary, e.g. wild fires require a random point and in addition a probabilistic direction, probabilistic size, (probabilistic?) shape, and the allocation can be made more complicated by introducing sloop, land use boundaries, existing vegetation conditions, etc.
b. Trigger for: These layers could trigger other events, such as salvage logging of fire-killed trees. If they do, it can be useful to define a raster tracking the event, e.g. a raster with “time since last forest fire”.
c. Transition: Various changes could be required as a transition after a probabilistic event. For example, the relative age of an area affected by pests could be reduced to represent the reduced growth, or the species might change as a result of natural regrowth after a fire. It can also be useful to define a raster with vegetation features e.g. “snags following a fire” for processing other events such as subsequent salvage logging.
d. Projection of the layer: Since the initial location is random or probabilistic, the process of allocating the next disturbance is repeated at every time step and does not have a mathematical relation with the previous occurrence. Although the allocation of the ignition point may not have a relation to the previous years’ events, the spread of the event (e.g. fire) can be affected by the presence of previously burned areas.

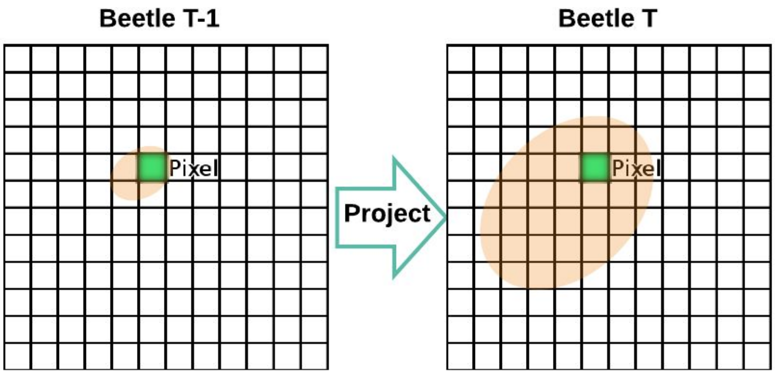
Probabilistic Events Over Time



7. Probabilistic and Variable Change Over Time: Disease, Pests
   a. Triggered by: Multi-time step disturbances can be triggered by specific landscape features (e.g. drought) or they can be allocated to the landscape using a probabilistic function or both (e.g. beetle infestation triggered by probabilistic occurrence in areas with certain tree species and ages, certain elevation, certain average temperature etc.)

b. Trigger for: Multi-time step disturbances are triggers for their own occurrence in the next time step. They might be triggers for other events.
c. Transition: As with random disturbances above, various changes could be required as a transition after a multi-time step event. For example, the relative age of the affected area might be reduced during a few years but if disturbance continues for several years, the trees might be considered dead e.g. a single defoliation event may reduce growth while multi-year defoliation can kill trees). It can also be useful to change species or just return age to zero and/or increase the raster with "snags" for processing other events.
d. Projection of the layer: While the first year's occurrence is based on a trigger or probabilistic function, during the next years the area disturbed can be redisturbed (e.g. multi-year defoliation) or the areas increases based on an algorithm representing the spread of the insect disturbance

Projection of Layer with Probabilistic Start and Change Over Time



8. Biomass Based: Growth and related carbon stocks, is a function normally executed by FLINT. To generate the inputs for the FLINT, biomass can in some cases be approximated by (relative) age of the vegetation. However, biomass and in particular merchantable volume, is so important for planning that most forest management tools duplicate this calculation that is also executed in FLINT. Consistency between the two systems has to be maintained by using the same growth curves in the Input Generator and in the FLINT.
   a. Triggered by: These are growth functions that are triggered every time step.
   b. Trigger for: Many forest management disturbances are triggered by biomass or age thresholds
   c. Transition: There are no transitions
   d. Projection of this layer: Every year the wood volume is increased according to the applicable growth curve.
9. Dead Wood Based: This type of layer includes stem snag, residue, and other dead wood volumes. These carbon pools require complex modeling by FLINT and can thus not easily be duplicated in the Inputs Generator. Since some management decisions require these dead wood volumes as triggers, they can be approximated by creating a layer that contains a value entered as part of a transition rule. This approximation is manageable because the layers once created do not need to change over time.

    a. Triggered by: These are values (volumes) that are triggered by disturbances (windfall, fire, harvest) but they do not change over time.
    b. Trigger for: Some forest or fire management planning might require dead wood volumes
    c. Transition: There are no transitions
    d. Projection of this layer: The values do not change over time
10. Dead Organic Matter (DOM) Based: This layer includes litter as well as dead wood. The changes require complex modeling by FLINT. So it is very hard to approximate them. The use of a compatible system with space as its inner loop might be required if DOM is required as part of the inputs for planning.
    a. Triggered by: DOM volume is influenced by a range of factors that include disturbances as well as changes at every time step.
    b. Trigger for: Some planning might require DOM volumes.
    c. Transition: Various transitions might be required but it is not efficient to calculate them outside the FLINT.
    d. Projection of this layer: The factors influencing this layer require a FLINT-system to model, so it is not useful to generate them as a FLINT-input

The Inputs Generator requires functionality to allocate under-defined events. Applying disturbances might require information in addition to spatial layers. Many of the types of spatial layers listed above are used to trigger disturbances. But the spatial information might not be sufficient to decide where disturbances should be located. In other words, the information available is resulting in more potential locations where a disturbance could happen than the total surface that should be affected by the disturbance. One example is firewood collection. The regions in which firewood collection can occur can be defined based on access and proximity to settlements, but the exact location of firewood collection is not likely predictable. One could say that the location is under-defined. There are a number of tools that could be developed to deal with this type of situation:

1. Adjacency and distancing of disturbances: This functionality requires the Inputs Generator to cluster disturbances in a particular area and/or keep a minimum distance between the clustered disturbances. (e.g. Harvest of 200 Ha plots and plots are surrounded by a buffer of at least 10 km)
2. Distance to spatial features: This functionality requires the Inputs Generator to apply disturbances at a minimum and/or maximum spatial feature (e.g. degradation applied between 0 and 1 km on both sides of all roads.)
3. Sort values: (e.g. Apply to oldest stands) This functionality requires the aggregation of the eligible raster information into a table format that can be sorted. The sorted information will provide a range of values to which the disturbances should be applied, e.g. forests aged between 45 and 80 years of relative age. It is possible that the location remains under-defined even after the required values have been sorted.
4. Additional Functionality: More complex functionality might be developed to deal with under-defined disturbances. The application developed by the Bureau du Forestier en Chef du Québec, applies a simulated annealing algorithm to find the best location for the disturbance.

The Inputs Generator has predefined regimes and algorithms to facilitate data entry. The user will only need to provide high level information to run the projections. To enable high level data entry, a range of regimes and algorithms should be predefined. The data entry process would roughly follow the steps below:

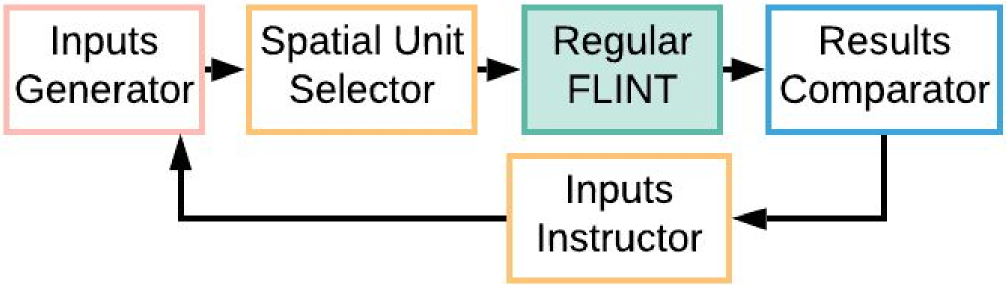
1. Definition of algorithms (several options might be available) and regimes (e.g. harvest or fire or drought, etc.)
    a. The input generator has a number of regimes and algorithms that can be selected during the data entry.
    b. Each algorithm or regime can have parameters that are obtained through data entry, e.g. number of ha harvested or reforestation distance from hamlets, etc.
2. Enter high level instructions: The only data the user actually has to enter. Below is a random example of an instruction set:
    a. Harvest all stands older than 45 years with a maximum of 500K Ha.
    b. Deforestation: Unplanned deforestation using Algorithm 2
    c. Reforestation: Replant 20K Ha of forest land at least 10 km away from the nearest hamlet and road
    d. Fire Regime: Apply Fire Regime number 3: i.e. increasing surface burned upto twice historical average
    e. Harvest Snags after fire: Harvest snags and deduct total surface from 500K Ha maximum harvest per year
3. The Inputs Generator will define the necessary layers based on triggers between these disturbances in addition to the layers FLINT needs for regular operations
4. The Inputs Generator runs all disturbances and regular projections for the first time step (i.e. going from time step today to today + 1) starting with the layers that are inputs for other disturbances.
5. Repeat 4 above for every time step required for the projection

Inputs Generator for Emissions Optimization

For optimizations based on iteration of scenarios, the Inputs Generator must be able to read parameters and high level instructions automatically. The data entry process outlined above needs to be automated for optimization processes.

Optimization involves additional tools that interact automatically. As discussed above, optimization requires a feedback loop from the FLINT results back to the Inputs Generator. The FLINT results are fed to a Results Comparator (discussed below) that checks whether the results meet the objective of the optimization. The result is handed over to the Inputs Instructor that uses the output of the Results Comparator to provide instructions to the Inputs Generator. The instructions should be similar to the definition of disturbances that are used for land management policy scenarios discussed above. The information received from the Inputs Instructor should preferably be read automatically. If the optimization cycle can be automated the number of cycles can be increased and hence more cycles will result in more accurate optimizations.

Tools Contributing to an Optimization Cycle



The settings of the Spatial Unit Selector do not need to be automated. The spatial unit selector can select a group of pixels that is representative for the changes to the whole landscape, so while the Inputs Generator changes the input files for the whole landscape, FLINT will only calculate the carbon fluxes and stocks for the pixels selected by the Spatial Unit Selector. Since this selection remains the same for each optimization cycle, the filter can be installed at the start of the optimization process and does not need to be updated every cycle.

The Inputs Generator should be able read parameters and high level instructions from the Inputs Instructor. Some optimization processes will require changes at the level of types of interventions that are normally selected through high level instructions. Other optimizations focus on refining the parameters of the regimes or algorithms (e.g. number of ha harvested per year). Automating the data entry requires the Inputs Generator to be able to read both parameters of regimes and algorithms as well as high level instructions. Developing the full functionality could be challenging but initial steps are easily implemented and should be tested.

Inputs Instructor

The Inputs Instructor uses feedback from the Results Comparator to adjust the parameters and high level instructions for the Inputs Generator. As discussed above, the Inputs Generator has predefined disturbances the user can select from. These disturbances are based on pre-defined regimes and algorithms. And these regimes and algorithms require parameters. Each of these inputs needs to be generated by the Inputs Instructor but for each cycle the inputs should be adjusted based on the feedback received from the Results Comparator.

The Inputs Instructor can adjust variables within the limits agreed at the start of the optimization. Not every optimization has to consider every possible option. At the start of the optimization process a number of limitations are set:

- The parameters and high level instructions that will be varied in this optimization will be selected. Not necessarily all variables will be optimized.
- The range of values for each parameter and options for each high level instruction will be established

- The mechanism to select from the available range will be selected from a range of options. These options can be elaborated over time but they will cover deterministic and stochastic approaches.

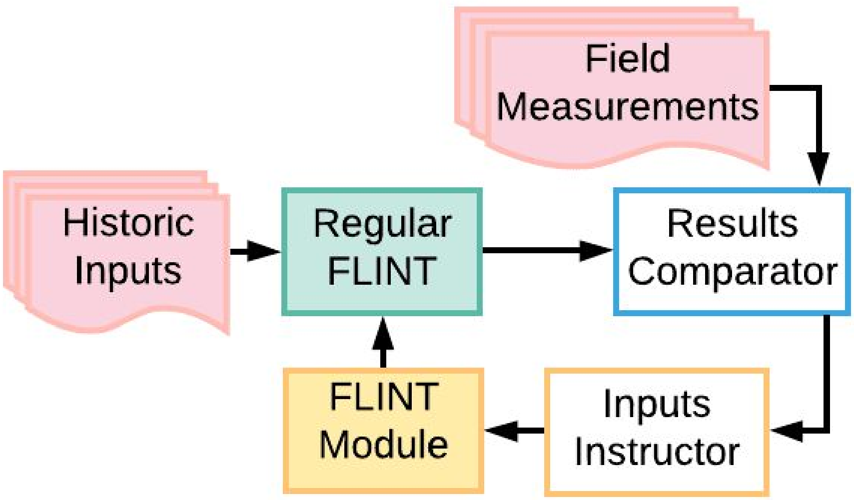
Deterministic approaches adjust variables based on precise correction instructions received from the Results Comparator. If the purpose of the optimization is to plan a harvest regime that keeps the long term carbon stock equal to the current levels, the first optimization cycle might be based on a rough estimate entered by the operator. The Results Comparator establishes that the result of the FLINT run indicates that the carbon stocks are dropping by about 15% every year. The Results Comparator feeds this information to the Inputs Instructor. This precise information is transformed by an algorithm in the Inputs Instructor into changes to the harvest rate and possibly to the species planted and the management regime. The new values are fed to the Inputs Generator which in turn triggers a new FLINT run. The Results Comparator now establishes that total carbon stock is going up by about 5% every year. Again the Inputs Instructor uses this value to adjust the parameters of the Input Generator and the next optimization cycle starts. The use of deterministic observations tends to optimize inputs in fewer cycles. There are two important disadvantages: First, the approach can get very complex very quickly as it requires an excellent understanding of the mathematical relationship between the observed results and the available input variables. Second, the mathematical relationship for the first estimate might have a bias that is repeated through every cycle and leads to a suboptimal result.

Stochastic approaches generate random values for each variable based on boolean feedback from the Results Comparator. If the purpose of the optimization is to get the highest possible value for the sum of the percent increase of yield, biodiversity, and carbon stock, the mathematical relationship between outcome and input is too complex to develop a useful formula. The solution is to start the process again based on a rough estimate (the prior). The values of this estimate are stored in the Inputs Instructor as a base line. Then the Inputs Generator and the FLINT (and tools for yield and biodiversity) are run. The Results Comparator stores the results (i.e. sum of increase of yield, biodiversity and carbon stock) as the baseline. And sends a “yes” signal to the Inputs Instructor. The inputs instructor keeps the current values and generates new values for all variables within the available range. The values are passed on to the Inputs Generator and the FLINT (and other tools). The results are compared with the previous run by the Results Comparator and imagine they are worse than the previous run. The instruction to the Inputs Instructor is a simple “No”. In response, the Inputs Instructor will discard the values and keep the values of the previous run as the best to date. Then new random (or semi-random) values are generated and the process is repeated. This approach does not require a mathematical relationship between output and input values so it can deal with complex systems. It also avoids bias. It might however require many cycles to reach an optimum. Methods are available to inform the selection of new parameters based on the performance of previous selections. This requires not only “yes / no” feedback but also information about the direction and magnitude in the changes of specific output variables.

A function similar to the stochastic Inputs Instructor has already been developed for the Monte Carlo Uncertainty Module. This Module compiles a list of inputs (coefficients, variables, parameters) and uses Probability Density Functions (PDFs) to generate values for each of these inputs. The PDF can be specified by the user for each input. Additional functionality will be developed in future versions of the module.

The Inputs Instructor is a separate tool as a similar process might be used for other processes. The functions of the Inputs Instructor might be geared specifically towards the requirements of the Inputs Generator (i.e. parameters and high level instructions specific for a particular optimization). A same tool with adjusted variables and ranges can be used for other optimizations. A simple example is module calibrations. A FLINT module has a number of parameters that need to be tuned to the situation in a particular location. There are field measurements for specific points in time for specific locations. The system uses the historic inputs to run the FLINT. The results from FLINT are compared with the field measurements. The Results Comparator sends a boolean or measured observation to the Inputs Instructor depending on whether a stochastic or deterministic approach is used. Next the Inputs Instructor will adjust the parameters of the module and the same input values are re-run through the FLINT and compared with the same field measurements. Again the difference is compared with the previous parameters and the next optimization cycle is started. The process is repeated until the required accuracy is obtained.

Use of Inputs Instructor for Other Optimizations



Results Comparator

The Results Comparator compares FLINT results and possibly other outputs with an optimization objective and produces a boolean or a measured output.

The Results Comparator is a separate tool as it must be easy to manipulate with little expertise. The Results Comparator compares FLINT results with the ideal that the optimization wants to achieve. This is the key policy question that will be changed as part of an ongoing policy dialogue. So the policy question (optimization objective) should be easy to change even by a person that only has limited coding knowledge. Keeping the tool separate from other tools reduces its complexity and allows operators with little training to make changes to the tool. This is technically feasible as the inputs and outputs of the Results Comparator are fairly simple, so interfacing with other tools will not be complicated. An additional advantage of having a separate tool is that it can be rewritten in different coding languages in line with the operators preference and the particular needs of the optimization. For example, a simple graphical user interface or a web-based input tool could be developed.

The Results Comparator can draw from the existing Reporting Tool. The Reporting Tool queries the FLINT output database and performs calculations on the results and reformats the results into simple output tables. This functionality could be replicated as a start for the Results Comparator.

Spatial Unit Selector

To increase efficiency, only a selection of all pixels need to be simulated by FLINT to get results that can be compared with the objective of the optimization in the Results Comparator. Running all pixels for each iteration of the optimization process is time consuming and will take away computer resources from the number of iterations that can be performed. Hence is it possible to select specific pixels that will produce representative outputs. If this is done, then an additional requirement will be a tool or script that allows for easy combination of existing results with those of the newly simulated pixels.

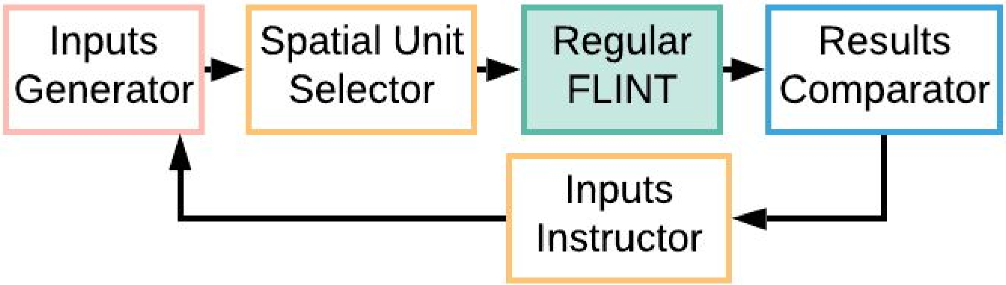
The Monte Carlo uncertainty analysis already has comparable functionality. The FLINT uncertainty module already has a simple random sampling approach to selecting pixels for a Monte Carlo simulation and more advanced probability-based sampling methods are planned including Restricted Random Sampling and Unequal Probability. More information is available in the Design for Monte Carlo and Propagation of Error Uncertainty in the Full Lands Integration Tool.

The optimization cycle can be made more efficient by only running a selection of pixels through the FLINT. FLINT is the most computer intensive part of the optimization cycle. Running a simulation for a regular country can take hours or even days as many millions or even billions of pixels need to be processed. And these simulations need to be repeated at each cycle of the optimization process. Since optimization does not require the output of every individual pixel, it is possible to select a representative sample of pixels that provides an output that approximates the actual outputs with a known error margin. If this error margin is many times smaller than the difference between the optimization results, the selection of pixels can result in considerable efficiency gains.

To ensure that the sampling error is several factors smaller than the optimization differences, the number of pixels in the sample can be adjusted. The approach to manage the error margin proposed in the Monte Carlo Module Design Document, is to do a pre-run simulation from which the variance of all pixels can be calculated. The variance is used to determine the sample size.

Open Source Building Blocks

Some basic versions are already available for each of the tools proposed to run spatially explicit projections. The code base will probably be relatively simple compared to the development of algorithms that predict changes to the landscape. Machine learning tools might be useful to assist with this challenge.



Each of the tools used for FLINT-based projections can build on some existing or evolving code base. The Input Generator can build on the Raster Calculator in QGIS. The Input Instructor can draw heavily from the inputs generator for the Monte Carlo Uncertainty Module. The Results Comparator can build on the Reporting Tool and the Spatial Unit Selector can build on the tool with a similar function from the Monte Carlo Uncertainty Module. Surely more tools and libraries are already providing useful functionality to build a minimum viable product to test the concepts proposed in this document.

The algorithms required are however not readily available and might need to be developed using research or machine learning. Each of these tools requires some algorithms to predict how certain features in the landscape will evolve in the future (e.g. hamlets). While various research projects are looking into these algorithms, applying machine learning methods might be a faster method to obtain results.