GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für Forstwissenschaften und Waldökologie
Burckhardt-Institut

AWF
Waldinventur • Fernerkundung

Abt. Waldinventur und Fernerkundung, Büsgenweg 5, 37077 Göttingen
Chair of Forest Inventory and Remote Sensing

Prof. Dr. Christoph Kleinn
Director

Tel. +49 (0) 551 / 39-3472
Fax +49 (0) 551 / 39-9787

ckleinn@gwdg.de

Göttingen, 30.12.2012

Ihre Nachricht vom | Meine Nachricht vom | Ihr Zeichen | Mein Zeichen

Observations and comments on the lead questions for "Issue Paper 3: MRVdesign …."

This submission is an individual submission by Christoph Kleinn, not representing any institution.

What follows here, is a series of observations regarding "Issue Paper 3: MRV Design …" that are based on the author´s experience in planning, implementing and analyzing forest inventories for quite a while in various regions of the tropics (and "non-tropics" as well).

This experience refers about equally to the various facets of forest monitoring, and that includes statistical sampling, forest mensuration, forest inventory, remote sensing, statistical modeling, capacity building and maintenance, the role of information in decision-processes, international networking, etc.

The language used in my observations is what has been used in forest monitoring sciences for decades - and is not necessarily entirely compatible with the language that the international processes have developed over the past years for their view on forest monitoring.

I am commenting on the questions raised to my best knowledge.

I am aware that my knowledge about the political processes where MRV is being part of is not complete. I, therefore, apologize for any ignorance; and kindly ask to concentrate in what follows on the scientific essence of my observations. Any hint to misunderstandings on my side is very welcome.

This submission starts with comprehensive methodological observations and comments on the lead questions raised; and ends with suggested elements of an "accuracy" assessment for carbon accounting.

Observations (lead questions in blue, my observations in black):

Issue Paper 3: MRV design: carbon accounting of Emission Reductions Programs, non-carbon, community role; and registries. Some key questions we seek ideas or advice on:

Q1: Considering the circumstances of the CF (piloting, in REDD+ phase 2), how accurate does the measurement and monitoring need to be?

- **Should the same minimum level of accuracy be required for all programs? If so, what level of accuracy should be required? What criteria would help determine such a level (e.g., how feasible a level is for most REDD+ countries to achieve; or a level that allows the fungibility of credits with other climate program standards)?**
- **Or should the CF be flexible, not prescribe a minimum level of accuracy, and be responsive to country circumstances?**
- **If so, what are the appropriate country circumstances?**

(Relatively) short "answer":

- Once "accuracy" is clearly defined as a common quality label for "units of carbon emission reduction" or for "units of sustainable forest management", it appears reasonable to have the same minimum accuracy requirements for all programs and for all regions.
- However, I doubt that "accuracy" can be defined in an equally scientifically sound and operational manner. It will very likely be a political decision and needs to be broken down into different elements.
  Various details will need to be defined in order not to open doors for easy mis-interpretation.
- If the major goal is to establish a market, homogeneous products maybe most important (= corresponding to equal levels of accuracy).

  If the major goal, however, is to enhance the forest resource and foster sustainable forest management, it will probably make sense to define specific goals under the given circumstances (=adjusted levels of accuracy, adjusted courses of action).

  Criteria could include, then: pattern of forest resource; pattern of change in forest resource; progress in establishing functioning forest research institutions; progress in establishing a permanent forest monitoring system; willingness of the government to invest resources into the enhancement of the forest resource.
- Focusing on carbon emission reductions: with the same monitoring efforts per unit of area, in different regions very different levels of "accuracy" will be achieved.

  "Accuracy" depends on the circumstances (that is: the forest conditions, the magnitude and spatial distribution of changes, the methodological approach, etc.).
- For large area forest monitoring, there has never been success in defining a general "good" level of precision of estimation. It is all based on convention and tradition.

  Therefore, we usually find "even" figures like 5% or 10% or 20% when the "level of accuracy/precision/ certainty/reliability" is being defined.

  Nor has it been achieved to scientifically establish the relationship between improvement of policies through improvement of the data and information base.

  At the end of this submission there is suggestion how to possibly deal with this situation.

Mostly, it appears, that the “happiness criterion” applies: if the contracting entity is happy with the results and considers them useful for the envisaged purposes, everything is o.k. and acceptable

And because it is difficult for the non-expert to understand in detail the meaning and principle of statistical errors, the range of errors that lead to “happiness” is wide.

Wider thoughts and observations:

- From a market and negotiating point of view, a general “quality label” of the product “carbon emission reduction” (and of the achievements regarding the other “products”) is probably required.
- The question about “accuracy of the measurement and monitoring” (as formulated in the lead question), however, is difficult to understand. What is at stake is probably a measure of reliability of the results and not so much just of the data/information collection process!
- A first step needs to be to very clearly define what is meant by “accuracy” and to which concrete products it refers. It is difficult to discuss further consequences of different “levels of accuracy” if it is not clear what the meaning of accuracy is.
- Regarding carbon emission reductions, accuracy measures may refer to estimates of areas (activity data), to estimates of carbon density (emission factors), to estimates of changes (in area / in emission factors), to the models used, for all pools together or for single pools, and to the overall result. It needs to be clarified what the target variable actually is – or if accuracy requirements refer to more than one variable.
- It needs also to be clarified what the geographical unit of reference is for any statement of a “level of accuracy”: for the entire country; for sub-national units, etc.
- Because all such calculations are based on a set of observations (including remote sensing, field samples, carbon models, a-priori information), one may also wish to analyze the role of each of these data sources in the course of the error propagation.
- Once these technical issues are clarified in principle, it will be necessary to define clearly the terminology.
- The term “accuracy” is clearly defined in sampling statistics. It is not so clear to me how it is used in the given context:

  When a sampling study is done, we distinguish precision and accuracy. “Precision” refers to the variability of the estimates around the expected value: “high precision” means that an imagined frequent repetition of the same sampling study would result in quite similar estimates. “Accuracy” refers to bias.
- When a remote sensing based mapping study is done, the concepts of “precision” and “accuracy” as of sampling studies do not apply. It is frequently stressed that remote sensing based mapping is not a sampling study but a full census and that there be no sampling errors. While this is entirely correct it does not mean that the product is eventually error free: like in any sampling study as well, there are measurement / observation errors also in remote sensing studies.

  The challenge to quantify the “quality” of a remote sensing based map is intensively researched; but there is – to my best knowledge – no generally accepted approach yet. What is frequently used is the simple measure of “overall accuracy” which gives a general idea of the quality of the product. It is essentially an estimate of a mixture of classification errors and geometric errors and has a completely different character from the standard error of estimation as of a sampling study.

That “overall accuracy” can only be recommended for a comparison and for an assessment of the quality of a map product if some more methodological details are defined (including the number of classes distinguished, the sample size and the data source for ground trothing, …).

- All studies that produce results on carbon stocks and carbon stock changes rely on models to determine “carbon per area unit”.

  The term “measuring carbon”, as frequently used, is somewhat misleading from the point of view of empirical sciences because forest carbon per area cannot be “measured”; rather, all is model based estimation / model based prediction.

  The available models have varying quality and introduce a – probably considerable – variability: using different models will produce different results; this is well known but difficult to predict / quantify.

  It is not possible to produce – with reasonable efforts – a “true value” for carbon stock on a forest plot area of, say, 1000m²; not even for above ground carbon and lesser so for the other carbon pools. But these quite uncertainly determined values are then used for all extrapolations and also for remote sensing based regionalization. In some cases, ground spatial resolution of satellite imagery used for the assessment of carbon stocks is much larger (like e.g. 250m and 500m for MODIS imagery) – and it is beyond the imagination of this author how to reliably determine the ground truth of carbon stocks for such reference areas!

  All reference values for carbon carry a considerable error.

- It is probably worthwhile and instructive to look at results of forest monitoring projects on national level and evaluate the experiences (NFI=National Forest Inventory).

  FAO has long-standing expertise in that field through (1) compiling forest information on national level for the Global Forest Resource Assessment (FRA Programme) and (2) in Supporting countries in setting up a National Forest Monitoring and Assessment System (NFMA Programme). Data from (1) are frequently used as reference also for comparison in scientific studies; however, it must be emphasized that these figures eventually are political figures and not for all countries are they substantiated by methodologically sound forest monitoring studies.

  Own experience: it may happen that, when asked about the methodological approach and the monitoring protocol for a national figure on forest cover, all the responsible country correspondent informs is that “those figures result from a remote sensing study”.

  It is known that not many countries have permanent NFIs that allow exact estimation of changes over time. But for many countries there are, in the meantime, (and thanks to FAO-NFMA and other projects of technical assistance) statistics-based NFIs where estimates can be reported with error margins

  Example 1:

  The 2nd German NFI (entirely sample based, no remote sensing integration) estimates forest area and growing stock per hectare (in terms of wood volume) for the entire country with a simple standard error of 0.7% and 0.5%, respectively (total area of Germany ca. 357´000km²; sample size ca. *n*= 45´000 field cluster plots).

  The cost for the German NFI was several million Euros.

  For the relatively small federal state of Saarland the values are 8.4% and 10.7%, respectively (total area of Saarland ca. 2´570km², sample size ca. *n*= 163 field cluster plots).

  Example 2:

In a research project in Burkina Faso, we carried out – as a side product – an NFI in order to produce large area land use data as an input to large area climate scenario modeling. Very difficult exercise as the distinction of forest and non-forest is not trivial given the particular tree vegetation types encountered in the drier regions of Burkina Faso. With our "low sample size and low budget"-approach, forest area and above ground carbon were estimated with a relative standard error of 9.9% and 18.3%, respectively - from an entirely field based sample of *n*=53 very large field cluster plots and a total budget of about Eur 50´000.

- In these examples, the simple standard error is given as an easily interpretable measure of precision that is unambiguously defined (still, some methodological observations can be made; for example: measurement errors and model errors are not built in).
- The question is: what precision of estimation is necessary for good decisions. That question is not answered in general terms in the forest monitoring domain.

  And there is surprisingly little research on it.

  Many publications (deliberately not referenced here) start with the statement "Good information is necessary to support informed policy processes"; and then a lot of efforts are done to improve the information from field sampling and / or from remote sensing, but there is hardly any publication that concludes with any sort of evidence that eventually the improved information has caused "improved" decisions.

  [That, however, would be essential when attempting to clarify the issue of cost-effectiveness, as addressed in Q3!]
- With the international policy processes on climate change (and possibly also on biological diversity and on desertification) a new era for forest monitoring has started; but not in what refers to the estimation approaches.

  The major difference to the former times is that all of a sudden, the information produced from large area forest monitoring potentially has an "economic value" and that there is an intensive discourse about "best" approaches.
- To my best knowledge there is no scientific solution to the question what a "good" precision for a large area forest inventory is.

  Everything is experience based.

  And it is difficult to understand for decision and policy makers what "information quality" actually means in the context of forest monitoring results.

  In forest monitoring itself, this issue of reporting precision is tackled by commonly publishing not only the point estimates (e.g. mean values like carbon stock / ha) but also the corresponding interval estimates (e.g. percent standard error).

  That is a good starting point to making a proper interpretation of the quality of the estimation.
- There are numerous detail questions which are scientifically not resolved yet. Estimating all carbon pools on ground reference plots, for example, as mentioned above; unambiguously defining what "forest" is (it is interesting that many forest mapping studies do not explicitly spell out the forest definition used): even if minimum crown cover and minimum area are given – this is not (!) sufficient to make it an unambiguous definition: for the same threshold values of, for instance 10% minimum crown cover and 0.5ha minimum size, considerably different figures may be produced for the same region and following methodologically sound monitoring approaches!
- In the above, we were referring primordially to the estimation of carbon. However, large area forest monitoring systems do usually cover between 100 and up to 250 variables. When

additional variables are to be integrated into the monitoring system, the discussion becomes much more complex. Estimating carbon is probably one of the easier exercises – and yet extremely complex.

- It may be worthwhile for the climate change community to more intensively look into the experiences in the field of forest monitoring, in particular on national level.

  For many years, this expertise was apparently not entirely exploited within the forest related UN-FCCC processes.

- It is maybe good to mention that from the 1970s onwards, forest inventories are being developed to multi-resource inventories integrating much more than timber; and in the 1990, NFIs started to also include into their design in a systematic manner for example the assessment of trees outside the forest and of interviews with forest users/owners on issues of forest use.

- The REDD discussions have given the dynamics in forest monitoring research a push; much progress has been achieved above all in the field of remote sensing through new sensors (and huge resources made available by space agencies) and a significantly growing research community.

  However, this submitter cannot see so far a major overall breakthrough. Rather – a very personal view – much continues to be technology driven; and results are sometimes presented less as scientific findings but rather in the form of marketing strategies.

  Sometimes, one wonders whether the focus is really in the first place on the great overall goal (= fostering forests and forest management and by that: considerably reducing emissions at reasonable cost and with as many co-benefits as possible).

- A brief observation on the Tiers 1-3: this is an excellently operational approach that can be checked and tested quite easily, because it mainly looks at the quite easily verifiable input data to the estimation approach: if general/global models are used (tier 1) the results are assumed to be less credible than when local models and a local inventory are implemented (tier 3).

  However, this approach is somewhat unusual, because the quality of the results is rated along the input only. It may well be that in a specific country and under certain conditions the global models fit perfectly resulting in excellent estimates; and it may well be that the local models and inventory are badly implemented and integrated so that the results are of low quality.

  > This may be nicely illustrated by the example of brewing beer: you need very good ingredients to brew a good beer: clear water, excellent quality malt and excellent quality hops. But just these excellent ingredients alone do not make a good beer: there is more required, and that is the expertise how to suitably process the ingredients.
  > Same thing with the input data to forest monitoring: expertise is required to properly process input data to high quality results.

- Some preliminarily concluding observations from the prior comments:

  1. Forest monitoring is complex and truly interdisciplinary. It is not enough to talk about GIS and remote sensing when referring to forest monitoring.

     Capacity building must be much broader than just GIS and remote sensing, the two of which are usually very much emphasized when referring to MRV-related capacity building.

2. Successful forest monitoring systems will need to be embedded into permanent national institutions and accompanied by research (also institutionalized in either forest research institutes or universities).

   All forest monitoring depends also on research – and the data gathered are excellent starting points for relevant research, also to improve forest monitoring.

   This long term aspect of capacity building (i.e. building forest research capacity) appears to be greatly ignored in the ongoing discussions.

- **How can the CF encourage countries to strive for higher accuracy, perhaps over time?**

(Relatively) short answer:

- Higher accuracy (and as a consequence more credibility) are achieved by:
  - Better biomass models / biomass estimation
  - Better field inventory
  - Better imagery and image processing / analysis
  - Better organization of the monitoring exercises, including quality control at all stages
  - Better analysis and reporting and better approaches to integration of data sources
- This is reached by better and accumulated knowledge and expertise; and by more, and more focused allocation of resources.
- South-south exchange of experiences as practiced by FCPF (and also implemented, for example, since 2001 by FAO-NFMA) appears essential and efficient.
- Managing the forest resource (including information procurement on the forest resource) must be understood as essential goal and mandatory to any government of forested countries such that there is also a budget line in the government for forest related research and information procurement.

Wider thoughts and observations:

- It is interesting to see that many nations appear to not fully acknowledge the relevance of the forest resource in itself and the relevance of information on it. There are few countries only (that I am aware of) that formally consider the forest resource as a national asset.
- Huge amounts of money are spent by many governments to collect other important information (for example about national security through secret services, or about taxes, or about the population structure in general); but there appears not to be the insight that the natural resources of a country are equally relevant and require a good information base.
- One might, therefore, look at forest monitoring systems as a standard element of good governance regarding forest related policies.
- Among other measures, improved curricula for the forestry schools and improved forest research programs in the partner countries appear essential.

  Not only monitoring, but sustainable forest management in general may benefit from also (!) investing in forestry related academic education to “build” the future generation of decision makers.

  That is a long term process (typical for sustainable measures …) with little short term successes. And maybe it is that the reason why has not been fostered so far (to my knowledge).

Q2: The Program monitoring system is expected to be consistent with the (emerging) national REDD+ forest monitoring system. What are appropriate criteria to assess this consistency?

I understand "Program monitoring system" here as "ER program monitoring system" in the sense of ER-PINs.

When a REDD+ forest monitoring system is installed it is to me a matter of course that an ER program monitoring system is consistent with it; and, to the degree possible, integrated into one single comprehensive system.

Q3: Are cost effectiveness, and country capacity, important considerations? I.e., should the MF stress a stepwise approach that is comfortable with early-stage approaches to issues (like early steps in developing the MRV system, short of a fully functioning system); or require potentially more expensive, higher capacity minimum approaches? (Recall the short timeframe of ERPAs, through perhaps 2020.)

- Partly commented on before.
- Cost-effectiveness is always important. However, as long as the technical goals cannot clearly be defined (e.g. accuracy level), it is the cost alone that counts and can be assessed. That is why in large area forest monitoring, for example, it is difficult to compare the "cost-effectiveness" of different approaches.
- Country capacity is, in my view, the key to most of the goals that the international processes wish to promote.

  I refer here explicitly to country capacity which goes far beyond project capacity. Country capacity building is a long-term process that focuses on sustainability of knowledge and institutions; project capacity building focuses on the success of a short-term project. Much of what I currently see appears to be project capacity building.
- I have not fully understood in which domains and by which means country capacity is being enhanced and improved by FCPF measures (this is not clear to me even after reading the FY2012 report of FCPF).
- When sustainable forest management is to be fostered in a sustainable manner, there must be a functioning permanent forest administration, a proactive forest research center, and forestry education at technical and academic level.

  It is probably not realistic to expect this set of institutions in each country so that regional approaches may be most promising here. Possibly supported by donors and long-term experts from donor countries.

  What is missing in many countries / regions is a long-term network of observation- and experimental forest plots, centralized data bases / information systems, and institutional knowledge.
- A methodological framework might consider a step-wise establishment of such a forest and forestry related infrastructure.
- In any case, a step-wise approach is indicated for monitoring systems. If only a fully functioning system is being acknowledged, too much time will be lost. And too many chances will be offered to slow down the process.

Q4: Non-carbon values should be monitored as feasible by REDD country Programs (and consistent with the national REDD+ forest monitoring system). How feasible is this for major non-carbon values? What are criteria for assessing feasibility?

- There is a quite some experience in forest monitoring in including non-carbon variables (values?) into the monitoring system; examples are the assessment of elements of biodiversity, of naturalness of the forest composition, of non-timber-forest products, and others.
  Major challenge is to agree on a set of indicator variables (that can be observed in terms of status and changes from remote sensing and in the field) and a model how to analyze these indicator variables towards the target variable (=the non-carbon value).

Q5: Are there best practices for the potential involvement of local communities in the MRV system design and implementation?

- Large area monitoring systems require consistency of the methodological approach so that much training efforts are required to maintain this consistency when overall responsibility or implementation responsibility is de-centralized and partly passed to local communities.
- However, it is important to stress the relevance of protecting, enhancing and sustainably managing the forest resource also on the local level. Involvement of local peoples into field monitoring activities may be one means to achieve that.
- Field teams of national level forest inventories have made the experience that they were the only forestry related “contacts to the government” for local communities. They have the chance to make clear to the forest users that the government cares about the forest resource and they may inform about forest related government programs.
- In national forest monitoring, involvement of local communities is frequently through integration into the field work because they know the area, they know the people and many do also know many tree species. In addition, they can be a very valuable source of socio-economic information about the use of the forest and tree resource, both in the past and what the plans are for the future.

  They can inform about human interventions and the corresponding background and may support insights into drivers of human activities in and around forests and trees.
- Experience shows that such involvement of local helpers requires careful and smart preparation and communication.
- When a functioning system of forest related institutions is in place (forest service, forest research station, forestry technical school, forestry faculty, …) the probability is high that students come from many different regions. These can then be contracted to introduce and implement monitoring work in or close to their home regions. The positive effect of such involvement must not be under-estimated.
- Referring to the formulation of Q5: in the general context of forest monitoring I find it generally critical to talk about “best practices”; there are <u>no best practices</u>, but usually <u>many appropriate ones</u> in the context of forest monitoring.

  By mentioning and writing on “best” or a selection of “good” practices, it is likely that approaches are formally excluded that work well under specific circumstances.

Q6: Is independent third-party verification essential for CF ER Programs; or should countries be able to propose how verification is performed, and by whom?

- Once there is a functioning system of forest related institutions in place (forest service, forest research station, forestry technical school, forestry faculty, …), I would suggest to let the countries propose how to do a methodologically sound verification.
- Given the complex nature of any forest related monitoring exercise, it may be a good idea to always have the national verification teams accompanied by independent external experts.

  Such an approach does also foster the mutual learning process.

Q7: Should a registry of REDD+ activities be required for a CF Program in a country? (The FMT is considering cooperating with others to develop a common registry platform that could be distributed at no cost to FCPF countries.)

- **If so, what key functions should it include?**
- **If not, how would the CF know that an activity or set of lands have not been double counted?**

I am not sufficiently informed about the details of registry idea.
My comments may be misleading, therefore.

A "forest/tree carbon cadaster" appears essential to me when an efficient monitoring of changes shall be implemented – and independently verified. Such a carbon cadaster would need to be spatially explicit and well defined and a decision must be made what the smallest mapping unit is on which the forest/tree carbon pools are documented.

All relevant forest and tree pools should be documented, and can then be independently verified per mapping unit.

A cadaster could be stepwise enhanced and then also include trees outside the forest e.g. on agroforestry systems.

Some overall conclusions:

- Forest monitoring (for whatever assessment of status or changes) is complex.
  Many questions are scientifically not conclusively resolved yet.
- However, monitoring science must not slow down the process by ever pointing to more and more problematic detail issues – and by that contributing causing the infamous "paralysis by analysis".
- Neither however, should be pretended by the policy processes and its advisors that everything is feasible and tractable and easy and resolved.

Suggested elements of for a general assessment of "accuracy":

- From the preceding pages it has become clear that this author does not see a simple and easy approach (such as the definition of one single uncertainty figure like "10%", or so) to transparent and credible quantification of an accuracy measure for the determination/estimation of changes in forest carbon stocks.
- The observations and statements in the preceding pages are to illustrate that forests (whether we look at it primarily as a resource or primarily as an ecosystem) are complex systems; and that any approach to describe this system requires a certain degree of simplification and a number of assumptions, regardless how advanced the technology is that is employed.

  Both simplifications and assumptions need to be spelled out if transparency is at stake.
- When analyzing and rating the uncertainty of forest carbon figures, a number of sources of variability need to be considered; some of them are taken into account in the suggestion below.

  Here, we use the term "accuracy" in a general and colloquial summarizing all sources of uncertainty / variability; the term is not used in the narrow and strict statistical meaning!

  It is important to understand the different sources of variability (in terms of sampling statistics: sources of error), because all of them play a role in carbon assessments:

  > There is, to the best knowledge of this author, no study yet that does a full blown "variance component analysis" in the context of REDD+ carbon monitoring from which it could be concluded at which methodological point one invested US-Dollar (or Euro) would lead to a maximum gain in accuracy.
  >
  > However, "error budget" approaches have been presented by Prof. George Gertner and his team [e.g. Wang, Gertner et al. 2009: A Methodology for Spatial Uncertainty Analysis of Remote Sensing and GIS Products. Photogrammetric Engineering and Remote Sensing Vol.71, No12, Dec. 2005, pp1423-1432;
  > Gertner G et al. 2011. Uncertainties of mapping aboveground forest carbon duet to plot locations using national forest inventory plot and remotely sensed data. Scandinavian Journal of Forest Research, 26(4):360-373]

  - Measurement errors (observation errors) are omnipresent in empirical studies.

    Measurement errors cannot be avoided. But measures like (1) good training of the operators, (2) intelligent supervision of data collection and processing, and (3) regular checking help reducing these errors; these measures need to be reported to support credibility.

    We refer here to random measurement errors; systematic errors (for example through non-calibrated measurement devices) can be avoided by careful work.

    It is important to realize that measurement errors are usually not quantified nor reported by default in forest monitoring.

    One marked exception: when, in remote sensing image analysis, for example, a land cover class is erroneously classified, this would also be a "measurement error" (observation error) and is subject of the so called "overall accuracy"
  - Whenever a target variable cannot be measured directly but needs to be calculated from other measurements, a model is required. One example is above ground biomass, which is very often determined from so-called allometric models. These models carry a number of assumptions and are based on a limited set of data. They do not give the true value of the target variable (e.g. biomass), but a model based prediction. And this prediction carries an error which we call here model error.

Using the example of allometric single-tree biomass models: the model error depends on the number of sample trees that were used to build the model, the care that was taken when doing the corresponding measurements, the selection of the sample trees and their coverage of the range to which the model is being applied to, and the suitability of the underlying mathematical model.

The inherent model error - which in case of a regression model is given as the standard error of estimation - is a characteristic of the model and needs to be specified together with the model coefficients. This error specifies the variability of the model results whenever applied to a population where this model holds.

When a model is applied to a study region / population of interest where it is not valid, an additional systematic error of unknown size may be introduced.

It, therefore, **appears essential to assess the quality and validity of any relevant model** that is being applied in forest carbon assessments in a specific reference area, and document the methodology of this validation; and this holds explicitly for field based sampling studies **and** to remote sensing studies where field "observed" biomass or carbon is usually taken as reference. In both types of studies it does frequently not become entirely clear how the carbon values were determined.

- When a sampling study is carried out, only a sub-set of observations out of a much larger population is being selected and observed. Any sample-based inference to the entire population carries a "sampling-induced" error, which is called "standard error of estimation".

  How close the estimated value from a sampling study is to the true value remains unknown. The standard error describes how close the estimate is on the average: it describes how close to each other the estimates *would be*, if the same sampling study would be repeated very often.

  When carrying out sampling along the principles of design-based sampling, the standard error can unbiasedly be estimated for most sampling designs.

  The standard error depends on the sample size, the variability of te sampling units in the population and the sampling technique applied.

  All these elements are part of a transparent reporting of methods.

- In the following, elements of a "quality control" system for the assessment of forest carbon stock changes are given as a suggestion.

  These elements would serve as "indicators" for the reliability of estimates of carbon stock changes and would need to be linked / combined to one assessment system, for example by introducing weights or a rating system.

  Again: it is not seen that one single "accuracy indicator" can be given – such as a 10% threshold for "overall accuracy". It is rather suggested to break the carbon estimation process down into "easier manageable" components (see Graph 1 below) and rate the reliability of these components.

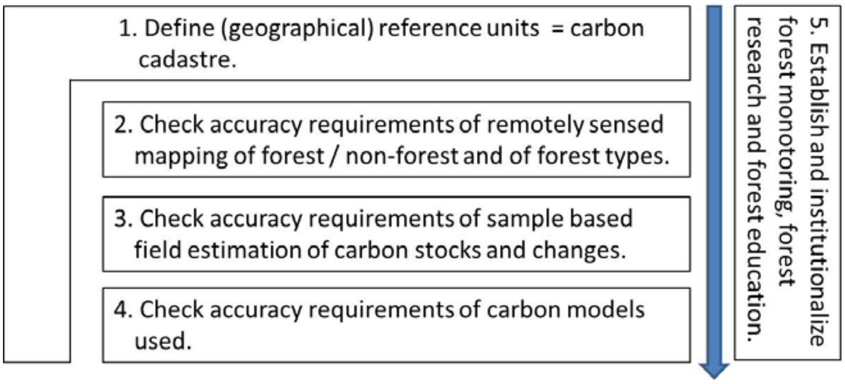
- These components refer to the assessment and monitoring of the true carbon development in the commitment period and not to the projected reference levels. Determining reference levels is a completely different exercise.

- It is important to make clear that these suggestions do NOT mean that there is scientific evidence that the figures mentioned are optimal.

  There is no such thing like scientific evidence in this context (to the best knowledge of this author).

However, the figures given here are likely to be acceptable for a broad majority of forest monitoring people, as they correspond to tradition, convention and experience (all such SUGGESTED figures are printed **in bold red** here to allow for easy identification of point s of discussion):

Graph 1 illustrates a possible system for an assessment of accuracy in the sense as defined before. The arrangement of the graph shall indicate that the carbon cadaster (1.) may be improved over time; number (5.) refers to the need to build efficient local (i.e. national or regional) forest management related institutions. Points (2.) to (4.) are components of a system of indicators for the accuracy of estimates of carbon stocks and changes.



Graph 1. Elements of a suggested system of quality check/accuracy assessment for estimates of carbon stocks and carbon stock changes. Building a reliable carbon cadaster (1.) and institutionalize an efficient forest management related infrastructure (5.) are "permanent tasks" while 2. - 4. are specific tasks for each monitoring project.

In detail:

1. Defining reference units = **carbon cadaster** = geographically clearly delimited areas for which separate carbon estimation, reporting and verification is being done.

   Such reference units could be smartly defined such that carbon accounting within them is optimized / streamlined in terms of the methodological approach and of cost; and that may include the criteria of statistical stratification = keeping variability as low as possible and delimiting also units entirely without forest (where no forest carbon accounting is being done).

   The size of such units could be in the order of **10´000-25´000km²** when talking about national forest carbon accounting.

   For the classes registered in this carbon cadaster, clear definitions must be spelled out; and these definitions need to be applied in a transparent manner also in the following points 2. to 4. to forest carbon estimation.

2. Defining an **accuracy measure for remote sensing** based forest / non-forest classification. That accuracy measure shall refer to the reference unit. The forest/non-forest classification shall have an overall accuracy of at least **90%**, where the sample taken to do the accuracy assessment shall be based on a probability sample (that is: not arbitrary, but, for example, randomized or systematic with or without stratification) of not less than ***n*=50** reference locations that are being independently verified in data sources of considerably higher reliability than the original imagery.

The approach to accuracy assessment needs to be documented step by step, including description and justification of data source and sampling design.

> Defining a sample size like *n*=50 is the easiest way of defining a requirement. However, with the same sample size and different sampling techniques very different precisions of estimation are being achieved. It would, therefore, be more reasonable to define a minimum standard error for the estimated overall accuracy, and it would then be up to the planners and implementers to devise an optimal sampling scheme to obtain such accuracy with lowest input of resources. A suggested simple standard error would be **SE%=5%** (meaning: absolute 5%, and not relative to the 90%, which would be 4.5%).
>
> Following statistical sampling, this would mean: a 95% confidence interval of about +/- 10% (absolute).

When accuracy shall be determined for more than the two classes forest/non-forest, the accuracy requirements should be less strict for the individual classes, for example **80%** in the case of 5 forest types; maintaining, however, the accuracy requirement for the distinction forest/non-forest.

3. **Field inventory** is required to produce estimates of carbon stocks and carbon stock changes and a series of further variables. For carbon stock estimation, a **simple standard error of 5%** (resulting in a 95% confidence interval of about 10%) appears appropriate and agreeable.

   The field inventory protocol would need to be reported including all elements of sampling design, plot design and estimation design.

   > Two further points should be addressed here:
   >
   > (1) The use of the 95%-confidence interval (CI) follows general convention and tradition. There is no scientific evidence whatsoever that the 95% CI is better/more appropriate/ etc. for any application than other probability levels!
   >
   > It is just that the 5% error margin (resulting from α=1.00-0.95) is being treated as a standard, and appears to be generally accepted for decades in the statistics science community.
   >
   > Deviation from the probability value of 95% for the confidence interval (e.g. suggesting 90% or 85%) would mean that a subject matter justification would need to be given why this new probability value is "better"; which this author considers impossible on scientific grounds.
   >
   > (2) The usually most efficient design-based sampling technique is systematic sampling.
   > However, there is no design-unbiased estimator for systematic sampling. When, as usual, applying the simple-random-sampling (SRS) estimator framework, a considerable over-estimation of the true standard error results in practically all cases; this is called: a conservative estimator for the standard error. As this over-estimation can be large, it may be recommended here to reduce the standard error requirements by, say, 30-50% when systematic sampling is used and the SRS estimator is applied for the estimation of the standard error. The above 10% CI requirement would then read about 15% or 20% CI for the application of systematic sampling.

4. **Biomass models / carbon models** are being applied in all forest carbon monitoring studies, because carbon is a variable that cannot be measured directly. Biomass models for forest carbon estimation are usually single tree models based on input variables such as tree species / species group, dbh and height; or they are gross models that give biomass per hectare for specific forest types.

In any case, the validity of the models used must be checked for any large area application, in order to reduce the possibility of systematic errors.

Also, the model error (standard error of estimation) must be known and shall be such that the projects requirements are met. Here, absolute values may be discussed and defined.

Gross expansion factors or generalized global biomass models will introduce potentially (!) large and unknown errors. In case they area applied, it is also advised to make a validity check.

However, checking the quality and validity of a carbon model for specific application in a study region is difficult because, as forest carbon cannot be measured directly, usually one model based estimate is compared to another model based estimate. The true value remains unknown practically always.

Giving evidence that a biomass model (or carbon model) is appropriate for a specific purpose is a generic research exercise: own observations need to be compared to what the model would predict and if the model predictions are well in line with the own observations one would conclude that the model applies to the given conditions. This verification can be done - for example for single tree allometric biomass models – by measuring a number of sample trees for volume, multiplying by known wood densities and then checking whether **95% of the sample tree biomass values are within the 95% confidence interval of the biomass model to be tested**. There should be expertise in the country to carry out such studies, which implies expertise in forest mensuration and statistical modeling.

5. This author is convinced that a sustainable (that is: successful on the long run) implementation of forest carbon payment schemes and REDD+ in general must go parallel to the institutionalization of a sustainable policy of forest management and the institutionalization of a forest monitoring system.

   By these “long-term measures of permanence” do governments give evidence that the goal of sustainable forest development is seriously pursued and seriously put onto the policy agenda for the benefit of the country, including carbon benefits.

   Only with a well trained and experienced forest monitoring institution, together with a forest research station will it be possible to efficiently further develop forest monitoring systems which are adapted to the specific national or regional circumstances, to give evidence of the success of these measures and to make such development independent of external interventions.

   One medium-term measure – deemed promising by this author, and urgently required – is to establish university curricula that are specifically geared towards the role of forests in the international conventions. There appears currently to be a lack of academic curricula that cover both the natural science elements of forest management and the policy/ governance/ institutions-related social sciences.

   By starting today educating and training the decision makers of tomorrow we do have the chance to make a difference in the future. This is a longer term undertaking which goes beyond normal project horizons.

   FCPF with its 37 partner countries could give an example and install, for example, a master course as a regional course in three partner countries. Students would mandatorily spend 2 semesters in one region and two in another region; locations in FCPF partner countries could be e.g. IPB in Bogor, Indonesia; for Asia, Sokoine University in Morogoro, Tanzania, for Africa; and CATIE in Turrialba, Costa Rica, for Latin America.

   Students of these courses could be supported by scholarships of their governments (or involved international donors) and could be actively involved in their national REDD programs.

   The number of graduated students from this course could be taken as one indicator for progress of a country towards the goals of REDD+.

The lengthy statements in this submission about the complexity of forest carbon monitoring and the many questions to be worked on, indicate that an efficient system of forest research is required to support the optimization of these monitoring systems; which is but one element of further developing the sustainable management of the forest resource of a country.

Christoph Kleinn

Göttingen, 30.12.2012