Collection: COST Action FP0903 (2010) - Rome (Italy)

"Research, monitoring and modelling in the study of climate change and air pollution impacts on forest ecosystems"

Guest Editors: E Paoletti, J-P Tuovinen, N Clarke, G Matteucci, R Matyssek, G Wieser, R Fischer, P Cudlin, N Potocic

Availability, accessibility, quality and comparability of monitoring data for European forests for use in air pollution and climate change science

Clarke N⁽¹⁾, Fischer R⁽²⁾, de Vries W⁽³⁾, Lundin L⁽⁴⁾, Papale D⁽⁵⁾, Vesala T⁽⁶⁾, Merilä P⁽⁷⁾, Matteucci G⁽⁸⁾, Mirtl M⁽⁹⁾, Simpson D⁽¹⁰⁻¹¹⁾, Paoletti E⁽¹²⁾

Data from existing monitoring programmes such as ICP Forests, ICP Integrated Monitoring and EMEP, as well as from large-scale international projects such as CarboEurope IP and NitroEurope, can be used to answer questions about the impacts of air pollution and climate change on forest ecosystems and the feedbacks of forest to climate. However, for full use to be made of the available data, a number of questions need to be answered related to the availability, accessibility, quality and comparability of the data. For example, how can these databases be accessed, *e.g.*, freely, over the internet, on request, by authorisation? How should intellectual property rights be protected, while improving access to data? What possibilities exist for harmonisation? Which quality assurance/quality control (QA/QC) procedures have been used and for how long? These and other relevant questions are discussed.

Keywords: Forest monitoring, Data availability, Data accessibility, Data quality, Data comparability, Data sharing

Introduction

Forests are expected to face significant pressures in the future from both climate change and air pollution (Fischer 2008). At present, research into and monitoring of climate change and air pollution impacts on European forests are rather fragmented, with a number of different networks existing (Fischer et al. 2011). COST Action FP0903 entitled "Climate Change and Forest Mitigation and Adaptation in a Polluted Environment" (COST 2010) creates a platform of experts from different fields and different networks, with the objectives of increasing understanding of the state and potential of forest mitigation and adaptation to climate change in a polluted environment and of reconciling process-oriented research, longterm monitoring and applied modelling at comprehensive forest research sites (supersites).

The large amounts of data already obtained within existing monitoring programmes and large-scale international projects can be used to help to fulfil the objectives of COST Action FP0903. The Action's Working Group 1 aims to investigate the availability and evaluation of data, with special emphasis on databases from long-term monitoring programmes and projects as it is these that have the greatest potential to provide the necessary information. However, in order to make best use of the large amounts of data that exist, we need to know the answers to a number of questions, including:

- what data are available?
- how accessible are they?
- what is their quality?
- how comparable are data obtained from different sources and by different methods?

In addition, COST Action ES0804 "Advancing the integrated monitoring of trace gas exchange between biosphere and atmosphere" creates a platform for analysis, harmonisation, synthesis, and assessment of future needs and further development of a European integrated monitoring program for comprehensive trace gas flux observations. The same issues are relevant for this Action.

The aim of this paper is to inform the scientific community about the availability of data relevant to the objectives of COST Actions FP0903 and ES0804 and to briefly discuss issues of their accessibility, quality and comparability. In an accompanying paper, the same transnational forest monitoring

(1) Norwegian Forest and Landscape Institute, P.O. Box 115, N-1431 Ås (Norway); (2) Johann Heinrich von Thünen-Institute, Instit. for World Forestry, Leuschnerstrasse 91, 21029 Hamburg (Germany); (3) Alterra, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen (Netherlands); (4) Swedish University of Agricultural Sciences, P.O. Box 7050, SE-75007 Uppsala (Sweden); (5) DISAFRI Università della Tuscia, v. S. Camillo de Lellis, I-01100 Viterbo (Italy); (6) University of Helsinki, P.O. Box 48, SF-00014 Helsinki (Finland); (7) Parkano Research Unit, Finnish Forest Research Institute, Kaironiementie 15, FI-39700 Parkano (Finland); (8) CNR/ISAFOM, v. Cavour 4-6, 87036 Rende (CS, Italy); (9) Federal Environment Agency Austria, Spittelauer Lände 5, A-1090 Vienna (Austria); (10) EMEP MSC-W, Norwegian Meteorological Institute, Blindern, N-0313 Oslo (Norway); (11) Chalmers University of Technology, Hörsalsvg. 11, S-412 96 Gothenburg (Sweden); (12) CNR/IPP, v. Madonna del Piano 10, I-50019 Sesto Fiorentino (FI - Italy)

a Nicholas Clarke

(nicholas.clarke@skogoglandskap.no)

Received: Dec 23, 2010 - Accepted: Mar 17, 2011

Citation: Clarke N, Fischer R, de Vries W, Lundin L, Papale D, Vesala T, Merilä P, Matteucci G, Mirtl M, Simpson D, Paoletti E, 2011. Availability, accessibility, quality and comparability of monitoring data for European forests for use in air pollution and climate change science. iForest 4: 162-166 [online 2011-08-11] URL: http://www.sisef.it/iforest/show.php?

id=582

and research networks in Europe are discussed in view of their potential to establish a transnational system of supersites for forest monitoring and research (Fischer et al. 2011).

Availability of data

A number of large-scale international databases exist both from monitoring programmes and research projects, as well as data from many small-scale projects and programmes, including:

- international databases from monitoring programmes (ICP Forests, ICP Integrated Monitoring, EMEP, Long-Term Ecological Research LTER, ICOS etc.);
- data from large-scale (CarboEurope IP, NitroEurope, IMECC etc.) and small-scale international research projects covering most of Europe;
- regional databases and projects (*e.g.*, Noltfox, the Northern European Database of

Tab. 1 - Overview of existing large-scale databases relevant for forest monitoring and research, including data access rules and manuals/pro-tocols.

Programme/project	Short name (Home page)	Database	Data access rules	Manual/protocols
International Coope- rative Programme on Assessment and Moni- toring of Air Pollution Effects on Forests	ICP Forests http://www.icp-forests.org	Central data base with raw data at vTI, Germany	Open after registration and information of countries, 2 years reserved for internal evaluation	http://www.icp- forests.org/Manual.htm
International Coope- rative Programme on Integrated Monitoring of Air Pollution Effects on Ecosystems	ICP Integrated Monitoring http://www.environment.fi/ default.asp? contentid=17110& lan=en	Central data base with partly raw data, partly aggregated data at SYKE, Finland	Open after registration, large data sets are approved by Task Force	http://www.environment.fi/de- fault.asp?node=6329&lan=en
The nitrogen cycle and its influence on the European greenhouse gas balance	NitroEurope http://www.nitroeurope.eu	Central data base with raw data at CEH, United Kingdom	Meta data: open access. raw data: open for partners and case by case decision for external users	Partly CarboEurope methods - Rainfall: EMEP methods - Own methods of partners, that need to be documented in me- thod sheets
Cooperative Programme for Monitoring and Evaluation of the Long- range Transmission of Air Pollutants in Europe	EMEP http://www.emep.int	Central data base with raw data at NILU, Norway: http://ebas.nilu.no/	All the data is open for web download	http://tarantula.nilu.no/projects/ ccc/manual/index.html
Integrated Project CarboEurope-IP Assessment of the European Terrestrial Carbon Balance. Data under the responsibility of IMECC from 2009	CarboEurope (IMECC) http://www.carboeurope.org (http://imecc.ipsl.jussieu.fr/) Ecosystem data: http://www.europe- fluxdata.org	 Central data base with meta data at MPI-BCG, Germany. Half-hourly Ecosystem data, Univ. of Tuscia Viterbo, Italy. Raw data under respon- sibility of site managers 	Metadata: open access. half hourly: open for partners and case by case decision for external users (but from 2011 open)	Measurement protocols under http://www.carboeurope.org
Integrated Carbon Observation System	ICOS http://www.icos-infrastruc- ture.eu	Raw data: - Ecosystem sites Univ. Tuscia, Italy - Atmospheric sites, LSCE, France	Metadata: open access Raw data: open access	 Partly CarboEurope and IMECC (7 FP) Partly under development Concept papers available
European Long-Term Ecosystem Research Network	LTER http://www.lter-europe.net	 Central data base with meta data at UBA, Austria. Raw data under respon- sibility of site managers 	Metadata: web-based, open. Raw data: links to other networks	Best practice guidelines at: http://www.lter-europe.net/doc- ument-archive/central/Best %20practice%20guideline %20v3-June1.pdf

Long-Term Forest Experiments, the Nordic flux tower network NECC);

 national databases and projects including national forest inventories (NFIs); the European National Forest Inventory Network (ENFIN) in collaboration with the FutMon project works to maximise the synergy between NFIs and other European and international level data collection systems, monitoring and reporting activities.

A number of sites have been used in several different networks as well as smaller-scale projects (*e.g.*, site Birkenes in southern Norway). For some very old sites (*e.g.*, site Zelivka in the Czech Republic - Z. Lachmanová, pers. comm.), however, older data may only exist on paper, limiting their accessibility. A major problem until now is that, in practice, data that are not included in databases are often only known by the scientist(s) who obtained them. As soon as these scientists leave or change their position the data are in practice lost. This stresses the importance of databases that can host data and related Metadata acquired by scientists/ scientific teams/scientific project teams not involved in large research programs.

For practical reasons, it makes sense to consider in this analysis mainly the data from the large-scale programmes and projects, covering all or most of Europe and with manuals or protocols that are to a large extent harmonised. For these datasets, metadata and processing procedures exist and are available together with the data, enabling the user to understand and evaluate the nature of the measurements and ensure their correct use. An overview of the most relevant largescale databases for European forests that can be used in air pollution and climate change research is given in Tab. 1.

Accessibility of data

In principle, data obtained using public funding should be available to the public. The OECD has recently published principles and guidelines for access to research data obtained using public funding (OECD 2007). Improved access is generally seen as benefiting the advancement of research, boosting its quality and facilitating cross-disciplinary research cooperation. However, that some data may remain inaccessible is acknowledged by the OECD guidelines, where it is stated that "Data access arrangements should respect the legal rights and legitimate interests of all stakeholders in the public research enterprise" (OECD 2007).

Rules for access to data in existing databases vary, and can be summarised as:

- free access (generally by internet), *e.g.*, EMEP;
- free access with registration (often for in-

ternet security issues or for additional communications with users about problems in the data), but no authorisation, *e.g.*, ICOS;

- access on request/by authorisation, *e.g.*, ICP Forests;
- inaccessible (information lost either technically or politically due to ownership/ property/security concerns).

Information about the accessibility of data in the main European forest monitoring and research databases is included in Tab. 1. In practice, though, scientific use of data beyond the specific networks gathering the information is in many cases limited (Fischer et al. 2011).

The data that are accessible in databases are seldom truly raw data, in the sense of being the signals that come from the sensors. Raw data needs to be defined: for example, in CarboEurope IP and IMECC raw data are 10Hz unprocessed measurements coming from the sensors that need to be elaborated and corrected to obtain the final, useful fluxes. These are at the moment not available in the database; however, the half-hourly data are available. The half-hourly data may be considered raw data or not, depending on standpoint: for modellers or general users they are raw data, for eddy covariance workers not. The "raw data" referred to in Tab. 1 are therefore not generally truly raw data, although there are exceptions: For ICOS truly raw data, 10 Hz measurements, will be available

Published results are of course generally available in international peer-reviewed journals, books, technical reports from programmes or projects, or national reports (often in the national language), but these are normally not the raw data, with some exceptions. More commonly they are the results of statistical treatment of the data.

Although the principle of free access to publicly funded data is important, there are legitimate concerns about intellectual property rights (IPR), including a right to first use of unpublished data for the scientist(s) responsible for obtaining them. "In current research practice, the initial data-producing researcher or institution is sometimes rewarded with temporary exclusive use of the data. The rules for such incentive arrangements should be developed and explicitly stated by the funding sources in co-operation with the affected research communities" (OECD 2007). IPR regulations should be mandatory in any project supported by public funding. A programme or project may have an agreed intellectual property policy with strict regulations governing the use of data collected by others, e.g., ICP Forests (Lorenz 2010). However, respect for IPR should be balanced with the fact that at one point the data often acquired using public money (EU or national) should become

available to the broad scientific community and for this reason another possibility is for first use to be guaranteed for a period of, *e.g.*, one to three years after the end of the project before open access. Financial security provided by the funding organisation would help to improve access, as would also a direct link between funding and data sharing or data use by the scientific community.

Not only is it important for society as a whole that data are made accessible, but it is also important for those collecting the data. Experience has shown that in the long term only the "active" sites in terms of data sharing have survived, *i.e.*, those sites with a wide use of their data by scientists that are not directly involved in the data collection. This provides justification for further funding. In addition, experiences from synthesis activities like FLUXNET (http://www.fluxdata.org) demonstrated that multi-sites analysis does not prevent the publication of site specific papers but instead adds new analysis and important scientific results otherwise probably not possible (e.g., Beer et al. 2010, Jung et al. 2010).

Data quality

The quality of a database is only as good as the quality of the data it contains. Strict quality assurance/quality control (QA/QC) procedures at all stages of an investigation (field sampling, transport to the laboratory, laboratory analysis, data treatment and reporting) are crucial and should be included in the manual of any programme or project supported by public funding.

Not all analytical methods are equally good, and it is possible to prepare a list of methods that give inaccurate or imprecise results, as has been done for example for deposition analyses in ICP Forests (Clarke et al. 2010). Even laboratories using the same analytical method can get very different results. This can be seen in the results from inter-laboratory ring tests, where the same samples are analysed by a number of different laboratories; the range in results obtained may be large even when the same method is used (e.g., Marchetto et al. 2009). In recent years, a better understanding of QA/QC has led to improvements in the quality of data obtained from laboratories. However, questions may in some cases remain about the quality of older data and these data should be flagged. Whenever possible it is important to associate an uncertainty value with all the data, uncertainty that can take into account all the different sources, from the measurement collection to the final OA/OC.

Data quality is generally considered relative to the standard and objectives of each single project. Thus different projects may have data of similar, high quality for the purpose of the project but not necessarily usable for the purposes of other projects or for meta-analyses. This relates to the question of data comparability, which is discussed in the next chapter.

Data comparability

Data comparability is related to data quality and should be seen in this context. In many databases, data have been obtained from a large number of different partners, raising questions about their comparability. Different methods/techniques are often used even within one programme, leading to difficulties in comparing results. Achievement of comparability between different programmes /projects is even more difficult as their aims differ.

An example of the use of different methods within one programme is the monitoring of deposition, especially throughfall, in ICP Forests. Different types of samplers are used; for throughfall sampling these could be either funnels or gutters. In addition to different sampler types, there are differences in a number of factors that might influence the representativity of the measurements, such as the number of samplers used, the surface area of the samplers, and their placing in the forest. Field intercomparisons of samplers for bulk precipitation and throughfall showed that there were large differences in the results obtained, and that these differences were most likely related to differences in the sampling strategy such as the sampler placement and the collecting area used (Bleeker et al. 2003, Erisman et al. 2003). Such difficulties create major problems in attempts to use the data for model evaluation (e.g., Simpson et al. 2006). Inaccuracy induced by the samplers appeared to be larger than that induced by use of different laboratories in chemical analysis.

In the laboratory too, use of different methods can lead to different results, even in cases where the methods use the same basic principle. For example, a comparison of five different methods for aluminium fractionation carried out by Wickstrøm et al. (2000), four of which used the same fractionation principle (cation exchange), showed differences caused by factors such as reaction time and pH even when the same fractionation principle was used. Method intercalibrations are therefore necessary and should be mandatory.

Method harmonisation, leading to the use of the same equipment and the same methods by all participants in a network, might be considered as a way of improving data comparability. However, there are a number of hinders to harmonisation:

- methods could be well-adapted for local conditions, leading to uncertainty about the benefits of a new, harmonised method;
- there are different historical schools in different countries (*e.g.*, soil type classifica-

tion);

- changing the methodology will break the time series; comparison between before and after the method change is difficult, as old and new methods must be run in parallel (for at least a year in the case of field equipment) to be able to compare data obtained;
- harmonisation is expensive: new equipment must be purchased and installed;
- harmonisation might inhibit the ability to develop new methods or improve on existing ones.

One possible solution could be the use of a standard method/set-up in parallel with the local, long term and established methodo-logy. In this way the common reference method would permit a clear and robust comparison and the paired methods applied at each site to evaluate the uncertainty or spot problems and errors. However this solution would be clearly more expensive.

In practice, how do we deal with these challenges, considering that it is important to have comparable data without losing too much?

- Use of the same acquisition protocols, where possible;
- in some cases it is possible to keep different methods but link them to the same reference (*e.g.*, transfer of different cover-abundance scales for ground vegetation to a standardized percentage scale);
- centralized raw data processing, where possible;
- bridging functions (as developed within the ENFIN community);
- inter-laboratory ring tests;
- field harmonisation tests;
- intercalibrations;
- detailed meta-information submitted with the data;
- early and user-friendly release mechanisms of the data to allow external usage and evaluation. Data buried in databases almost always contains significant and often easily-fixable errors;
- definition and organization of coordination centres that would help in the development of protocols and assist those responsible for the measurements;
- in many cases external co-financing has helped a lot to enable harmonisation processes, because transnational harmonisation in many cases does not provide an immediate national benefit.

Laboratory improvement after participation in ring tests has been demonstrated in the ICP Forests programme (Marchetto et al. 2009). Field harmonisation has proven more difficult to achieve. In the LIFE+ FutMon project, a harmonisation experiment is currently being carried out, in which national deposition samplers are compared with standardised deposition samplers on forest monitoring plots throughout Europe. A preliminary paper on this experiment is published in this journal (Zlindra et al. 2011). Centralized data processing has been tested in FLUXNET with the "LaThuile synthesis activity" (http://www.fluxdata.org) leading to a large number of synthesis papers based on harmonized and standardized datasets collected in more than 200 sites globally (*e.g.*, Beer et al. 2010, Jung et al. 2010).

The question of scale is also important when comparing data from different sources. The scale at which to operate is chosen according to the questions to be answered. When the scale at which to operate is chosen, it will determine the study design and also the amount of precision that is possible. For example, detailed biological studies might require exact and extensive data on a single tree, but such data may have limited usefulness to assess the functioning of the whole ecosystem. At the other end of the scale, good measurements might be made of an extensive area (for example with masts sampling large footprint areas), but with little information on individual species or trees. The main problem for the users is then to integrate the different measurements considering the related uncertainties.

Conclusions

Main conclusions are:

- a lot of data exists, but it is hard to get a good overview of what is available and where;
- much is accessible, but there is room for improvement. Internet-based solutions should be the standard; these have been very successful with some networks;
- a major challenge is how to improve access while safeguarding both IPR and also publicly-funded data collection;
- networks set up for one purpose often provide data which are valuable for other research communities (*e.g.*, CarboEurope's data on friction velocity have proven valuable for air pollution modelling), and such co-benefits need to be encouraged and enabled, *e.g.*, by making access easy;
- understanding of QA/QC and assessing the uncertainty is crucial. This has improved in recent years;
- comparability of data from different sources, often obtained by different field and laboratory methods, is a serious challenge and needs to be frequently tested;
- cross-comparison of data from different databases offers valuable opportunities for a better exploitation of present data including validation of models and understanding of real-world forest responses to air pollution and climate change.

Acknowledgments

Joint publication of the COST Actions FP0903 "Climate Change and Forest Mitigation and Adaptation in a Polluted Environment" and ES0804 "Advancing the integrated monitoring of trace gas exchange between biosphere and atmosphere". We thank also two anonymous reviewers for helpful comments on the manuscript.

References

- Beer C, Reichstein M, Tomelleri E, Ciais P, Jung M, Carvalhais N, Rödeneck C, Arain MA, Baldocchi D, Bonan GB, Bondeau A, Cescatti A, Lasslop G, Lindroth A, Lomas M, Luyssaert S, Margolis H, Oleson KW, Roupsard O, Veenendaal E, Viovy N, Williams C, Woodward FI, Papale D (2010). Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329: 834-838. - doi: 10.1126/ science.1184984
- Bleeker A, Draaijers G, van der Veen D, Erisman JW, Möls H, Fonteijn P, Geusebroek M (2003).
 Field intercomparison of throughfall measurements performed within the framework of the Pan European intensive monitoring program of EU/ICP Forest. Environmental Pollution 125: 123-138. doi: 10.1016/S0269-7491(03)00142-8
- Clarke N, Zlindra D, Ulrich E, Mosello R, Derome J, Derome K, König N, Lövblad G, Draaijers GPJ, Hansen K, Thimonier A, Waldner P (2010). Sampling and analysis of deposition. Part XIV. In: "ICP Forests. Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests". UNECE, ICP Forests, Hamburg, Germany, pp. 66. [ISBN: 978-3-926301-03-1] [online] URL: http://www.icpforests.org/pdf/FINAL_Depo.pdf
- COST (2010). Action FP0903 [online] URL: http://cost-fp0903.ipp.cnr.it/
- Erisman JW, Mols H, Fonteijn P, Geusebroek M, Draaijers G, Bleeker A, van der Veen D (2003). Field intercomparison of precipitation measurements performed within the framework of the Pan European Intensive Monitoring Program of EU/ICP forests. Environmental Pollution 125: 139-155. - doi: 10.1016/S0269-7491(03)00082-4
- Fischer R (2008). Forest ecosystems in a changing environment: identifying future monitoring and research needs. COST Strategic Workshop. Istanbul (Turkey) 11-13 March 2008. [online] URL: http://www.icp-forests.org/pdf/COST.pdf
- Fischer R, Aas W, de Vries W, Clarke N, Cudlin P, Leaver D, Lundin L, Matteucci G, Matyssek R, Mikkelsen T, Mirtl M, Öztürk Y, Papale D, Potocic N, Simpson D, Tuovinen J-P, Vesala T, Wieser G, Paoletti E (2011). Towards a transnational system of supersites for forest monitoring and research in Europe an overview on present state and future recommendations. iForest (in press) doi: 10.3832/ifor0584-004
- Jung M, Reichstein M, Ciais P, Seneviratne SI, Sheffield J, Goulden ML, Bonan G, Cescatti A, Chen J, de Jeu R, Dolman AJ, Eugster W, Gerten D, Gianelle D, Gobron N, Heinke J, Kimball J, Law BE, Montagnani L, Mu Q, Mueller B, Oleson K, Papale D, Richardson A, Roupsard O,

Running S, Tomelleri E, Viovy N, Weber U, Williams C, Wood E, Zaehle S, Zhang K (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467: 951-954. - doi: 10.1038/nature09396

Lorenz M (2010). Objectives, strategy and implementation of ICP Forests. Manual Part I. In: "ICP Forests. Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests". UNECE, ICP Forests, Hamburg, Germany, pp. 21. [ISBN: 978-3-926301-03-1] [online] URL: http://www.icp-forests.org/Manual.htm

Marchetto A, Mosello R, Tartari GA, Derome J,

Derome K, König N, Clarke N, Kowalska A (2009). Atmospheric deposition and soil solution working ring test 2009. Laboratory ring test for deposition and soil solution sample analyses for the laboratories participating in the EU/Life+FutMon Project. Report 04.09, CNR-ISE, Italy.

- OECD (2007). Principles and guidelines for access to research data from public funding. [on-line] URL: http://www.oecd.org/dataoecd/9/61/ 38500813.pdf
- Simpson D, Fagerli H, Hellsten S, Knulst J, Westling O (2006). Comparison of modelled and monitored deposition fluxes of sulphur and nitrogen to ICP-forest sites in Europe. Biogeosciences

3: 337-355. - doi: 10.5194/bg-3-337-2006

Wickstrøm T, Clarke N, Derome K, Derome J, Røgeberg E (2000). Comparison study of five analytical methods for the fractionation and subsequent determination of aluminium in natural water samples. Journal of Environmental Monitoring 2: 171-181. - doi: 10.1039/a909139d

Zlindra D, Eler K, Clarke N, Simončič P (2011). Towards harmonization of forest deposition collectors - case study of comparing collector designs. iForest (in press) - doi: 10.3832/ ifor0586-004