

Enviro-Net: A Network of Ground-based Sensors for Tropical Dry Forests in the Americas

Gilberto Z. Pastorello ^a, G. Arturo Sanchez-Azofeifa ^{a,*}, Mario A. Nascimento ^b

^aDepartment of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB, T6G 2E3, Canada
gilbertozp@acm.org, arturo.sanchez@ualberta.ca (*corresponding author)

^bDepartment of Computing Science, University of Alberta, Edmonton, AB, T6G 2E8, Canada
mario.nascimento@ualberta.ca

Abstract – Monitoring ecosystems is necessary to understand the development and the response of forest ecosystems to climate and human disturbance. With the increasing availability of sensor data, new techniques for data analysis are being introduced. This paper presents the Enviro-Net Project, which involves several deployments of sensor systems throughout the Americas. Such deployments use various ground-based monitoring solutions to assess the conditions in tropical dry forests over long periods of time. This work describes these deployments and our efforts in managing the systems themselves and the data generated by them, also discusses issues related to data cleaning, and the Web portal developed to aid in data management, visualization and analysis tasks.

Keywords: micro-climate monitoring, phenology monitoring, wireless sensor networks, Web-based sensor data management

1. INTRODUCTION

The Enviro-Net Project offers a Web-based platform for exploring data from a network of ground-based sensors deployed throughout sites in tropical dry forests in the Americas. Involving eight research institutions, but coordinated at the University of Alberta, Enviro-Net's original goal was to allow research on the phenology of tropical dry forests at selected control sites. Current research on phenology, the study of climate effects on periodic biological activity (Schwartz 2003), has been conducted mostly in an ad-hoc, site-by-site, manner. In order to achieve a more systematic approach to studying phenology across sites, sensor systems and computational infrastructure is essential. With that in mind, environmental monitoring stations have been deployed in several sites in Argentina, Brazil, Costa Rica, and Mexico. The variables monitored include air and soil temperature and relative humidity, incoming and reflected solar radiation, incoming and reflected Photosynthetically Active Radiation (PAR). Data from these stations are fed into Enviro-Net's database using online tools where they are quality-checked and made available to system users (see <http://www.enviro-net.org>). Once available, these data can be explored using a set of visualization and selection tools. Among those tools we single out: graphing of individual sensor measures through time, trend graphing of vegetation indexes (such as different versions of NDVI and EVI) through time and across sites, visualization of spatial distribution of measures and their level of confidence, detection of missing data, simple quality checks, and filtering based on time frame, time of the day, values of interest, error removal, raw/calibrated values, among many other possibilities.

The contributions from the Enviro-Net Project include the experience in maintaining deployments that are: (1) long term (multi-year); (2) distributed across different sites and ecosystems; (3) rely on a network of collaborators from each

site; (4) use low cost equipment; and, (5) offer a centralized and simple data processing and access solution for all project members. This results in an inexpensive platform for ground-based environmental monitoring, allowing not only phenology research, a field of growing importance in climate related studies, but also the combination of other sources and types of data from various research fields such as carbon and water vapor flux, validation and calibration of satellite-based data collection, etc. Furthermore, the online tools allow better provenance and quality control over the data, improving reproducibility of the research results, a critical requirement in current environmental and climate related research.

2. DEPLOYMENTS

Currently, we have 36 operational deployments, plus temporary deployments in Edmonton, Canada, for equipment testing and calibration. In more detail, there are three deployments in a site in Argentina (with five more being installed in two other sites in 2011); 29 deployments in three sites in Brazil, two in a site in Costa Rica; and two at one site in Mexico. Wired and wireless deployments are both used. Wireless deployments use wireless communication to either transmit the data to a logger or interact with its control software in a laptop computer. For wireless deployments, the number of nodes is usually 12 – some with as few as five and some with as many as 20 nodes – each node having between three to six sensors each. Planning for new deployments are underway, and a total of 60-70 deployments should be operational by 2013 (mostly on Boreal sites).

All of our deployments use commercially available equipment from a variety of manufacturers. These include wired systems from Onset Computer Corp. (<http://www.onsetcomp.com/>), wireless systems from Olsonet Communications Corp. (<http://www.olsonet.com/>) and MicroStrain, Inc. (<http://www.microstrain.com/>). Among the sensors we used solar radiation sensors provided by Onset and Apogee Instruments, Inc. (<http://www.apogeeinstruments.com/>); air temperature and relative humidity sensors from Onset; and, soil moisture sensors by Decagon Devices, Inc. (<http://www.decagon.com/>). These offer a relatively inexpensive way for monitoring phenology at an ecosystem level, i.e., a high spatial resolution, with an also high temporal resolution, for long periods of time (using solar power to recharge batteries for some of the equipment). Alternatives to achieve comparable resolutions include eddy covariance carbon flux monitoring and high resolution, on-demand, satellite remote sensing images, both of which orders of magnitude more expensive.

Phenological responses can be monitored with solar radiation flux sensors installed above canopy levels (in what we call a Wireless Optical Phenology Station (WOPSTM), measuring incoming and reflected PAR (wavelengths between approximately 400 to 700 nm) and broad spectrum solar

radiation (wavelengths between approximately 300 to 1100 nm). This configuration is shown in Figure 1, which also depicts the 85° field of view of the sensor and the minimum distance of five meters from the canopy to the sensors measuring reflected radiation. This gives a minimum radius of around 50 meters for measurements. Ratios of these measurements can be used to derive vegetation indexes such as NDVI (Normalized Difference Vegetation Index) or EVI (Enhanced Vegetation Index) – see, for instance, **Huemrich et al. 1999**, **Wilson and Meyers 2007**, **Jenkins et al. 2007**, and **Rocha and Shaver 2009**. Auxiliary sensors, such as temperature and humidity, support analysis of other factors affecting phenology. A second type of deployment involves monitoring conditions under a forest canopy, its understory. These deployments allow assessing a different range of micro-climatic conditions and also soil condition – e.g., temperature and moisture levels. Understory deployments usually offer easier access, being useful for validating the readings observed in a tower and also as a backup for certain variables in case of sensor malfunction in a tower. Using wireless systems, increases the spatial coverage of understory deployments with much smaller increases in cost and effort for deployment, system maintenance and data retrieval.

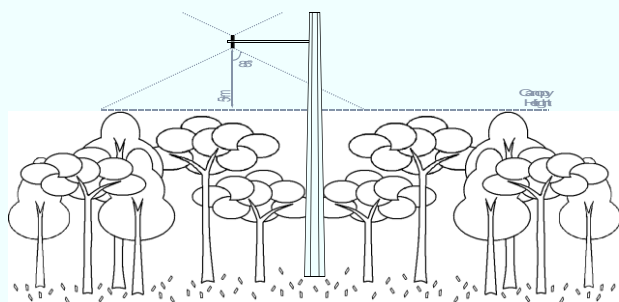


Figure 1. Schematics of phenology tower deployment

Except for test installations, our deployments are intended to be long term, collecting data for a minimum of 2-3 years. Our first deployments date back to mid 2007, and our first wireless deployments were installed in mid 2008. Some level of interruption on the data collection affected all of our deployments, with gaps in data ranging from a few days up to a couple of months, depending on how early the problem was detected. Many problems were related to unexpected interactions of environmental conditions with the equipment, such as weatherproofing failing on account of differences of pressure between the sealed interior of the case and external environment (usually caused by extreme variations on temperature), or lightning electrical discharges more powerful than supported by the grounding equipment. However, most of the deployments are still operational today, with secured funding for maintaining them operational until at least 2013.

Having long term ground-based deployments requires a good understanding of the ecological processes and drives that control a given ecosystem, and also familiarity with the sensor equipment and what to expect from it in terms of both data products and error handling, knowledge about the environmental conditions surrounding the deployment and its accessibility, and a clear attribution of maintenance tasks. Moreover, a number of other issues need to be addressed in scenarios as those considered in this paper. The preparation for

deployment expeditions involved coordination of efforts in home laboratory, where most of the equipment was tested, and the partner research teams at the deployment sites, where supporting equipment was installed. Sensor calibration was mostly conducted against a reliable source at the lab and, where no reference was available, comparing sensors against each other's readings also proved effective with a sufficient number of sensors. Testing of logging system and control software was extensive, not only to test the equipment itself, but also as a way to gather experience with it, especially in non-standard conditions (e.g., low battery, high humidity, etc).

The installation of the phenology tower is also a potential problem. If the disturbance to the site is too significant, the readings from the tower will be representative only after the ecosystem recovers the original vegetation coverage. Furthermore, the height of the tower has to be enough to cope with vegetation growth for at least the period of the study.

For wireless systems, testing the radio signal coverage in the deployment site is essential, since vegetation coverage and topography significantly affect this parameter. One aspect to keep in mind is that changes in vegetation coverage can greatly impact the radio signals, particularly in deciduous forests, where the deployments are usually installed in the dry season and most interference occurs during the wet season. Similar concerns were reported by **Barrenetxea et al. 2008** and **Welsh 2010**. These usually involved equipment and software developed (at least in part) by the research group. Also, these efforts have been mostly short term, and focused on specific aspects of monitoring, while our work is mainly long term and has an integrated approach from systems deployment to offering easy access to derived data products.

Data retrieval also involves a few choices. Although automation of this task is possible, e.g., by using cellular data communications or satellite-based transmissions, it might not be the best solution in many cases. However, for very remote sites that are seldom visited, this might be the only feasible solution. However, sites that are visited often (because of other experiments, for instance), might be better served by manual data retrieval. This is particularly true with new equipment that was not previously tested in the environment it is deployed on, yielding better chance of correcting problems.

Having this level of distribution of deployments it is essential to have multiple teams carrying out maintenance efforts. This presents a challenge from the point of view of training members of these partnering research groups, since the time available for training sessions is limited. The most effective method for us was to allow hands on experience with the equipment from the beginning, with early exposure to the most common problems faced in the field. Besides the equipment documentation and manuals, it was also very helpful to create documentation of our own, describing our deployment protocols and procedures for dealing with exceptions. Personal backgrounds and interests also affected the rate of success for our deployments; however, solution to this type of problem is more subjective and case-by-case.

The last aspect directly related to deployments is the long term aim for the deployments. Regular maintenance is key in ground-based monitoring, with some systems requiring interference as often as twice a month. The site itself might present problems such as fallen vegetation blocking sensors, animals or insects obstructing or damaging equipment, etc. Sensors can also deviate from their correct calibration, requiring adjustments;

they also have a limited life span, which leads to regular replacement needs. Clear distribution of these and other maintenance tasks helps prevent gaps in resolution of equipment problems and conflicts. Keeping regular cycles to perform maintenance tasks also helps, usually having one short term cycle for data retrieval and simple maintenance (e.g., every 2-8 weeks) and a longer cycle for sensor re-calibration and equipment replacement.

3. DATA MANAGEMENT

The potentially large volume of data generated by ground-based sensor systems demands some level of automation with respect to data management tasks, preferably within proper computational infrastructure (Teillet 2010 and Rundell et al. 2009). Real datasets generated by these systems, however, are subject to problems and exceptions, which setback automation efforts from the start. Many of such issues are usually dismissed as “implementation details”, but might affect not only data quality, but also models to store and distribute the data. More expensive and homogeneous equipment are affected in a limited way by most of these issues. In a distributed, heterogeneous, and low cost approach as ours, these are actually common issues.

Examples of problems with raw datasets range from time related issues to proper documentation of correction procedures. Regarding time, two issues may arise: time synchronization and time zone determination. Time counting in logging equipment drifts, especially if it is subject to drastic changes in temperature. Without regular corrections, the precision of the timestamps in the data might be compromised. Determining the correct time zone might also be a challenge. Although our deployment protocols establish that the local standard time is to be used at all times, the number of people involved in the deployments (increasing the probability of inadvertent changes in time options), misconfiguration (mainly regarding daylight savings settings) and failures in laptops with control software, may introduce uncertainty to determining which time zone is actually used in a deployment.

Changes in the formats of the raw datasets extracted from loggers is a small problem when dealing with a single model from one manufacturer. However, when this changes to several different models, versions of models, versions of software, from a number of manufacturers, this kind of changes become a challenge. Appropriately documenting the corrections applied to a dataset becomes significant to understand the data provenance, which can be a key factor in assessing the quality of the data.

The implementation of corrections to these and other problems with raw datasets are usually carried out within a data pre-processing (or data cleaning) phase. Although these terms usually encompass explicit data quality verification or removal of erroneous readings (e.g., values outside the scale measured by a sensor), in our framework, this phase is limited to addressing problems that actually prevent, or are difficult to trace after, the ingestion of the data into a data management system.

Data manipulation resources and ease of use are the two main guidelines for building a data management system for large scale in-situ monitoring. A pre-processing step, as discussed

above, allows datasets to be uploaded to the system, storing and making them available in a uniform fashion. Users can interact with the system by applying filtering, aggregation and other processing operations to the data. Data visualization and downloading are also available for original measurements and processed data.

Uploading the pre-processed data into the system has to offer some flexibility to account for changes in sensor configuration and other deployment options. This aspect is addressed by allowing mapping of individual parts of datasets into pre-defined storage configuration within Enviro-Net's portal. When uploading new datasets, properly handling errors and exceptions is essential for both user experience and data quality. Documentation of the upload process is also a good practice regarding data provenance. This was addressed through automatically collecting metadata and offering a historical view of the process.

With the data ingested into the system, a user can modify the data according to their goals. Operations available include filters (three of them shown in the upper part of Figure 2) and aggregation operators (e.g., using daily averages instead of an entire dataset). We also integrated more specialized operations, such as automatically deriving vegetation indexes from data collected by solar radiation sensors. These and other options (including error checking and data quality information) are available to the users through a simple Web-based interface. Furthermore, the original measurements and derived data can be either downloaded to be further analyzed or visualized within the portal.



Figure 2. A little over one year of NDVI data for two in-situ monitoring sites in South-East Brazil

Good data visualization tools significantly help in assessing data quality, trends, and events. The screen capture in Figure 2 shows plots of NDVI, comparing more than a year of data for two sites in South-East Brazil. As one can see by the parameter setting in the interface, the data are shown only for the period of interest, in daily averages, limiting the data points to measurements between 10am and 2pm local time, and removing data points for seemingly cloudy periods by only including reading whose incoming PAR values are between 900 and 3000 microeinsteins. To be able to handle all these aspects, from pre-processing to generating the plot, a researcher might have needed hours or even days. Other visualization tools are also

available within the Enviro-Net's portal, including spatial distribution and reliability of wireless sensor data, and data availability with associated quality information. Automation and computational infrastructure efforts similar to what has been described here are of the essence to fully realize a global network of sensors as envisioned by Sensor Web related proposals (Delin et al. 2005 and Teillet 2010). Integration and processing aspects precede publication and interaction aspects.

4. CONCLUSIONS

This short paper discussed the efforts underway within the Enviro-Net Project, from deploying and maintaining networks of sensor systems throughout the Americas, to managing the large amounts of data generated by them, to offering easy and flexible access to these data. Our current efforts involve adding other types of datasets to our system, including carbon and water vapor flux data and MODIS vegetation indexes for streamlined comparison of results. Other processing tools are also planned for future versions, as well as implementing programmatic APIs following Web standards for allowing sensor data access from other systems.

REFERENCES

- M.D. Schwartz (ed.). "Phenology: An Integrative Environmental Science," Kluwer Academic Publishers, 2003.
- K.F. Huemmrich, T.A. Black, P.G. Jarvis, J. H. McCaughey, and F. G. Hall. "High temporal resolution NDVI phenology from micrometeorological radiation sensors," *Journal of Geophysical Research*, vol. 104, num. D22, p.p. 27,935–27,944, 1999.
- T.B. Wilson, and T.P. Meyers. "Determining vegetation indices from solar and photosynthetically active radiation fluxes," *Agricultural and Forest Meteorology*, vol. 144, num. 3-4, p.p. 160–179, 2007.
- J.P. Jenkins, A.D. Richardson, B.H. Braswell, S.V. Ollinger, D.Y. Hollinger, and M.-L. Smith. "Refining light-use efficiency calculations for a deciduous forest canopy using simultaneous tower-based carbon flux and radiometric measurements," *Agricultural and Forest Meteorology*, vol. 143, num. 1-2, p.p. 64–79, 2007.

A.V. Rocha, and G.R. Shaver. "Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes," *Agricultural and Forest Meteorology*, vol. 149, num. 9, p.p. 1560–1563, 2009.

G. Barrenetxea, F. Ingelrest, G. Schaefer, and M. Vetterli. "The Hitchhiker's Guide to Successful Wireless Sensor Network Deployments," *Proc. of the 6th ACM Conf. on Embedded Network Sensor Systems (SenSys'08)*, p.p. 43–56, 2008.

M. Welsh. "Sensor networks for the sciences," *Communications of the ACM*, vol. 53, num. 11, p.p. 36–39, 2010.

P.M. Teillet. "Sensor Webs: A Geostrategic Technology for Integrated Earth Sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, num. 4, p.p. 473–480, 2010.

P.W. Rundel, E.A. Graham, M.F. Allen, J.C. Fisher, and T.C. Harmon. "Environmental sensor networks in ecological research," *New Phytologist*, vol. 182, num. 3, p.p. 589–607, 2009.

K.A. Delin, S.P. Jackson, D.W. Johnson, S.C. Burleigh, R.R. Woodrow, J.M. McAuley, J.M. Dohm, F. Ip, T.P.A. Ferré, D.F. Rucker, and V.R. Baker. "Environmental Studies with the Sensor Web: Principles and Practice," *Sensors*, vol. 5, num. 1, p.p. 103–117, 2005.

ACKNOWLEDGMENTS

The Enviro-Net Project is funded by the Canadian Foundation for Innovation and the Inter-American Institute for Global Change Research (IAI) CRN II # 021 which is supported by the National Science Foundation (Grant GEO-0452325). It is also partially supported by Cybera and Canarie (through the GeoChronos Project), as well as the National Science and Engineering Research Council of Canada (NSERC) Discovery Grant Program. The authors acknowledge and appreciate the contributions to the Enviro-Net Project, received in various forms, by numerous members of the local and remote research groups, in particular the students and staff at the University of Alberta, and our research partners at Federal University of Minas Gerais (UFMG), State University of Montes Claros (UNIMONTES), and State University of São Paulo (UNESP), in Brazil; Autonomous National University of Mexico (UNAM), in Mexico; Technology Institute of Costa Rica (ITCR), in Costa Rica; University of Buenos Aires (UBA), and National University of San Luis (UNSL), in Argentina.