Near-real-time environmental monitoring and large-volume data collection over slow communication links

Misha B. Krassovski[1], **Glen E. Lyon**[2], **Jeffery S. Riggs**[3], **and Paul J. Hanson**[1]

[1]Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6290, USA
[2]Campbell Scientific, Inc., 815 W 1800 N, Logan, UT 84321-1784, USA
[3]Integrated Operations Support Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6290, USA

Correspondence: Misha B. Krassovski (krassovskimb@ornl.gov)

Abstract. Climate change studies are one of the most important aspects of modern science and related experiments are getting bigger and more complex. One such experiment is the Spruce and Peatland Responses Under Changing Environments (SPRUCE; http://mnspruce.ornl.gov, last access: 16 October 2018) conducted in northern Minnesota. The SPRUCE experimental mission is to assess ecosystem-level biological responses of vulnerable, high-carbon terrestrial ecosystems to a range of climate warming manipulations and an elevated CO_2 atmosphere. This manipulation experiment generates a lot of observational data and requires a reliable on-site data collection system, dependable methods to transfer data to a robust scientific facility, and real-time monitoring capabilities. This publication shares our experience of establishing a near-real-time data collection and monitoring system via a satellite link using the not very well-known possibilities of PakBus protocol.

1 Data acquisition

1.1 SPRUCE experimental site

The SPRUCE experimental field site is an 8.1 ha *Sphagnum–Picea mariana* (black spruce) bog within the US Forest Service (USFS) Northern Research Station's Marcell Experimental Forest (MEF; 47°30.17′ N, 93°28.970′ W). Access to experimental plots is provided by three 2.5 m wide, aluminum-framed, composite-decked boardwalks above the bog surface. Boardwalks are supported by helical piles anchored into mineral soil to a depth between 12 and 18 m. The boardwalks provide support for electrical power, propane and CO_2 distribution lines, and data transmission infrastructure. In total, 17 plots were initially established (Fig. 1). Each plot contains a 10 m high instrumentation tower with environmental sensors at 0.5, 1, 2, and 4 m above the ground (nominal heights of typical shrub and tree foliage and branches). Ten plots are isolated from the aboveground environment by 8 m high sidewalls and can be supplied with warmed air at ground level and subsequently mixed throughout the vertical space of the enclosure. Of the 10 enclosed plots, 2 are operated without heating (control plots), and the remaining 8 plots are equipped with propane-fired heating and air handling units to deliver duplicate elevated temperature treatments at 2.25, 4.5, 6.75, and 9.0 °C above recorded readings in control plots. Five of the 10 plots, including each of the temperature treatments, are equipped with CO_2 injection equipment to deliver an elevated CO_2 (eCO_2) treatment (500 ppm above ambient). The seven remaining non-enclosure plots are also instrumented and monitored to provide information on undisturbed background ambient conditions. More details about the experiment, warming manipulation studies of ecosystems, and the parameters being measured can be found in Hanson et al. (2017), Krassovski et al. (2015), and on the project's website at http://mnspruce.ornl.gov (last access: 16 October 2018).

1.2 On-site LAN

The SPRUCE on-site data acquisition system is a local area network that is wired using fiber-optic and Cat6 Ethernet cables. This system was designed to handle high-volume sources of in situ observational data and function in a harsh,

Figure 1. Aerial view of the SPRUCE experimental footprint.

remote environment. Air temperature in northern Minnesota can be up to $+40\,^\circ\mathrm{C}$ in summer and drop down to $-40\,^\circ\mathrm{C}$ in winter. Although all network and data acquisition equipment is located inside enclosures that provide protection from the elements, they still may suffer extreme temperature and humidity variations. All networking equipment deployed outside the control building is industrial grade and has extended operating temperature and humidity ranges. Ethernet switches are rated for an operating environment of -40 to $+70\,^\circ\mathrm{C}$ and $10\,\%$–$90\,\%$ relative humidity. Fiber and Ethernet switches are rated for -20 to $70\,^\circ\mathrm{C}$ and $5\,\%$–95% relative humidity.

1.3 Data collection

Data from an instrument or a sensor are recorded by data loggers located in the data acquisition panels that are assigned to each plot. Logger data are then copied via the LAN to the data storage server located in the control building. All data loggers deployed are Campbell Scientific, Inc. (CSI) CR1000 models commonly used for environmental science experiments and long-term meteorological observation stations. The main software component is the LoggerNet product produced by CSI. It supports programming, communication, and data transmission between data loggers and PCs. Use of this package is especially convenient in applications that require telecommunications and scheduled data retrieval within large data logger networks. LoggerNet runs on a computer in the control building and collects data from all data loggers every 30 min. Collected data are copied to an on-site file server and then transferred to ORNL servers through a satellite link. The Fig. 2 diagram shows all components together and gives a full overview of the SPRUCE data acquisition system. A more detailed description of the entire data acquisition system including measurements, in-

struments, sensors, software, and hardware deployed can be found in Krassovski et al. (2015). This publication will look in detail at how the data are transferred from the experimental site to ORNL.

2 WAN data transfer

2.1 Satellite connection

As mentioned above, the experimental site was built in a remote location that had no utility lines in the surrounding area. About 5 km of power cable was buried to supply the site with electricity. Because of low regional population density, the region has no communication lines or reliable cellular coverage. The closest wired Internet access point is more than 16 km away. Estimates showed that the expense to extend wired Internet access to the site was initially too high but may be provided in the future. This reality left only one possibility for Internet connection – a satellite link. Modern technologies can provide fast satellite connections up to $1\,\mathrm{Gb\,s^{-1}}$, but they are very expensive and their use as a permanent connection for data transfer was cost prohibitive for SPRUCE. Consumer-oriented satellite Internet access, however, was affordable and available in low-density population areas. The drawback of consumer-grade connections is the upload speed. Most consumer users browse the Internet, download movies and music, and do not upload a lot of data. The fastest advertised upload speed that was found on the affordable market was $1\,\mathrm{Mb\,s^{-1}}$. In reality it fluctuates between 600 and $800\,\mathrm{Kb\,s^{-1}}$. Another requirement was to find a service that provides a static IP address. Although there is a staff member on-site to monitor and manage the experiment, the ability to have remote access to the local area network is very important, and, as will be shown later, a static IP is required
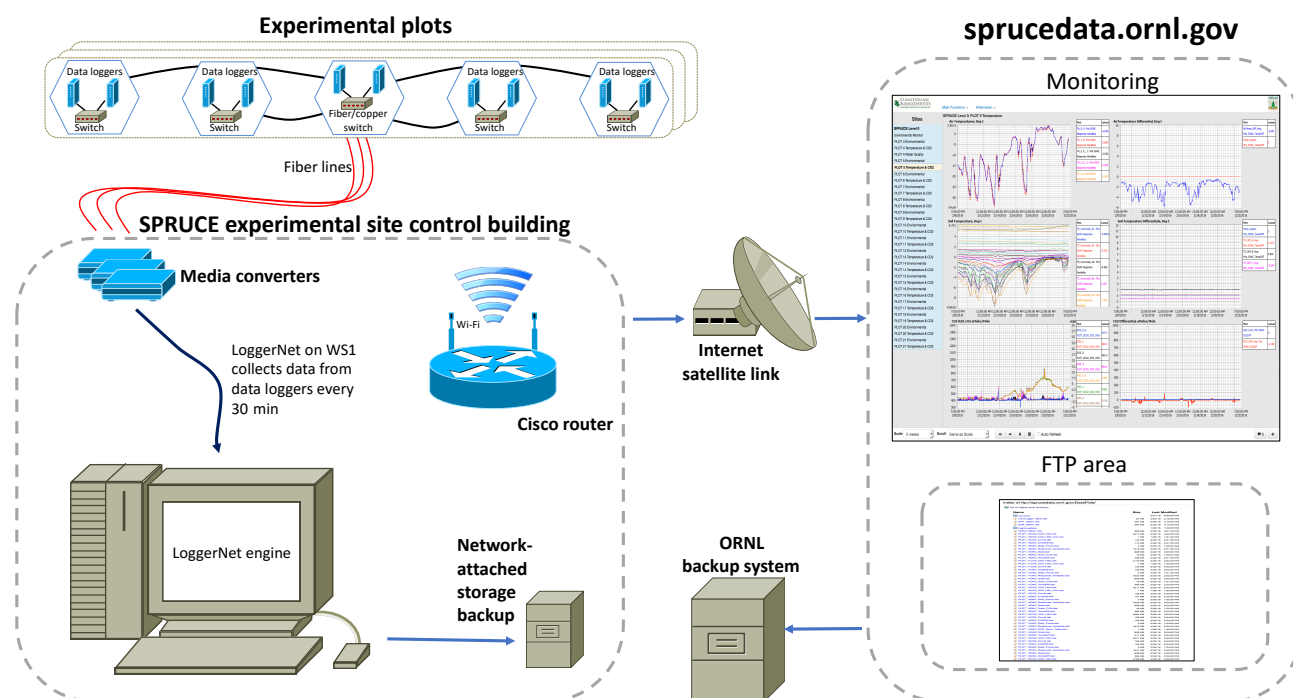
Figure 2. SPRUCE data system.

to set up data collection over the Internet. We found the satellite connection to be very reliable. The data collection system described below can handle interruptions and on-site backup procedures can preserve data when the link is down, but our experience shows excellent connection service availability. During 5 years of use, we had only one major down time caused by on-site satellite equipment failure.

2.2 FTP transfer

FTP is one of the most popular ways to transfer data. It is easy to implement, quite reliable, and easy to automate. A limitation is that it deals with files as a whole. Experimental output data files are managed in 1-year increments (i.e., they contain data from 1 January to 31 December). The amount of data generated by the project is about $25 \, \text{MB} \, \text{h}^{-1}$. In the beginning of a year data files are small, but as the year progresses, files sizes grow, and transfer time gets longer. Another limitation is that satellite connections are sensitive to weather conditions. Heavy rain, snow, or very dense clouds can interrupt the link and data transfer. Every interruption forces the FTP to reinitiate interrupted transfers. Copying only the changed blocks could save time and bandwidth. Many commercial and free software programs implement partial transfers, but only for block-oriented file types like database files, drive images, and virtual disk images. Stream-based files, on the other hand, will usually cause all blocks to be changed whenever they are modified (for example, text

documents, spread sheets, zip files, and photos). Data loggers generate comma-separated value (CSV) text files, which cannot be partially transferred without special software that runs on both sides of the communication line.

2.3 PakBus protocol

As stated before, the main control and monitoring tool at the experimental site is the LoggerNet software product distributed by CSI. The LoggerNet software product offers the tools to remotely accomplish many data logger tasks, including writing data logger programs, transmitting programs to data loggers, collecting data, and analyzing data either in real time or after the data have been saved to a computer. It consists of a communication server application, several client applications and support telecommunications, and scheduled data retrieval used in large data logger networks. LoggerNet supports TCP/IP and can transmit the CSI PakBus protocol over it (LoggerNet, 2015). PakBus is a switched network protocol developed by CSI that works with PakBus-enabled data loggers and other devices. It is a routing protocol and a data logger, or the LoggerNet software can be configured as a router. PakBus is mostly used when a situation does not require setting up a TCP/IP network, but can be implemented over it as well. The LoggerNet server uses PakBus to communicate with data loggers and collect data. The data collection is based on an implemented schedule (every 30 min in this particular case; it is a good resolution for recording diur-
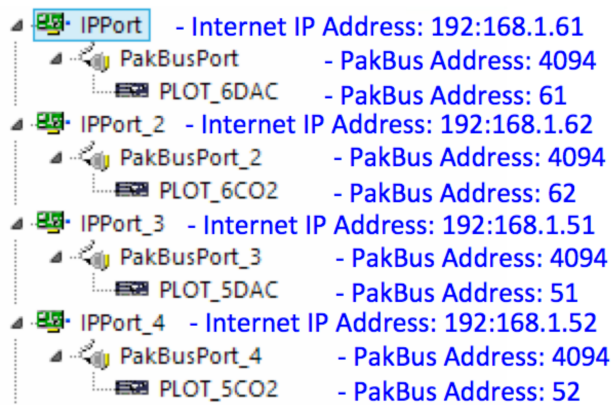
Figure 3. Example PakBus and IP addresses assigned at experimental site.

Figure 4. Example PakBus addresses assigned at ORNL.

nal patterns) and works in the following way. Measurements from sensors and instruments connected to a data logger are written to the internal memory of the data logger. The LoggerNet communication server contacts data loggers, retrieves data from their memory, and writes them to files. The communication server determines what data are new since the last data collection and retrieves only uncollected data, although data logger memories may have days' worth of data stored. This way the data files grow incrementally and only the last 30 min of data are transferred over the LAN.

2.4 PakBus over WAN

After exploring the routing capabilities of the PakBus protocol, it was decided to use it over the satellite connection and take advantage of the incremental data transfer provided by the LoggerNet communication server. Each device in a PakBus network has a unique address. Valid addresses are 1 through 4094. Communications from PakBus devices in the PakBus network with addresses greater than 3999 cannot be ignored by other PakBus devices in the PakBus network and must be responded to. The LoggerNet default PakBus address is 4094. To simplify orientation, data loggers at the experiment site have PakBus addresses assigned according to the plots they are installed in and the LoggerNet server is assigned the default address of 4094. Port 6785 is the default port used for PakBus/TCP communications in Logger-Net. Port forwarding for port 6785 is set up at the SPRUCE router, and the LoggerNet server is configured to bridge Pak-Bus ports, thus functioning as a PakBus router. When bridging PakBus ports in LoggerNet, all data loggers in the network must have unique PakBus addresses, which they do in this case, or a PakBus address conflict will exist. Figure 3 details the LoggerNet data logger network settings at the experimental site. Additional LoggerNet server software was installed at ORNL and assigned a PakBus address of 4093 to avoid PakBus address conflict with the LoggerNet soft-
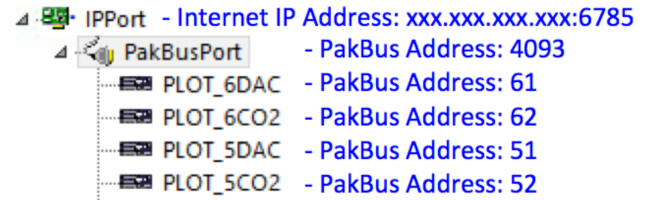
ware at the experimental site. The configuration depicted in Fig. 4 is implemented at ORNL. The IP address of the IP port in Fig. 4 is the WAN IP address of the SPRUCE router. All of these arrangements allow the LoggerNet server at the SPRUCE site to act as a PakBus router and allow the Logger-Net server at ORNL to communicate with devices on remote LAN and retrieve data.

2.5 Near-real-time data collection and monitoring

As previously mentioned the LoggerNet communication server gathers only data stored in the data loggers since the last scheduled data collection and transfers only those small data packages over the satellite link. At the present time, it takes about 30 min to move all new data from the experimental site to ORNL. For redundancy and as an additional backup, LoggerNet data collections are scheduled on both on-site and ORNL servers and data are accumulated in both locations. Although technically unnecessary, to optimize data collection simultaneous data collection by the two Logger-Net communication servers is avoided by offsetting their data collection schedules by 15 min. At the SPRUCE site, it happens twice in an hour at XX:00 and XX:30, and at ORNL it is scheduled at XX:15 and XX:45. It allows for the monitoring of data with a 30 min delay on-site and about a 1 h delay at ORNL (it varies slightly based on the satellite link speed). The data received are also used as a constant feed to data visualization software (Vista Data View, Vista Engineering) for plotting all site data in variable time series formats. Such a visualization package enables remote assessments of equipment performance and data quality assurance. Figure 5 shows an example of measured events at experimental plot 19.

3 Conclusion

Ecosystem-scale manipulation experiments are getting more complicated and require innovative approaches that help manage high volumes of in situ observations. New large-scale experiments in remote locations will become common in the future and will require reliable data acquisition systems. Most of those systems will be based on slow, limited-bandwidth satellite or cellular links. The use of Campbell Scientific equipment and bundled software is very common
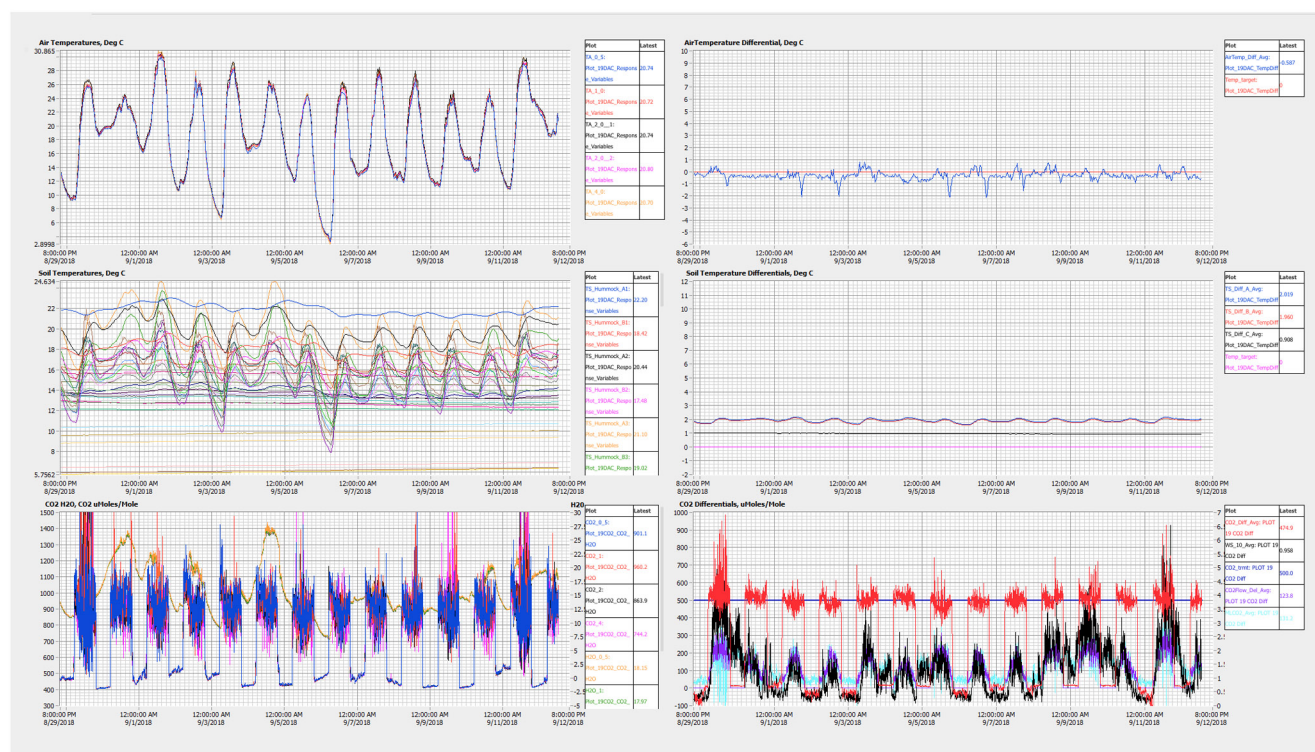
Figure 5. Monitoring processes at plot 19 via VDV software at ORNL.

and almost a de facto standard for environmental studies. The presented approach shows an example of using existing nonstandard network technologies to overcome difficulties caused by slow communication links. We hope that the provided details about network technologies, hardware and software configuration, and the logic behind these choices will help the future design and development of similar systems for other experiments.

Data availability. Data is publicly available at https://mnspruce. ornl.gov/public-data-download (last access: 16 October 2018).

Appendix A

Abbreviations

SPRUCE: Spruce and Peatland Responses Under Changing Environments
MEF: Marcell Experimental Forest
DOE: US Department of Energy
TES: terrestrial ecosystem science
ORNL: Oak Ridge National Laboratory
CSI: Campbell Scientific, Inc.
LAN: local area network
WAN: wide area network
CVS: comma-separated value
FTP: file transfer protocol
TCP/IP: transmission control protocol/internet protocol

References

Hanson, P. J., Riggs, J. S., Nettles, W. R., Phillips, J. R., Krassovski, M. B., Hook, L. A., Gu, L., Richardson, A. D., Aubrecht, D. M., Ricciuto, D. M., Warren, J. M., and Barbier, C.: Attaining whole-ecosystem warming using air and deep-soil heating methods with an elevated CO_2 atmosphere, Biogeosciences, 14, 861–883, https://doi.org/10.5194/bg-14-861-2017, 2017.

Krassovski, M. B., Riggs, J. S., Hook, L. A., Nettles, W. R., Hanson, P. J., and Boden, T. A.: A comprehensive data acquisition and management system for an ecosystem-scale peatland warming and elevated CO_2 experiment, Geosci. Instrum. Method. Data Syst., 4, 203–213, https://doi.org/10.5194/gi-4-203-2015, 2015.

LoggerNet: Instruction manual, Version 4.3, Campbell Scientific, Inc., 815 W 1800 N, Logan, UT, 84321-1784, 2015.