Basin-scale Hydrologic Experiment by Means of a Wireless-Sensor Network System

by

Ziran Zhang

A dissertation for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Steven D. Glaser, Chair
Professor Roger C. Bales
Professor Kristofer J. Pister

Fall 2016

Basin-scale Hydrologic Experiment by Means of a Wireless-Sensor Network System

Abstract

Basin-scale Hydrologic Experiment by Means of a Wireless-Sensor Network System

by

Ziran Zhang

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Steven D. Glaser, Chair

The use of wireless sensor networks(WSN) to study the mountain hydrologic cycle has been proposed as a supplement to existing systems used for ground-based hydrologic and ecological monitoring. Historically, the study of mountain hydrology and the water cycle has been largely modeling with limited data, with meteorological forcing and hydrological variables extrapolated from a few infrequent manual measurements. Recent developments in Internet of Things (IoT) technology are revolutionizing the field of mountain hydrology. Low-power WSNs can now generate denser data in realtime and for a fraction of the cost of labor-intensive manual measurement campaigns. This research details the requirements and different technical options, describes the technology deployed in the American River basin, and discusses the methods associated with modeling large-scale environmental monitoring in extreme conditions. The American River Hydrologic Observatory (ARHO) project has deployed fourteen low-power wireless IoT networks throughout the American River basin to monitor California's snowpack. A network of sensors for spatially representative water-balance measurements was developed and deployed across the 2154 km^2 snow-dominated portion of the upper American River basin, primarily to measure changes in snowdepth and soil-water storage, air temperature, and humidity. The WSNs, each has 10 measurement nodes that were strategically placed within a 1-km^2 area, across different elevations, aspects, slopes and canopy covers.

The research evaluates the accuracy of a machine-learning-based path-loss model to estimate the expected radio-transmission distances. The model is trained on 42,157,324 RSSI samples collected over seven months from the ARHO WSN. The 2218 links in the network span across the upper portion of the American River basin and are deployed in a complex environment, with large variations of terrain attributes and vegetation coverage. The model is based on an ensemble regression-tree machine-learning algorithm (Random Forest). The independent variables used in the model include path distance, canopy coverage, terrain variability, and path angle. The accuracy of this model is compared to several well-known canonical and empirical propagation models. This model showed a 37% reduction in the average prediction error compared to the canonical/empirical model with the best performance. The research presents an

in-depth discussion on the strengths and limitations of the proposed approach as well as opportunities for further research.

Compared to existing operational sensor installations, the ARHO WSN reduces hydrologic uncertainty in at least three ways. First, redundant measurements improved estimation of lapse rates for air and dew-point temperature. Second, distributed measurements captured local variability and constrained uncertainty in the air and dew-point temperature, snow accumulation and derived hydrologic attributes important for modeling and prediction. Third, the distributed relative-humidity measurements offer a unique capability to monitor upper-basin patterns in dew-point temperature and better characterize elevation gradient of water vapor-pressure deficit. Network statistics during the first year of operation demonstrated that the ARHO WSN was robust for cold, wet and windy conditions in the basin.

Using daily dew-point temperature and the amount of snow accumulation at each node to estimate the fraction of rain versus snow resulted in an underestimate of total precipitation below the 0 ^{o}C dew-point elevation, which averaged 1730 m across 10 precipitation events. Blending lower-elevation rain-gauge data with higher-elevation sensor-node data for each event provided precipitation estimates that were on average 15-30% higher than using either set of measurements alone. Using data from the current operational snow-pillow sites give even lower estimates of basin-wide precipitation. Given the increasing importance of liquid precipitation in a warming climate, a strategy that blends distributed measurements of both liquid and solid precipitation will provide more-accurate basin-wide precipitation estimates. The distributed, representative sensor-network measurements also improve upon operational estimates of snowpack water storage, snowmelt amount and snowmelt timing across the basin.

To the love and support of my parents Ming Zhang and Shaoli Chen who made this possible and meaningful.

Contents

List of Figures

List of Tables

Acknowledgments

I would like sent my deepest appreciation to my graduate advisor Steven D. Glaser's unconditional support and guidance throughout my years of exciting life as a graduate student. I would like to send my warmest gratitude to Professor Roger C. Bales's instructions that guide me through the research. In addition, I would like to thank Professor Martha Conklin, Drs. Danny Marks and, Robert Rice, who gave me unfettered support from their experience and expertise.

I would like to thank my ongoing collaborator, Dr. Thomas Watteyne and Dr. Branko Kerkez, for their on-demand feedback and support.

I would also like to thank Kristofer Pister for dedicating time, interest, and effort in reviewing my research.

I thank my fellow students Sami Malek, Carlos Oroza, Zeshi Zheng, Chris Sherman, Mario Magliocco, and Paul Selvadurai for the insightful discussion and assistance throughout my years as a graduate student.

Chapter 1

Introduction

Quantitative information on mountain water cycles at the basin scale almost non-existent; the limited measurements that are available are largely provided by a few operational precipitation, snowpack-storage, and stream-gauging stations. These tend to represent a limited range of elevations, weather and snow conditions, with precipitation and snow measured in the upper regions and streamflow in the lower regions of the mountain basin. The data from mountain basins is limited, and limited information about the massive gradients that occur from basin outflow to headwaters across the western United States. Research networks include a few selected headwater catchments where a complete set of meteorological and hydrological attributes are carefully measured. While these catchments offer specialized information on some aspects of mountain hydrology, they provide only a limited understanding the hydrology of larger mountain river basins. Climate warming is raising the rain-snow transition elevation and changing the hydrologic response of mountain basins. While we can see these changes in data from research catchments, extending these findings to the larger mountain basin remains a challenge. At the scale of large mountain basins, evapotranspiration, soil moisture, and groundwater recharge/discharge are generally not measured and the spatial properties of energy-balance variables driving the mountain hydrologic cycle are poorly measured across the region.

Snowmelt from the Sierra Nevada provides more than half of the developed water supply for the state of California. A severe drought in the recent years has motivated an increased stakeholder interest in a finer monitoring of this resource. There is a growing need for better streamflow predictions to allow more-efficient dam operations, for hydropower generation, flood control, and drought mitigation. Different strategies have recently gained attention from both governmental agencies and utilities companies, and are being employed to improve snowpack monitoring. They include remote-sensing (MODIS), aerial Light Detection and Ranging (LiDAR), in-situ sensors, weather forecasts and ideally their synergistic combination. Though satellite remote sensing typically covers wide areas, it lacks fine spatial resolution, and thus is complemented with finer resolution limited range techniques such as LiDAR and in-situ sensors. Snowpack and soil-moisture storage, and the fluxes into and out of these reservoirs, are not precisely measured; and the spatial properties of energy-balance variables driving the

mountain water cycle are poorly quantified across the region.

A basin-scale instrument cluster was developed, as part of the goal of this dissertation, to make comprehensive water-balance measurements across the rain-snow transition in the American River basin of the Sierra Nevada. This is the American River Hydrologic Observatory (ARHO, arho.org). Quantifying the water balance at greater spatial and smaller temporal resolutions is critical to research in atmospheric science, biogeochemistry, ecosystem science, and water resources, and is central to research in hydrologic science. It has enabled new science for which no adequate infrastructure currently exists. Such a rich data set will enable development of new classes of modeling tools to produce quantitative climate assessments, influence hydrologic forecasting, probe system response to climate and land-cover perturbations, increase process understanding of basin-scale water cycles, and provide defensible scenarios for infrastructure planning all over a scale currently not possible. New measurement will promote new models to be developed for more accurate representation of the processes governing snow and snowmelt.

1.1 Document organization

Three research questions motivated the research reported in this document:

- What are the requirements for WSNs to become a viable ground-based measurement system, and what is the methodology needed for designing and deploying WSN systems for snow monitoring?
- How can a new network connectivity model help guiding future WSN deployments?
- What hydrologic insights can a spatially distributed and temporally resolved information provided by WSN offer?

The dissertation is organized to answer these questions.

Chapter 2 examines the requirement, design, and deployment methods of using WSNs. This chapter provides a detailed description of the technology and the system concept, architecture, hardware, and software. Some existing (wireless) sensor network platforms were compared to the system used in this research in this chapter to illustrate how they are different.

Chapter 3 uses real-world data, collected from the ARHO, to develops a connectivity model of wireless signals. The results are compared to existing propagation models to show a 37% reduction in the average prediction error using the proposed machine-learning method.

Chapter 4 gives a formal description to the methodology of using WSN as a main tool for in-situ snow monitoring in the American River basin. It discusses the methods for using WSN technology in mountain hydrology studies through an analysis of an extensive monitoring campaign. Data-use scenarios are also given in this chapter. An evaluation of the field performance of our WSN system devices is also provided.

Chapter 5 provides insights into mountain precipitation and snowpack through analysis of the data. This chapter describes how to better characterize the rain/snow transition zone on

the mountain slopes using the WSN. It also provides a quantitative estimate of the amount of snowpack on the mountain and discusses the snowmelt patterns.

Chapter 2

Wireless sensor network preliminaries

2.1 Introduction

There is an increasing amount of attention and resources spent on developing WSNs in scientific and operational research. The momentum of adopting wireless solutions is fueled by reduced production cost of wireless equipment, advances in the development of networking protocols, and the synergies wireless technology creates by combining traditionally wired sensing equipment with a wireless platform [4, 33, 120]. Among different kind of wireless applications, wireless platforms/test-beds in outdoor environments are especially enticing to researchers because of the challenges and the opportunities offered; it is simply too impractical to run wires for hundreds of meters through the woods. Over the last ten years, many WSNs have been deployed over various spatial scales, showing some success in monitoring environmental phenomenon with distributed arrays of sensors. Dozens, sometimes hundreds of wireless sensor nodes hosting non-trivial numbers of sensors were put into outdoor environment to measure and monitor air temperature [66], relative humidity, snow depth [56], soil moisture [9, 14, 65], permafrost [44], forest fire [42], potential landslides [104]and volcanic activities [112]. The use of WSNs and actuators for monitoring and controlling various aspects of agricultural activities had also gained tremendous attention over the years [11, 57, 40].

WSN is a field that has grown out of the initial work done at the University of California, Berkeley [85, 70, 47]. Coming from the concept of "smart dust", the expectations is extremely low power consumption so the network nodes can run off batteries for extended periods of time, which means low power radio since the transceiver is the largest user of the sensor node, known as a "mote". The motes should form a reliable self-managed self-assembling redundant mesh network. Each mote carries computation, memory, and all the necessary signal conditioning to operate the needed sensors. The motes on the network are closely time-synchronized so that temporally-based analyses are possible. Given a large enough power budget all these tasks are not difficult.

There is special interest in developing WSN platform to perform sensing and real-time monitoring tasks in remote mountainous watersheds. Historically, the study of mountain hydrology

and the water cycle has been largely observational, with meteorological forcing and hydrological variables extrapolated from a few or even individual measurement sites situated at non-representative locations [89, 93, 94, 24]. Traditionally, very few point measurements were made over usually very large area by sensors hosted on met towers situated at non-representative locations [19]. Large deployments of WSN in the wilderness represent an increasing demand for detailed information about hydrological and meteorological events. The data WSNs can provide are crucial for advancing understanding of watershed science [8]. Distributed measurements of hydrological variables were done by painstaking manual surveys at a far coarser temporal resolution (monthly) compared to what WSNs can provide. Samples in a WSN can be taken and transmitted at a sub-hourly frequency and the motes are dispersed in patterns that capture the variance and mean of the catchments to provide detail insight needed for accurate modeling and predictions.

2.2 Previous studies of outdoor WSN applications

In situ hydrological and meteorological information is difficult to obtain from distributed locations due to logistic constraints associated with harsh weather. Many studies have been performed to examine the possibility of using WSNs to obtain more detailed knowledge from remote watersheds. Early campaigns of using WSN for habitat monitoring lead by Berkeley engineers had shown both the both the opportunities and technical challenges in applying this new technology to environmental monitoring [70, 103]. A 12 stations deployment, from June to October 2009, in a 20-km^2 catchment of the Swiss Alps measured the spatial variability of meteorological forcing parameters such as temperature and precipitation [101]. Recently, densely deployed WSNs have been scaled to a size comparable to the mountaintops being studied. Ninety-nine sensor-loggers, within three 40-180 km^2 basins, were deployed to monitor snow cover dynamics in southern Germany for one winter [86]. Bogena et al.[14] installed 150 wireless nodes, with over 600 soil moisture sensors in the forest catchment of Wustebach, Germany to study the spatiotemporal distribution of soil moisture over complex terrain. Over 300 sensors hosted by 60 wireless nodes are deployed at the Southern Sierra Critical Zone Observatory to study complex interaction of water within the snowpack, canopy, and soil [56].

2.3 Requirements and options

Remote environmental sensing places unique requirements on the network interconnecting the sensors. Section 2.3 starts by listing those requirements; Section 2.3 then details the networking options in light of those requirements. The goal of the ARHO project is *not* to develop a new networking technology, but rather to “bring it all together” by integrating the appropriate off-the-shelf networking solution with environmental sensors.

Requirements for remote environmental sensing

Placement of the sensing nodes is generally expertly determined and depends mainly on some hydrological characteristics of the terrain. The networking technology must be able to support the placement of nodes at arbitrary locations. Moreover, for the collected data to be physiographically representative, the network of sensors must be able to cover 1-2 km^2 and be comprised of up to 60 network devices.

Sites of interest are scattered throughout the Sierra Nevada, and reaching some sites requires hours of driving through forest roads that are closed during the winter. The networks must, therefore, be able to run "on their own" without any mandatory user intervention. They must be able to withstand large temperature swings, self-heal in the case of environmental changes, and hopefully run for several years without human intervention.

Each node must be able to report sensor measurements . Assuming a 60-node network, this translates in the network generating one packet every 15 seconds. The network must successfully and securely transport the sensors data to a gateway device [38, 98]. The user must be offered remote monitoring capabilities, i.e. the ability to remotely assess the health of the network. Hydrologists must be able to access the data gathered by the sensors within minutes of when it was measured.

Present solutions for WSNs : A comparison of existing technologies

Environmental data loggers such as the Campbell CR1000, very well know in the hydrology community, store time-stamped sensor data on the internal memory, which can be retrieved manually. While off-line logging does not satisfy the "real-time" requirement, it is an interesting option, at least for keeping a copy of the data locally in case of networking problems.

Wired equipment

A first networking option is to use wired technology to interconnect the different sensors. This is the technology used by commercial products such as Campbell. This is not an option for the ARHO for the following reasons. First, there is significant additional cost and time associated with covering a 1-2 km^2 with wires. Second, the state of California and the US Forest Service has strict rules on deploying equipment in the Sierra Nevada; interconnecting sensors using wires would be too intrusive. Finally, the weight of snow-pack, falling objects, and animal curiosity are likely to inflict damage to wires laying on, or shallowly buried, in the ground [94, 57].

Cellular, WiFi, BlueTooth technology

Wireless is the technology which fits the requirement of Section 2.3 the best. Cellular coverage on some sites is non-existent, spotty on others; equipping each sensor with a cellular modem is therefore not an option. An IEEE802.11 (low-power) WiFi network would force a star topology; given the size of the deployments, WiFi doesn't offer the range required. WiFi consumes a comparatively very large amount of power compared to the mote. The range of Bluetooth and the

size of piconet prevent us from using this possible solution. To offer a 2-year battery lifetime, assuming a 17,000 mAh Li-Ion battery, a communicating node must draw at most 970 uA, on average.

Sensor networks are made up of motes. This term, from its definition "a small speck of dust," was coined at UC Berkeley to describe a very small, low-power device that incorporates a radio transceiver, computational power, data storage, and sensors. There are many commercially available motes with many different standards. Different motes were designed for different purposes such as research, education, hobbyists, indoor industrial or outdoor monitoring and control applications. The question is why to pick one hardware solution over another? Our primary consideration focuses on how well the networking protocols were implemented to ensure performance and robustness. We also look into the hardware flexibility to satisfy needs for interfacing with different sensors. We focus on the most popular of motes that comply with the IEEE 802.11.15.4 standards for power-savings reasons and availability. IEEE802.15.4 [52] is a wireless technology which can satisfy the requirements. It offers a good balance between data rate (250 kbps), communication range (50-100 m outdoors) and power consumption (a radio chip consumes 5-25 mA when on, depending on the vendor). Communication protocols enable the radio to be heavily duty-cycled (enough to reach the 970 uA on average current draw above) and form a multi-hop topology. The challenge is that IEEE802.15.4 is "just" the physical layer; a complete flurry of protocols, products, and open-source implementations build on top of it. Properties and specifications of two main families of motes along with our solutions were investigated.

Mica-II, Iris, TelosB, Lotus

Memsic Inc. provides a number of low-power motes (MICA, TelosB, and Lotus) that are 802.11.5.4 compatible. The MICA and TelosB design go back almost twenty years to the early UC Berkeley work. The LOTUS mote with a Cortex M3 processor and ZigBee radio provides the most internal memory for the OS and application software among Memsic motes. Several operating systems (RTOS, MoteRunner and TinyOS) can be ported to LOTUS. However, some problems with the network-routing protocol and channel-management protocols remain. Operating on a single channel makes them vulnerable to network instabilities resulting from signal interference and multi-path effects. The current draw from the LOTUS and MDA300 board is estimated to be around 17 mA at 3V when transmitting. The mote depletes two AA batteries in approximately five months when set to transmit for 3 seconds every 15 minutes. The data-acquisition board (MDA300) provides seven single-ended and one (multiplexed to four) differential 12-bit ADC channels. In addition, the digital I/O support is very limited on this board.

ZigBee and Xbee

ZigBee is a protocol stack that builds on top of IEEE802.15.4 radio. ZigBee motes represent a popular family of wireless motes that share common communication protocols and specifications (network layer and application layer) by implementing a ZigBee software stack [21].

In practice, ZigBee operates as a star network. Although it is theoretically possible to form a mesh-like network topology, it has two major shortcomings. Firstly, a subset of motes has to be pre-selected and programmed as dedicated routing nodes to relay data from the end/leaf node to the coordinator. The RX channel of those router elements has to be constantly powered, which results in high overall energy consumption of 5-25mA [49], well above the 970 uA limit. Secondly, ZigBee operates on a single channel, making it difficult to avoid channel interference due to other ISM sources such as WI-FI, and multipath [2]. In Gungor 2009[37] and Huang 2012[50], the authors discuss how this prevents ZigBee from meeting industrial requirements, and is therefore also our requirements. Recent large-spread hacker attacks have shown the ZigBee network security to be non-existent.

Digi International maintains a family of motes called XBee. The XBee 2.4 GHz-band mote has its own proprietary protocol called DigiMesh that seriously suffers from environmental interference and varying effects of multipath because it does not implement channel hopping [53]. In order to achieve low power in a DigiMesh network, the system needs to enter a synchronized sleep mode. Due to the lack of a central network coordinator, a subset of the DigiMesh motes needs to be constantly running to serve as sleep coordinators (i.e., network manager). Those motes continuously broadcast sync message to the surrounding nodes to keep the network assembled, otherwise the message transmitted to a mote during the sleeping period will be permanently lost [53]. Another possible issue for DigiMesh network in synchronized sleep mode is that when a new mote is added to the network, it needs to be physically near a sleep coordinator to receive a sync message in order to join the network [118]. If a node temporarily drops out, it is permanently lost, and an extended trial and error installation is difficult. XBees are commonly interfaced with Arduino single-board computers to provide facilities to host sensors, compute, and store data. The Arduino Uno R3, with no external load from sensors and other components, consumes about 40 mA at 5V. The Uno R3 uses a slow and outdated Atmega328P (8-bit/16MHz) microprocessor with a 10-bit analog to digital converter that provides only six analog pins and fourteen digital pins to interface with the sensors and other equipment. Similar issues with power and flexibility can be found with solutions provided by Raspberry Pi, which consumes 700 mA at 5V, making it impractical to operate with a battery. Systems with Arduino and Raspberry Pi are best kept indoors where sufficient power input is provided. They are not recommended for long-term outdoor deployments, as they were designed for hobbyist use.

SmartMesh IP

SmartMesh IP is an off-the-shelf low-power wireless networking product, developed by the Dust Networks product group at Linear Technology, which satisfies all requirements listed in Section 2.3. A SmartMesh IP network is multi-hop in nature, with a single gateway device (called "manager") and up to 100 motes. The protocol stack it uses is rooted in IEEE802.15.4e Time Synchronized Channel Hopping (TSCH), a new networking technology in which nodes are tightly synchronized, and where a communication schedule orchestrates all the communication in the network [109]. This results in over 99.999% end-to-end reliability, and router nodes drawing less

than 50 μA on average [107]. The capacity of a network is 25 packets per second, which is well above the one packet every 15 s requirement of our application.

A SmartMesh IP network is composed of a manager and up to 100 motes. The motes form a redundant multi-hop and self-healing wireless mesh network around the manager.

The core of a manager or mote is the LTC5800 chip, specifically designed for SmartMesh IP, and which features an ARM Cortex-M3 microcontroller and an IEEE802.15.4-compliant radio chip. Because it is specifically designed for SmartMesh IP, the LTC5800 features hardware acceleration for common operations, resulting in ultra-low-power operation.

45,000 SmartMesh networks are deployed today, in application domains ranging from industrial process monitoring and city-wide parking solutions to building monitoring and remote sensing. SmartMesh IP provides over 99.999% end-to-end reliability and less than 50μA average current draw on routing nodes, translating to over a decade of lifetime on typical batteries.

The network stack used in SmartMesh IP is rooted in the IEEE802.15.4e TSCH standards. In a TSCH network, nodes are tightly time-synchronized, and time is cut into timeslots. All communication is orchestrated by a schedule that indicates to each mote what to do in each timeslot: transmit, receive or sleep. The schedule is built in such a way that nodes only switch their radio on when they actually communicate, yielding radio duty cycles well below 1%. When communicating, nodes sent successive packets at different frequencies. The resulting channel hopping is known to combat external interference and multi-path fading efficiently [2, 12, 23, 69, 84, 106], the two main sources of unreliability in low-power wireless networks.

The network stack used in SmartMesh IP combines the performance of TSCH (through the IEEE802.15.4e standard) and the ease-of-use of IPv6. Each SmartMesh IP devices can be assigned a globally unique IPv6 address, significantly simplifying Internet integration. Through several sets of keying material, and the use of an AES-128 cipher, all frames exchanged in a SmartMesh IP network are secured. At any point in time, a device can be administratively removed from the network. Thanks to the underlying TSCH technology, nodes are tightly synchronized (a handful of μs maximum de-synchronization between neighbor nodes). This means the events measured can be accurately timestamped, or that several nodes in the network can coordinate actuation.

The manager plays an important role. First, it serves as the gateway between the low-power wireless mesh and the Internet. Second, it is responsible for building and maintaining the TSCH schedule. The manager monitors the topology of the network and the application-level communication requirements of the nodes, and continuously adjusts the schedule accordingly. The result is that the network adapts to topological changes. Finally, the manager is responsible for coordinating the join process of new motes, which includes a security handshake, and handing out the appropriate keying material.

In a SmartMesh IP network, each node has at least two routing parents. In case one of the parents is switched off or removed, the second parent is used to reliably route the node's packets. To be able to exploit this path diversity, nodes need to be deployed in a sufficiently dense manner. The rule by Dust Networks is that each node should be able to "hear" at least three other nodes.

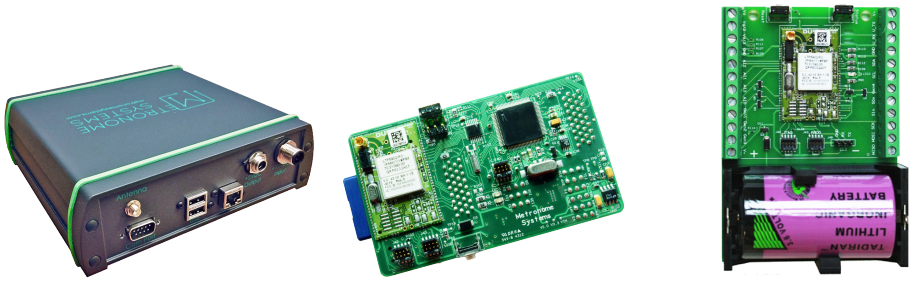
The TSCH schedule provides a clean trade-off between bandwidth and power consumption. The more data a mote has to send, the more activity it will have in its communication schedule, and hence the more it consumes. Based on the communication requirements, it is possible to accurately predict the current draw of a node [107], which typically ranges between 10 μA and 50 μA.

2.4 Metronome Systems neoMote

Metronome Systems provides a comprehensive solution for the sensor node called the neoMote, which was developed at UC Berkeley. It combines the DUST Networks Eterna radio module with a Cypress Programmable System on Chip (PSoC5) into a two-chip solution. While DUST Network radios provide robust and reliable wireless networking capability, PSoC provides full support to any peripheral. The PSoC offers an array of configurable system blocks that can be dynamically added to a project for a particular application. For instance, the board can interface up to 40 analog and/or digital sensors at once, providing all analog and digital signal conditioning and excitation. The PSoC building blocks are available to a drag-and-drop interface and are reprogrammable over the radio. The neoMote provides 3.3, 5, and 12Vdc excitation to sensors. Interfacing with an SD-card slot provides additional storage for data and system parameters. In addition, the board is ultra-low power. Power consumption is 30 mA, 60 in 20-bit A/D mode, with transmission adding 10 mA, two to three orders of magnitude lower than the previous solutions. The network is controlled by a Metronome Systems network manager, which also interfaces the data with the outside world. It is based on a full LINUX computer, while only consuming 50 mA at 5V. It runs a full database and sends the data out through a variety of modems. DUST Networks, a division of Linear Technology, provides an industrially rated ultra-low power fully-meshed wireless networking platform. The dynamic network allows seamless joining and rejoining by any mote or hopper. A few technical details properties of the DUST mote make it superior to other choices. The SmartMesh IP software utilizes time synchronized mesh protocol (TSMP) that maintains complete network synchronization to 10 mS, which minimizes the “on-time” to listen for the beacon.

Incorporating TSCH reduces interference within the communication channels through the diversity of frequency at which each packet is sent [85]. Adding diversity to the channel selection reduces the adverse effect of multipath fading in the wireless network. The typical duty cycle of the DUST radio module is < 1% while keeping communication reliability 99.999%. The DUST network is unique in that it constantly collects a wide variety of network statistics, which allows for the later optimization of a network.

The Metronome system provides for Internet-of-Things capability, such that one can deliver programs remotely to sensor nodes to resync real-time clock settings, change firmware, sampling interval, sensor gain, etc.



(a) Manager (b) NeoMote (c) repeater node

Figure 2.1: The Metronome Systems hardware used in the ARHO project.

The Metronome Systems Ecosystem

Linear Technology sells generic communication modules (a small 24 mm × 42 mm board), which can be easily integrated into a complete solution. Metronome Systems (http://metronomesystems.com) has done this integration, and offers an ecosystem of SmartMesh IP-enabled boards, which the ARHO project is using. Section 2.4 details the Metronome Systems ecosystem we use. Metronome Systems' "ecosystem" of low-power wireless devices is chosen in this application.

The Manager (Fig. 2.1a) is a single-board Linux computer connected to a SmartMesh IP manager module, encased in a robust and water resistance enclosure. Besides the advantage of being built to cope with the extreme Sierra Nevada weather, the Manager consumes only 50 mA at 5 V on average, much less than the raspberryPi (700 mA) or Arduino boards (500 mA). In the ARHO project, the Manager is connected to a cellular or satellite modem, as detailed in Section 2.4.

The NeoMote (Fig. 2.1b) is the sensing heart of the ARHO network. It is composed of a SmartMesh IP module and a Cypress Programmable System on Chip (PSoC5). The SmartMesh IP module provides robust and reliable wireless networking capability, the PSoC interfacing capabilities to any sensor and actuator. The PSoC offers an array of configurable system blocks that can be dynamically added to a project for a particular application. For instance, the board can interface up to 40 analog and/or digital sensors at once, providing all analog and digital signal conditioning and excitation. The PSoC building blocks are available to a drag-and-drop interface and are reprogrammable over the radio. The NeoMote supports 3.3V, 5V and 12V sensors. Interfacing with an SD-card slot provides additional storage for data and system parameters. In addition, the board is ultra-low power. The PSoC consumes 30 μA at 3.3 V, 60 μA when the 20-bit A/D mode is active. In the ARHO project, the NeoMote is connected to sensors and mounted on a sensor node as detailed in Section 2.4.

The repeater node (Fig. 2.1c) consists of a SmartMesh IP mote and a carrier board with a D-cell 17 Ah battery. In the ARHO project, the repeater node is used as a relay in case the sensor nodes are spread too far apart in the deployment field, as detailed in Section 2.5.

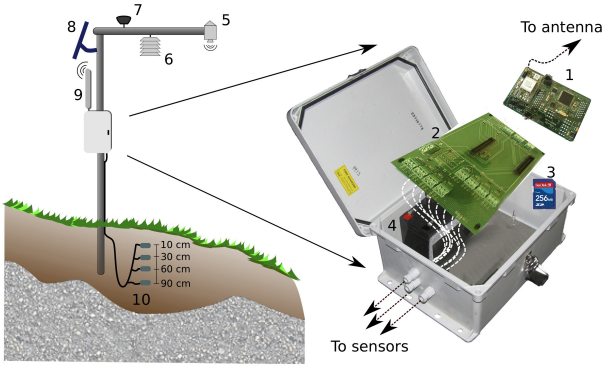


Figure 2.2: A sensor node is composed of: (1) a NeoMote low-power wireless mote; (2) a breakout board; (3) a memory card for local data logging; (4) a rechargeable battery; (5) a snow depth sensor; (6) temperature and relative humidity sensors; (7) a pyranometer (solar radiation) sensor; (8) a solar panel; (9) an omni-directional antenna; and (10) underground volumetric soil moisture sensors.

The sensor node: Interfacing sensors

The NeoMote is the "sensing heart" of an ARHO network. Through the PSOC microcontroller, it can connect to any sensor or actuator, through digital or analog interfaces. In the ARHO project, the NeoMotes with sensors attached were enclosed in a waterproof enclosure mounted on a mast. These poles, typically 5 m high, were secured into the ground at strategic locations throughout the deployment site.

Fig. 2.2 depicts such as sensor node. A waterproof box contains all the electronics, including a rechargeable battery. This box is typically mounted 2 m above the ground, so it can be accessed both when there is, and when there is no snow. A number of elements are mounted on a cross arm at the top of the pole. These include solar panel, as a well solar radiation, temperature, humidity and snow depth sensor. The latter is an ultrasonic range finder oriented down; it sends "chirps" and measures the time it takes for the sounds to bounce on the snow and return to its built-in microphones. Soil moisture sensors are placed at different depth in the ground to monitor the water seeping through the soil. All these elements are connected to the main electronics box through kevlar tubing, after animals, including bears, chewed and broke traditional electric wiring in early deployments.

The sensor nodes were pre-assembled in the laboratory, and transported by truck to a deployment site. At the site, the cross arm and mast were fastened, and the pole was secured into the ground using a U-channel. Installing a single pole takes approximately 1 hour for a team of two.

The base station: Internet connectivity

The manager is attached to a sturdy 3 in diameter, 4 m long steel mast secured on a concrete foundation. The manager is connected to a cellular modem (Sierra Wireless RavenX) or satellite modem (Inmarsat BGAN 9502). When available, we prefer cellular connectivity as it is cheaper and provides more stable services. Five sites are connected using cellular, five sites using satellite. The other four sites have their gateways yet to be installed.

The main role of this cellular/satellite connection is to upload the sensor data and network statistics onto the servers located in the back-end at UC Berkeley (see Section 2.6). In addition, it allows one to log into the Linux computer of the manager remotely to inspect the network, reprogram the network, and update the software running on the manager.

2.5 Deployment strategy

An ARHO network is composed of one manager, 10-11 sensor nodes, and a number of relay nodes in case some sensor nodes are deployed too far apart from the rest of the network to have radio connection.

We deployment an ARHO network as follows. The locations of the sensor nodes are chosen based on their hydrological importance and representativeness. In a typical deployment, the sensor nodes cover an area between 1 km^2 and 1.5 km^2. The location of the manager is chosen so it has good cellular connectivity, or in an open area for good satellite connectivity.

Once the sensor and gateway poles are in place, and the devices turned on, we use the monitoring tools built in the manager to verify the connectivity between the sensors. We place relay nodes in an iterative fashion, each time verifying the updated connectivity information. The rules provided by Dust Networks are for each node to have at least three neighbors with which it has a connection with a Received Signal Strength Indicator (RSSI) higher than -85 dBm. Replay nodes are added until that rule is satisfied. In a typical network, 20 to 30 relay nodes are added to ensure connectivity and path-redundant.

2.6 Back-end system

The back-end system consists of a server running at UC Berkeley, and which is responsible for receiving the data from the deployed ARHO networks, storing it in a database and presenting the data over a web interface on a first-come-first-served basis. Each manager of an ARHO network sends the sensor data and topological information to the back-end server, using the cellular or satellite connection, periodically. Sensor data is also stored locally on each NeoMote's internal memory as backup.

The size of the American River Basin and the number of devices deployed as part of the ARHO project confirm the clear advantages of using low-power wireless technology over a wired solution. The idea of using an off-the-shelf low-power wireless solution has proven a successful

one, although the ARHO project has demonstrated that the networking aspect is only a very small piece of the challenge.

This project has highlighted the importance of network monitoring and management tools. Although the SmartMesh IP manager offers a complete interface to query the state of the network, receive alerts, and verify the performance of the network, what is missing is a multi-network management interface. Such a solution would complement an existing SmartMesh IP network by offering a unified interface to assist during network deployment, visualize the network, run network health routines, display the sensor data, and log maintenance activity. This tool is developed through collaboration with the REALMS associate team (https://realms-team.github.io/).

Chapter 3

The method and evaluation of a basin-scale wireless sensor network for mountain hydrology

3.1 Introduction

In environmental monitoring applications, sensors are often deployed in remote regions. In the ARHO deployment, it takes hours of driving and hiking to get to a deployment site, and installing a single node takes about about an hour. When the snow season starts, the deployment sites are usually inaccessible for 3-4 months. A trial-and-error deployment, in which nodes are added/moved over the course of a couple of days to obtain the right connectivity, is hence not an option. When we install a new node, we need to be able to make an informed guess at how well that node will connect to already installed nodes. A key tool for making that guess is a connectivity model. Given different features of the deployment (the distance between the nodes, the amount of vegetation, etc), that model must produce the best possible prediction of the Receive Signal Strength(RSS) between that node and different other nodes in the network. Note that the term RSS and Receive Signal Strength Indicator (RSSI) are used interchangeably in this chapter.

Such a connectivity model is a basic building block for planning the physical connectivity of a deployment, regardless of the type of networking technology being used (star topology, multi-hop redundant mesh network). Even though, in the ARHO networks, we are using a particular type of low-power mesh network (see Section 3.2), the methodology developed in this chapter is not tied to that networking technology, and applies equally well to star networks. Similarly, even through we focus on low-power wireless networks deployed in forested mountainous areas, the methodology can apply to any deployment area, as long as the data used to train the model resembles the data to predict (see an in-depth discussion in Section 3.6).

Propagation models (equations) are a natural choice for predicting connectivity. Yet, canonical path-loss models are ill-suited in complex terrain due their simplifying assumptions of

plane earth or free space environments. In order to address these limitations, empirical path-loss models that parametrize the effects of vegetation on path-loss in a variety of settings. But although these models attempt to capture the excess power loss as a function of frequency and foliage depth, they are otherwise univariate (i.e. they are only a function of distance).

The ARHO networks produce on average 10 times more network statistics than sensor measurements. Part of these statistics are RSSI measurements: each node reports the RSSI of the link to each of its neighbors, every 15 minutes. Over the course of 1-year, we have collected 42,157,324 measurements, together with the topographic and other features about each of the 2218 wireless links in the deployments. This dataset and gives us a unique opportunity to quantify the performance of the propagation models.

Perhaps more importantly, having this dataset allows to think about connectivity models in a radical new way. In traditional propagation models, one creates an equation which approximates the observed data, and fine-tunes the parameters in a univariate equation (multipliers, exponents, etc.) so it matches the data points best. We proposed to use a completely agnostic "big data" approach by associating to each of RSSI measurements a set of features (distance, vegetation, terrain, etc. between the communicating nodes) and train a multivariate non-parametric model so it learns which features are most important in predicting the RSSI. The result of the this machine learning approach is a predictor: given a new set of features (e.g. a new node is added), is it able to predict the RSSI over that link. The connectivity dataset allows us to quantify the performance of this approach through cross-validation. This allows us to evaluate the accuracy of the machine learning method used, which is based on an ensemble of regression trees (Random Forest). This is a multivariate, non-parametric method in which an ensemble of decision trees are trained on existing data.

The contributions of this article are threefold:

- We present a connectivity dataset consisting of 42,157,324 RSSI measurements gathered on 2218 wireless links in the ARHO networks.
- We develop a machine learning approach to predict link quality by training on this dataset.
- We evaluate *(i)* the suitability of traditional and this Machine Learning-based model for predicting the RSSI in complex environments, *(ii)* whether features other than distance play a role in the prediction of RSSI in complex terrain, and *(iii)* the strengths and limitations of our machine learning-based methodology.

The remainder of this article is organized as follows. Section 3.2 describes the large-scale connectivity dataset. Section 3.3 discusses related connectivity models for forested environments. Section 3.4 presents the machine learning model we propose. Section 3.5 evaluates the performance of the machine learning and canonical models. Section 3.6 summarizes the key findings, discusses the strengths and limitations of the proposed approach and presents the opportunities for further research.

3.2 A large-scale real-world connectivity dataset

The connectivity data we use in this article is gathered from the deployed network discribed in Section 3.

Besides generating sensor measurements, the network continuously produces network statistics for an operator to be able to assess its "health". The network generates approximately 10 times more network statistics than sensor measurements. Each node in the network produce a "Health Report" every 15 min which contains the list of neighbors it is communicating with, and – among other things – the average RSSI of the packets it received from that neighbor over the past 15 min. Over the course of 1 year, we have collected 42,157,324 such RSSI measurements from the 2218 wireless links that make up the ARHO deployments.

Note that we chose to use two different types of omni-directional antennas with different gain level to adapt with the terrain challenges. Pairs of 6-dB gain antennas were used on "flat" areas with less than 10-degree slopes. When the slope between to radios are greater than 10-degree, we switch to pairs of 4-dB antennas. The lower-gain antennas has a wider beam width in the vertical direction, hence the performance is improved on severe slopes.

Every entry in the dataset of RSSI values is annotated with a set of "features" to characterize the topographic and canopy structures between the two nodes which have exchanged the data used to make that RSSI measurements. The location of the deployed nodes are logged with a handheld GPS unit (Magellan Explorist 710). Features at the associated locations are extracted from two digital raster maps. We use a 30 m resolution digital elevation model (DEM) from the National Elevation Dataset (www.nationalmap.gov/elevation.html). The DEM stores elevations above sea level in meters for the latitude and longitude coordinates of each pixel. Each pixel is 30 m wide in the North-South direction. The spacing varies in the east-west direction, depending on latitude. The clusters are separated with distances from 10 to 50 km. Features associated to canopy density are extracted from a percent-tree canopy cover raster developed by the National Land Cover Database (NLCD) with 30 m resolution (www.mrlc.gov/nlcd2011.php). The NLCD map gives a relative canopy density value ranging from 0 to 100, representing the percentage of tree canopy cover for each pixels [48].

The features we annotate each RSSI measurement with are:

1. **Path ground distance** (a number in meters): the distance between the two radios communicating for this RSSI measurement, calculated from their GPS locations and elevations.

2. **Mean percent tree canopy cover** (a number between 0% and 100%): the average pixel value from the NLCD vegetation map along the line-of-sight path between the two communicating nodes.

3. **Terrain complexity**: the standard deviation of the raster values from the DEM along the line-of-sight path between the two communicating nodes.

4. **Vegetation variability**: the standard deviation of the raster values of the NLCD vegetation map along the line-of-sight path between the two communicating nodes.

5. **Path angle**: the angle between the line-of-sight path between the two communicating nodes and horizontal.

6. **Source canopy coverage**: the bi-linear interpolated values of the NLCD vegetation map pixels at the source and receiver locations, a number between 0% and 100%.

7. **Receiver canopy coverage**: same calculation as the "Source canopy coverage" feature, but at the receiver node.

3.3 Related Connectivity Models for Forested Environments

The aim of a propagation model is to predict the expected connectivity. Models based on the physics of the diffusion of electromagnetic waves in an ideal medium (Friis propagation) provide a first-order approximation of the expected connectivity. Second order models, such as "plane earth" account for constructive/destructive interference based on the height of the transmitter and receiver nodes, and the assumption of an ideal flat and empty environment. These propagation models are described in Section 3.3.

In forested environments, signal strength is attenuated by vegetation. Related work has focused on empirically modeling the excess signal strength loss due to canopy, based on field measurements. These models are described in Section 3.3.

Canonical Propagation Models

The simplest path-loss model is "free space" propagation. It assumes unobstructed, line-of-sight decay of an electromagnetic wave, based on the Friis transmission equation, see (3.1) (and equation (1) in [62] "free space").

$$\frac{P_r}{P_t} = \frac{G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{3.1}$$

In (3.1), P_t is power transmitted by the transmitter, P_r is power received by the receiver, G_t and G_r are the gains of the antennas at the transmitter and receiver, respectively, λ is the signal wavelength (m), d is the distance between the transmitter and receiver (m), L is the system loss factor (equal to 1 for free space, but modified in the empirical models detailed Section 3.3).

"Free space" propagation does not capture the effect of ground reflection. The "plane earth" is a second canonical model which takes into account the effect of ground reflection, under the assumption of an infinite ground plane, see (3.2) (or (2a) in [62]).

$$\frac{P_r}{P_t} = 2\frac{G_t G_r}{L}\left(\frac{\lambda}{4\pi d}\right)^2 [1 - cos(k_\omega \frac{2h_t h_r}{d})] \tag{3.2}$$

In (3.2), h_r and h_t are the heights of the transmitter and receiver (m), respectively, k_w is the wave number (m^{-1}).

Empirical Propagation Models in Forested Environments

While canonical plane earth and free space path loss models provide a first-order approximation of path loss, their simplifying assumptions make them unrealistic in the general case. We are looking for a model representative of our deployments in forested environments. Models specific to this environment model the excess loss induced by the foliage characteristics between each link. They modify the loss factor (L) from the canonical models. Well-known propagation models for forested environments include Weissberger's modified exponential decay model [110], the ITU Recommendation (ITU-R) model [54] and the COST235 model [41].

The Weissberger [110] model assumes that propagation occurs through a dense body of dry trees (see (4) in [62]). It assumes that propagation only occurs through the trees, and is not diffracted over the top of the trees. The loss factor is given by (3.3).

$$L_{Weiss} = \begin{cases} 1.33 f^{0.284} d^{0.558} & 14 \text{ m} < \text{d} < 400 \text{ m} \\ 0.45 f^{0.284} d & 0 \text{ m} < \text{d} < 14 \text{ m} \end{cases} \tag{3.3}$$

In (3.3), L_{Weiss} is the loss due to foliage, f is the transmission frequency (GHz), d is the distance between transmitter and receiver (m).

Another common propagation model is ITU-R [54]. Like Weissberger, it assumes that the majority of the signal propagates through a body of trees (See (5) in [62]). The measurements for ITU-R were primarily made in the Ultra High Frequency (UHF) range. The loss factor is given by (3.4).

$$L_{ITU-R} = 0.2 x f^{0.3} d^{0.6} \tag{3.4}$$

In (3.4), f is the transmission frequency (MHz), d is the distance between transmitter and receiver (m).

The Weissberger and ITU-R models do not account for seasonality (i.e. differing amounts of vegetation when trees have leaves or are bare). In order to account for this effect, the COST235 model [41] was developed based on measurements carried out over two seasons when trees are "in-leaf" and "out-of-leaf". Measurements were made in the millimeter-wave frequencies (9.6 to 57.6 GHz). The resulting parametrization is given by (3.5).

$$L_{COST235} = \begin{cases} 26.6 f^{-0.2} d^{0.5} & \text{out-of-leaf} \\ 15.6 f^{-0.009} d^{0.26} & \text{in-leaf} \end{cases} \tag{3.5}$$

In (3.5), f is the transmission frequency (MHz), d is the distance between transmitter and receiver nodes (m).

Several other models have been developed to take into account specific characteristics. In [39], the authors derive path loss as a function of trunk height gain k for a pine tree environment. The authors in [6] model RSSI inside a forest, based on factors such as the average density of trees, or the average trunk diameter. They find that the path loss coefficient decreases linearly with the average tree density multiplied by the trunk diameter. The authors in [20] use high-resolution LiDAR data to calibrate a log-normal path loss-model.

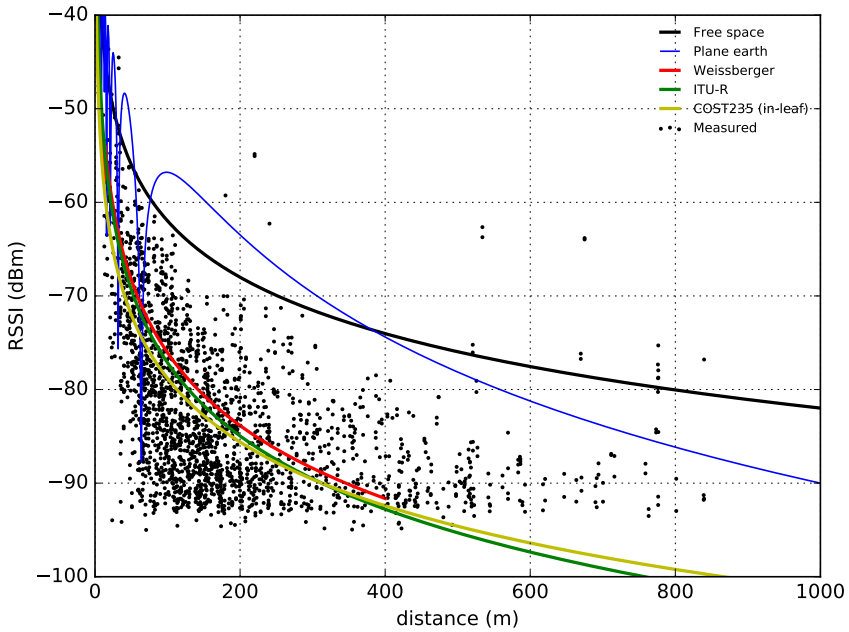


Figure 3.1: Comparing the different canonical (Section 3.3) and empirical propagation models (Section 3.3) against the measurements gathered on the 2218 wireless links of the ARHO networks (Section 3.2).

Comparison to Real Data

All of the canonical and empirical propagation models in Sections 3.3 and 3.3 are equations. Starting from ideal physics in the canonical models, the empirical propagation models fine-tune some parameters of those equations (e.g. the system loss factor L) so the resulting equation better matches some experimentally gathered data. These models use the distance between nodes as the only variable.

Since we have collected a large number of real-world RSSI measurements on the 2218 wireless links from the ARHO networks, we are interested in seeing how well the models surveyed in Section 3.3 are able to match the measurements. We plot in Fig 3.1 the RSSI as a function of distance predicted by the different models, and overlay our measurements. In the models, we set the parameters to match that of our deployment: $G_t = 4$ dBi, $G_r = 4$ dBi, $\lambda = \frac{c}{2.4\ GHz}$ (c being the speed of light), $h_t = h_r = 5$ m. Fig. 3.1 suggests the models do not match our empirical data well, and over-estimate the measured RSSI. Moreover, the shape of the signal strength decay is not exponential: at short distances, there is much greater signal strength variability than predicted by the models.

To quantify the error between the models and the measurements, we use each of the models on each of the 2218 wireless links in the dataset. We record the difference between the RSSI predicted by the models, and the average RSSI of that link over the year of measurement. Fig 3.2

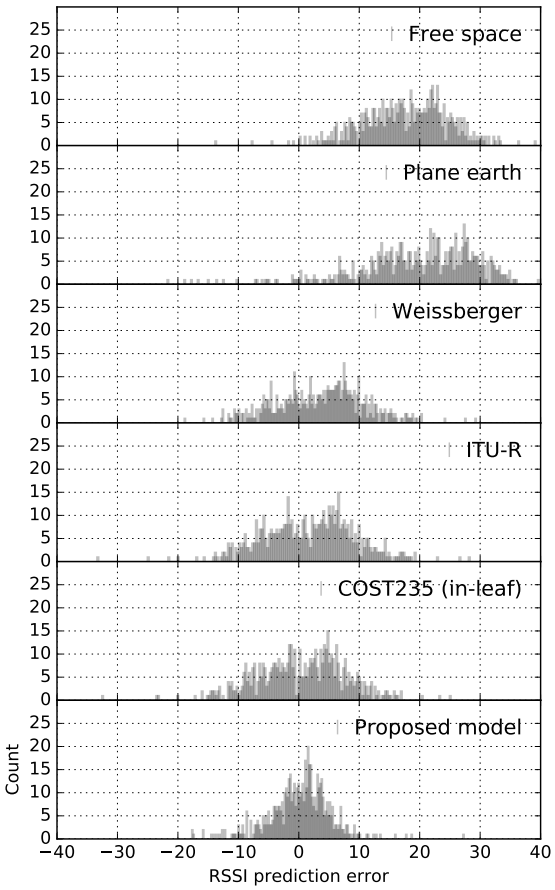


Figure 3.2: Distribution of errors under canonical and empirical models (top panels), compared to proposed model (bottom panel) for year-averaged RSSI data.

shows the results as a histogram. Table 3.1 contains the average and standard deviation of the prediction error for the distributions shown in Fig 3.2.

It is clear from Fig. 3.1 and Table 3.1 that the models cannot be used as-is to accurately predict the RSSI between two nodes deployed in the field. Other phenomena besides distance and vegetation affect signal strength, including specifics about the environment we deploy in (e.g. terrain), or the hardware we use (e.g. transmission power, antenna matching, antenna alignment, radiation pattern). We could produce another model, in which we take into account these phenomena, and create an equation that best matches our data. Yet, we believe that such matching makes the model more specific to a particular set of empirically-gathered data, and less generally usable. We question whether using a simple equation is the right approach.

Table 3.1: Performance evaluation of the propagation models to the ARHO dataset.

Model Name	Avg. Error, dBm	SD Error, dBm
Free space [116]	20.5	8.60
Plane earth [116]	17.8	6.92
Weissberger [110]	6.65	4.70
ITU-R [54]	6.37	4.64
COST235 [41]	5.91	4.37
Proposed model	3.74	3.40

Our conclusion is that trying to model every physical phenomenon in an equation is a non-starter. Minute changes to the environment (e.g. node position [108], antenna alignment, the quality of the antenna connector) can cause the RSSI to vary by over 10 dB. Understanding, measuring and modeling each of these phenomena is unfeasible for any real deployment. Our intuition is that we should instead *learn from the wireless links deployed now and predict the performance of wireless links installed in the future.* Section 3.4 develops this idea, and presents a model based on an Ensemble Regression-Tree algorithm (a common algorithm used in Machine Learning). We show how this model achieves a 37% reduction in the average prediction error compared to the canonical/empirical model with the best performance.

3.4 A Machine Learning Model

Given the limitations of the canonical and empirical models, we want to determine whether a multivariate connectivity model could be trained on the RSSI measurements and associated features described in Section 3.2, and used to predict RSSI at un-instrumented locations. We describe machine learning and ensemble regression-tree in Section 3.4, and the proposed model in Section 3.4.

Overview of Machine Learning

Machine-learning algorithms are trained to identify patterns in historical data. This is very different from the canonical and empirical models surveyed in Section 3.3, which fit a specific function “a priori”. Patterns learned by these algorithms can be non-linear, multivariate, and can be used both for predicting which category a piece of data belongs to (classification), and predicting continuously-valued outputs from a set of inputs (regression). Algorithms in machine learning are broadly divided into two categories: “supervised” and “unsupervised.” Supervised algorithms require observations of the output to learn patterns. Unsupervised algorithms learn patterns in the space of independent variables without observations of the dependent variable (e.g. clustering). The field of machine learning has seen a number of recent applications to low-power wireless networking, including for localization and routing [1]. In this

study, we have observations of the output (the RSSI), and try to predict a continuously valued function. We hence design a "supervised regression" machine learning solution.

A number of algorithms can be used to solve supervised regression problems, including: Support Vector Machines, Neural Networks, Nearest Neighbors, Gaussian Processes, and Decision Trees. Of these, Decision Trees are considered to be one of the best off-the-shelf algorithms because they are not sensitive to independent variable scaling or the inclusion of irrelevant variables. [31] Also, decision trees are not "black-box" models: each split in the decision tree can be inspected once the model is trained. A single decision tree is known overfit data [30]. To address this, Random Forests combine estimates from multiple trees using a random selection of features to arrive to a consensus of the true output [16]. This process prevents the model from over-fitting the data (i.e. fitting the noise rather than the trend). The accuracy of the algorithm is affected by parameters of the estimator such as the maximum tree depth and the size of the ensemble. Decision tree depth controls the maximum depth of the decision tree (i.e. how many splits on the independent variables are made). The size of the ensemble is the number of decision trees the outputs are averaged over. In general, a small ensemble with deep decision trees has a greater tendency to overfit than a shallow ensemble of many decision trees. These parameters must be tuned for the RSSI model, which is discussed in Section 3.4.

Once the model is trained, its indicates which features are more important. This is calculated by computing the out-of-bag sampling error (MSE) during training, then permuting each predictor variable and computing the difference in sampling errors. The process is described in detail in [16].

Model Implementation

We develop the connectivity model using Scikit-Learn version 17.1, an open-source machine learning package implemented in Python (http://scikit-learn.org/). We first divide the average annual RSSI data along the 2218 links into three subsets using randomized sub-sampling ("test train split" in Scikit-Learn). We use standard splitting ratios: training (50%), cross-validation (25%), testing (25%). Training and cross-validation sets are used to determine the optimal parameters for the model (decision tree depth and number of estimators). We use the Random Forest Regressor module and perform a grid-search process over the set of potential parameters for n estimators (the number of estimators) and max depth (the maximum decision tree depth). We use mean squared error as the criterion of split quality, and use bootstrapped samples when building the decision trees. Min samples split is set to 2. Min samples leaf is set to 1.

3.5 Model Validation and Results

To validate the model, we use data not used during the training process. We train the model using the best parameters determined in Section 3.4, and evaluate the model's accuracy on the cross-validation dataset selected at random from the available data. We discuss the accuracy of

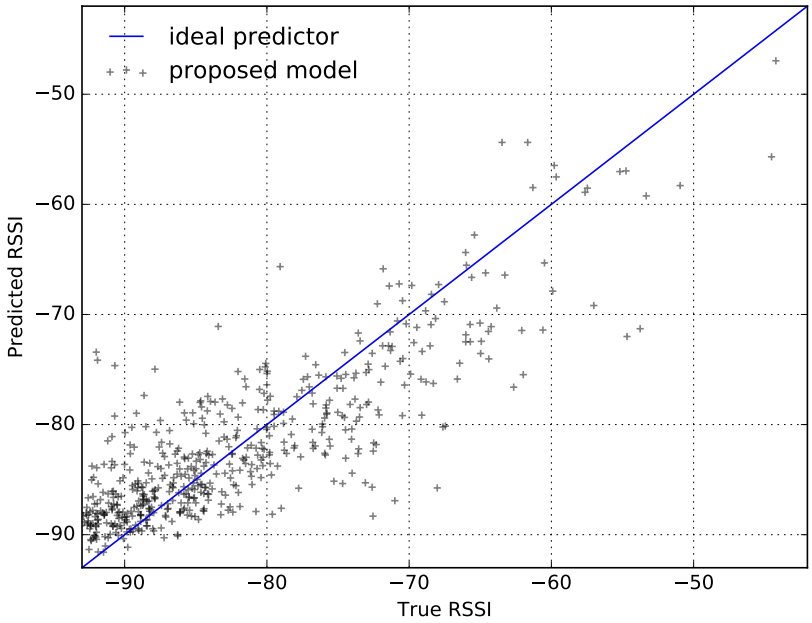


Figure 3.3: RSSI predictor accuracy compared to an ideal predictor (blue line) on the testing dataset of 555 RSSI measurements.

the model on the cross-validation dataset in Section 3.5. We then discuss the relative contribution of each independent variable in Section 3.5.

RSSI Prediction Accuracy

The accuracy of the predictor on the 555-sample cross-validation dataset is depicted in Fig. 3.3. The blue line represents an ideal predictor; black points is the predicted data for the 555-sample cross-validation dataset. The coefficient of determination (R^2) of the predictor is 0.70. The predictor exhibits a slight positive bias at very low values of RSSI (less than -85 dBm) and a slight negative bias at values greater than -85 dBm. Overall, the predictor exhibits near-zero bias (0.18 dBm).

Fig. 3.2 shows how this predictor compares to the predictions done with the canonical and empirical RSSI models from Section 3.3. All of the canonical and empirical propagation models exhibit positive bias on the testing dataset (the bias is 19.5, 17.7, 3.61, 1.75, and 0.49 dBm for plane earth, free space, Weissberger, ITU-R, and COST235, respectively). The models also show a higher mean absolute error and higher error variability (error standard deviation) than the proposed model (Table 3.1). Of the existing models, the COST235 (in-leaf) model shows the highest accuracy. The proposed model, however, exhibits a average prediction error 37% lower than that of the COST235 (in-leaf) model.

Table 3.2: Independent variable importances inferred from the ensemble regression (normalized).

Feature	**Mean importance**	**SD importance**
Path ground distance	0.594	0.026
Mean percent canopy	0.096	0.017
Terrain complexity	0.077	0.015
Vegetation variability	0.076	0.015
Path angle	0.070	0.014
Source canopy	0.043	0.012
Receiver canopy	0.044	0.010

Feature Contribution

One important aspect of this study, which is made possible by the machine learning approach, is to determine which additional independent variable ("feature") is important to predict the RSSI. Random Forests provide a natural ranking of features in the model, based on the degree to which splits on each variable improve the split quality criterion (mean squared error – MSE – in the current study). In Scikit-Learn, this value is averaged across the estimators (i.e. each decision tree) in order to determine a "mean importance" for each feature. Table 3.2 shows the normalized contribution of each independent variable used in the model. Path ground distance and mean percent tree canopy cover are the two most important features in the model, followed by terrain complexity and vegetation variability. Attributes related to the local characteristics of each node (e.g. source/receiver canopy coverage) exhibit lower significance.

3.6 Discussion and Opportunities

This paper introduces a radical new way of thinking about wireless connectivity models. Instead of relying on an *expert understanding* of the physics of wireless propagation, we propose a *agnostic computational* approach in which patterns are identified in recorded data. The machine learning tools used are generic: they are not aware that the data is related to wireless connectivity.

The main result, which is counter-intuitive, is that this agnostic approach yields better results than the expert approach. The proposed approach reduces the average prediction error by 37%, when compared the expert model with the best performance. The quantitative study is based on a large connectivity dataset of 42,157,324 measurements gathered for one year on 2218 wireless links: we have confidence in the results. These findings open up many new possibilities for understanding, planning and diagnosing wireless networks.

Discussion

On top of the numerical results, we want to conclude this article with a discussion about the strengths and limitations of the proposed approach.

Being able to compute which features are important is a key benefit of the methodology developed. First, it confirms the importance of distance as a key feature for predicting the wireless connectivity between devices. But it also highlights that other features are important. To quantify how these additional features help in predicting connectivity, we repeat the analysis with only distance. When we train the predictor with distance alone, the R^2 of the predictor is 0.6, compared to 0.7 when all the features are used. So, even though distance is the most important feature (with an mean importance of 0.594, see Table 3.2), annotating the data with addition features makes a real difference.

An immediate drawback is that these extra features are not common to all deployments. While the mean percent tree canopy cover is an important feature for our networks, it doesn't apply to for example a smart factory application. This means the model created during the learning phase only applies to the particular environment it was created in. That being said, the *methodology* can be used in all environments, and the same remark applies to canonical and empirical models.

Perhaps the main drawback of the approach is that one needs a lot of data to train the model. This has two main implications.

First, you need training data, which leads to a chicken-and-egg situation in new deployment environments. When deploying in a completely new environment, you need to build up a dataset by measuring the connectivity between nodes that were deployed without assistance from a model. As the dataset grows, and as more and more nodes/networks are deployed, the model can be refined to start helping you with the deployment.

Second, you need dataset to be gathered in an environment that has similar distributions of independent variables as the environment used train the model. The 14 low-power wireless networks from the ARHO are similar in that they are deployed outdoor in a mountainous forested areas, and composed of the same devices and radio technology. Fig. 3.4 quantifies this similarity. It shows the standards deviation of the RSSI of the different links over the one-year period of the dataset. It is less than 5 dB for 90% of the of links. The model would be less accurate if the different networks would be deployed in very different areas.

Another limitation of the proposed approach is that the feature selection needs to be done well. Table 3.2 shows the 7 features the training data is annotated with. Even though the machine-learning tool which crunches the data is agnostic to the semantics of the data, it still takes expert knowledge and "intuition" to select which feature to use.

Further complicating the problem, the importance of the feature can evolve depending on the setup. For example, the "path angle" feature has an mean importance of 0.070 (see Table 3.2), which is low. This , in part, can be explained by the choice of using a lower-gain 4 dBi antennas for nodes on the slope. If a 8 dBi antenna were used, the spread of the vertical radiation pattern would be narrower, probably leading to an increased mean importance of the "path angle" feature.

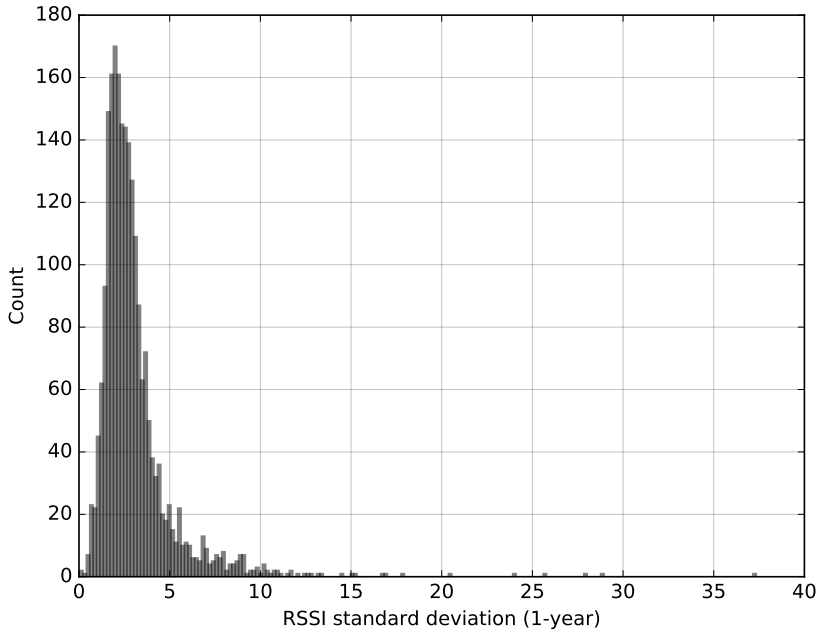


Figure 3.4: Standard deviation of RSSI over all links over a one-year period.

To conclude the discussion, the model developed in this article achieves very good results, with a 37% reduction in the average prediction error compared to the canonical/empirical model with the best performance. That being said, the methodology is applicable only to cases where (1) you already have gathered training connectivity data in similar deployments and (2) you have some intuition about which features are important, and hence which features to annotate the data with.

Opportunities for Further Research

The methodology presented in this article is a radical new way of predicting connectivity in wireless networks. It opens up numerous opportunities of further research, including to understand/minimize the drawbacks highlighted in Section 3.6.

First, develop a methodology to guide feature selection. Such a methodology would guide feature selection in a systematic way, and reduce the amount of expert guidance needed. The goal would be limit the possibility of "missing" important features.

Second, explore other machine-learning methodologies. While Random Forest is considered one the best "out-of-the-box" machine learning tools, other methods may be better suited to the prediction of RSSI. A study could evaluate the accuracy of multiple machine learning algorithms on the same dataset.

Third, evaluate how much training data is required to build a model with good enough ac-

curacy, and how that accuracy evolves as the size of the data set increases. The ultimate goal is to be able to assess how the model behaves if the training data is built-up as more and more nodes are deployed. This study could be done with the same dataset, by considering data from an increasing number of wireless links.

Fourth, build a prediction placement tool. This study would use the model proposed in this article to optimize network topologies in complex terrain over the set of feasible signal repeater placements. This would facilitate the automated deployment of new networks to ensure they are robust to path-loss.

Fifth, apply the same methodology in different environments. This study would gather a similar connectivity dataset in a different environment (e.g. a smart factory), verify that the methodology applies equally, and quantify the difference in connectivity with the model presented in the present article.

Chapter 4

A data-driven machine-learning based connectivity model for complex terrain low-power wireless deployments

4.1 Introduction

Currently, in situ measurements of mountain water cycles at the basin scale are limited in both spatial coverage and temporal resolution, with data largely provided by a relatively small number of operational precipitation, snowpack, climate and stream-gauging stations [8, 24].dv In the Sierra Nevada, measurement sites supporting operational water-resources decision making are also biased to middle and lower elevations and flat terrain in forest clearings [77].

Hydrologic prediction, particularly when constrained by the practical demands of water-resources management, relies heavily on calibrated models to mitigate both limitations in model formulation and inadequate data for rigorous model testing [97, 61]. There are increasing demands on distributed models as predictive tools for situations in which lumped models may fall short, such as non-stationarity in catchment conditions or climate; however, their use in water-resources management is limited by the level of field data available [92]. The need for improved coverage by in situ measurements is both local and global, and new network designs should complement satellite data [117]. Ground-based sensors provide critical ground truth for remotely sensed satellite and aircraft data, and offer a wide suite of independent data that can help provide much-needed gains in predictive modeling. Realizing gains in accuracy from the next generation of spatially explicit models at the scale of water-resources decision making will require both the broad spatial coverage of remotely sensed data and the accuracy of in situ measurements [64]. An adaptive rather than one-size-fits-all approach is needed to realize these gains [29].

WSNs are an efficient and economical solution for distributed sensing. It is often costly and disruptive to create networks of spatially representative wired sensors at the scale desired since it might require kilometers of cables placed either above ground or buried. Reliable wireless so-

lutions are now enabled by reduced production costs of wireless equipment and by advances in networking protocols, effectively combining traditionally non-wireless sensors with a wireless platform. [4, 120, 33].

A few WSN solutions, using different network technologies, were developed specifically for applications in hydrology. These studies have not provided quantifiable assessments of network design, operation and hydrologic results at the river-basin scale. A review of these prior deployments, and a comparison of three existing WSN solutions that have been used, is provided in supporting information[13, 95, 101, 56, 86]

In 2007, a WSN with a few dozen nodes was deployed to a golf course near Almkerk Netherland to monitor soil moisture. The study claimed to be successful. However, the description and discussing over the key component of the system, the wireless network infrastructure, was very brief[95]. In 2009, a 12-station, 4-month deployment in a 20-km^2 catchment in the Swiss Alps measured the spatial variability of meteorological forcing, including temperature and precipitation. The study was only conducted over a short time period with rather sparsely distributed stations [101]. Recently, densely deployed sensor arrays have been scaled to a size comparable to the mountain areas being studied. Ninety-nine sensor loggers, within three 40-180 km^2 basins, were deployed to monitor snow properties in southern Germany for one winter. The system deployed used data loggers rather than a WSN [86]. In another study, 150 wireless nodes with over 600 soil-moisture sensors were installed in a forest catchment at Westebach, Germany to study the spatiotemporal distribution of soil moisture over complex terrain [14]. The study used a variation of ZigBee motes developed by JenNet Ltd. Over 300 sensors hosted by 60 wireless nodes have been deployed at the Southern Sierra Critical Zone Observatory since 2008 to study heterogeneous interactions of water within the snowpack, canopy and soil influence on the water cycle [56]. This installation suffered from network optimization issues that limited locations of the sensor nodes, and hardware problems with the cold.

While sensor networks that are deployed in headwater catchments for short durations offer lessons for local-scale WSNs, they provide limited guidance for WSN design, performance and hydrologic benefits for systems in larger mountain river basins, characterized by steep gradients in temperature, precipitation, rain-versus-snow fraction, growing season, vegetation density and evapotranspiration. The proposed approach to scaling WSN measurements to larger basins involves strategically placing local clusters to capture the variability in hydrologically important basin attributes [111].

The research described in this technical report addresses three questions. First, to what extent can a basin-scale distributed wireless-sensor network with a limited number of sensors arrayed in local clusters sample hydrologic variables across a representative range of landscape attributes in a seasonally snow-covered mountain basin? Second, to what extent can this low-power, distributed wireless-sensor network reliably provides hydrologic data during harsh winter conditions? Third, what types of gains in hydrologic information may result from this network? Further development and more-detailed analysis of the third question is also the subject of Chapter 4.

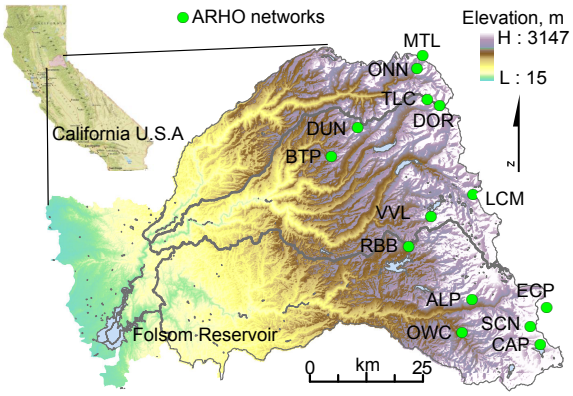


Figure 4.1: Location of American River basin and 14 sensor clusters deployed in the upper part of the basin.

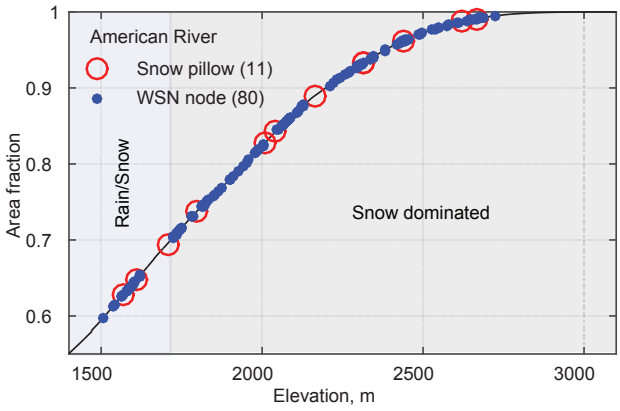


Figure 4.2: WSN nodes on hypsometric curve with existing snow pillows.

4.2 Methods

The network was deployed in the ARHO, in the upper, snow-dominated portion of the American River basin on the western slope of the Sierra Nevada in California (36.069 N, 120.583 W). The basin is incised with steep river canyons and is comprised of three sub-basins: the North, Middle, and South forks, which combine to form a drainage basin of 5311 km^2 above the Folsom Reservoir, the main impoundment on the river. Basin elevations range from 15 m at Folsom to 3147 m at the Sierra crest, with precipitation transitioning from rain to snow at about 1400-1600 m elevation [89, 59]. Forty percent or about 2154 km^2, of the basin is above 1500 m, the lowest elevation for siting our WSNs. About 0.5% of the basin is above the highest node that was sited (2678 m). In 2013-2015, 14 clusters of 140 wireless nodes were deployed (Figure 4.1), with locations selected to represent the range of elevation, aspect, canopy coverage, and solar loading in the basin (Figure 4.1 and 4.3).

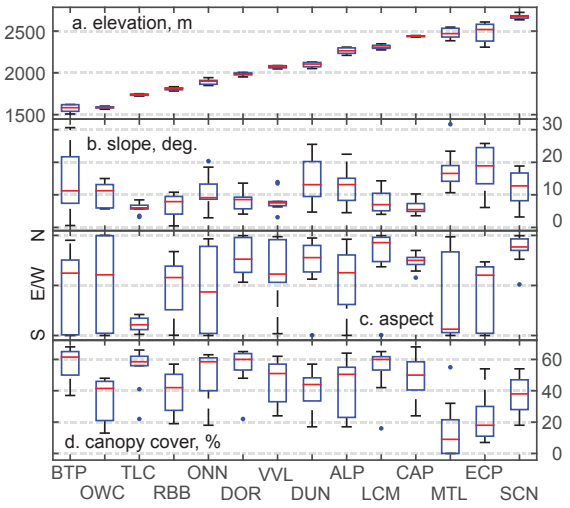


Figure 4.3: Characteristics of individual sites, arranged from lowest to highest elevation. See Table 1 for site abbreviations: a) elevation in m, b) slope (^{o}), c) aspect (S to N) and d) canopy cover (%). The central red mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extended to the extreme points not considered outliers, and the blue dots are the outliers.

4.3 Physiographic attributes of cluster

Elevations were extracted from a 30-m DEM. Slopes and aspects were calculated using ArcGIS spatial analyst toolbox. Percent canopy cover was extracted from NLCD 2011 data layer [55]. Overall, sensor placement reflects a close correspondence between site characteristic of the sensor nodes and the features within the 1-km^2 area for most of the sites. Mean site elevations range from 1590 to 2680 m, with considerable overlap between some sites (Figure 4.3a). Some sites are relatively flat (e.g. CAP) and some on relatively steeper terrain (e.g. MTL, ECP) (Figure 4.3b). It was possible to get a range of aspect at most sites, with the notable exception of TLC and CAP (Fig 4.3c). All other sites have both north and south aspects. Sensor placements capture the range of canopy covers, shown in Figure 4.3d.

The number of local clusters was based on results of Welch et al.[111]and constrained by project budget. The Welch et al. analysis used spatial time-series data over 11 years and a rank-based clustering approach to identify measurement locations that will be most informative for real-time estimation of snow depth, and derived a set of regions that remained relatively stable over time. They found a point of marginal return at about 15 measurement locations, after which placing more local sensor networks did not significantly improve estimation performance. The Welch et al. study also showed that there is some flexibility in placing the local clusters to capture representative parts of the basin, and thus all sites, except MTL and DOR, were co-located with existing snow pillows and met stations. Each cluster consists of ten measurement nodes, limited due to budget, seven to 35 signal-repeater nodes, and a network

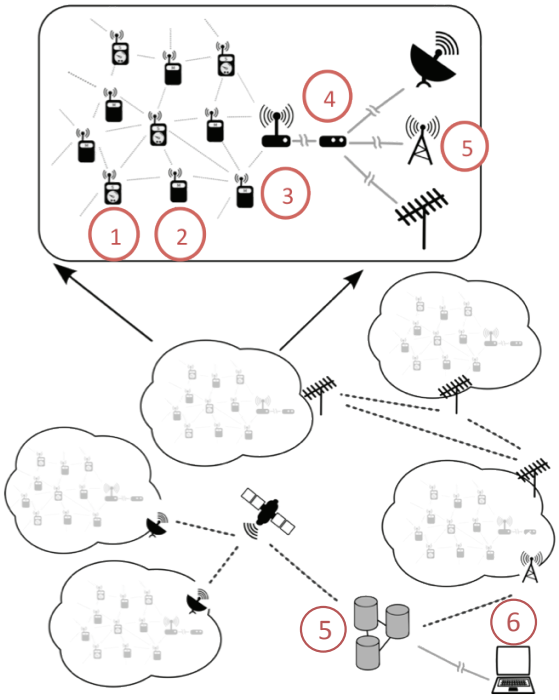


Figure 4.4: American River basin system hierarchy: local WSN clusters connect to central server through data links provide by cell phone or satellite modems. (1) wireless sensor nodes (a neoMote, sensors and external power infrastructure); (2) repeater nodes; (3) Metronome network manager/base station; (4) external connection to the Internet; (5) central-site data server; and (6) real-time visualization engine and data portal.

manager (see Table 4.1 for details; and Figure 4.4 for system hierarchy).

4.4 WSN Systems

Each sensor node (Figure 4.4) is equipped with an ultrasonic snow-depth sensor (Judd Communication Depth Sensor) and a temperature/relative humidity sensor (Sensirion SHT-15). A selected subset of the nodes at five of the sites measure soil moisture and soil temperature (Decagon GS3) at depths of 10 and 60 cm. Nine sites include measurements of total incoming solar radiation using an upward-pointing Hukseflux-LP02 pyranometer at node locations on a separate mast with a concrete foundation. The solar-radiation sensors at these locations are located in the open, without obstruction by either canopy or the terrain to capture the total available incoming solar irradiance. At nine of the 14 sites, co-located with the clear-sky irradiance, solar radiation is measured in a partially canopy-covered location, providing representative solar irradiance-measurements underneath the canopy structure. It should be noted

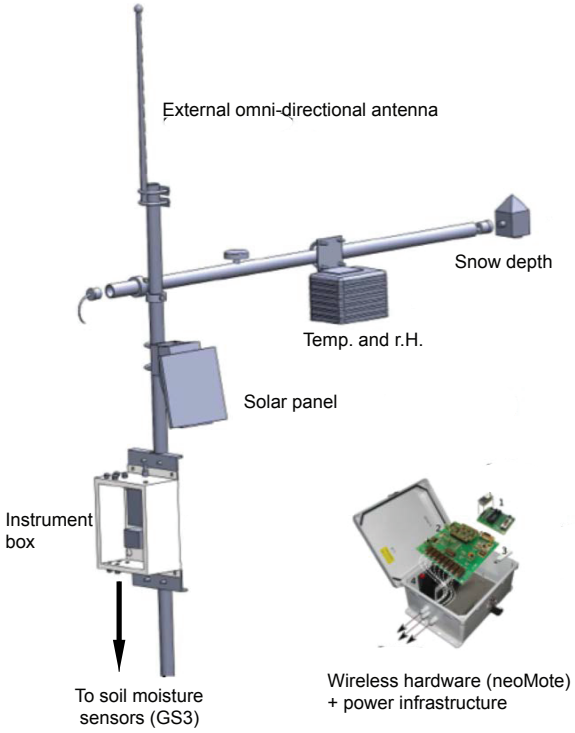


Figure 4.5: Sensor node detail. The 4.5 m vertical mast is bolted to a U-channel driven into the ground. Sensors are either buried under ground or mounted on a 1.2-m long cross arm 4 m above ground.

that our wireless nodes are not limited to these sensors, which were chosen based on past performance, cost and consistency with other networks.

Table 4.1: Equipment installed for sites

Abbrev.	Site name	Lat, Lon	Temperature, relative-humidity	Snow depth	Soil moisture	Solar radiation
SCN	Schneiders	38.745,-120.067	10	10	10	2
ECP	Echo Peak	38.848,-120.079	10	10	0	2
MTL	Mt. Lincoln	39.287,-120.328	10	10	0	1
CAP	Caples Lake	38.711,-120.042	10	10	0	2
LCM	Lost Corner	39.017,-120.216	10	10	0	0
ALP	Alpha	38.804,-120.216	10	10	10	1
DUN	Duncan Peak	39.154,-120.510	11	11	10	2
VVL	Van Vleck	38.944,-120.306	10	10	10	1
DOR	Dolly Rice	39.149,-120.369	10	10	0	0
ONN	Onion Creek	39.274,-120.356	10	10	0	2
RBB	Robb Saddle	38.912,-120.379	10	10	0	2
TLC	Talbot Camp	38.944,-120.306	10	10	0	0
OWC	Owens Camp	38.736,-120.241	10	10	0	0
BTP	Bear Trap	39.095,-120.577	10	10	0	0

Measurement-node placement consisted of three steps. First, major physiographic variables that affect the water balance were characterized in a 1-km^2 area around each site [10, 26, 5, 27, 102, 25, 71, 8]. Second, at each site, ten points representing different physiographic attributes were selected by a random-stratified technique, and the attributes aggregated to assess their representativeness in the larger basin. Rice and Bales [94] showed that a 10-sensor network could capture the mean and distribution of snow depths at this scale. Third, final location adjustments were made in the field to a small subset of sensor nodes, ensuring a complete sampling of the physiographic features together with a strong WSN connection mesh.

The network statistics presented were evaluated over a period of seven months. Each node provided 15-minute data for snow depth, air temperature and relative humidity. Hourly and daily products were developed for periods where no less than 75% of data were present and valid within the averaging window. Extreme values in the data were removed following Daly et al.[19]. Operational data were downloaded from the California Department of Water Resources (http://cdec.water.ca.gov/). SNODAS data, obtained from NSIDC, were used as an additional point of comparison with our snow measurements (http://nsidc.org/data/). SNODAS data were extracted from cells overlapping WSN clusters using a simple weighted-average scheme (Figure 4.6). The results of the extracted values from all 10 nodes are averaged to yield the SNODAS mean for each local cluster. Hourly dew-point temperature for each node was computed based on an empirical equation [63].

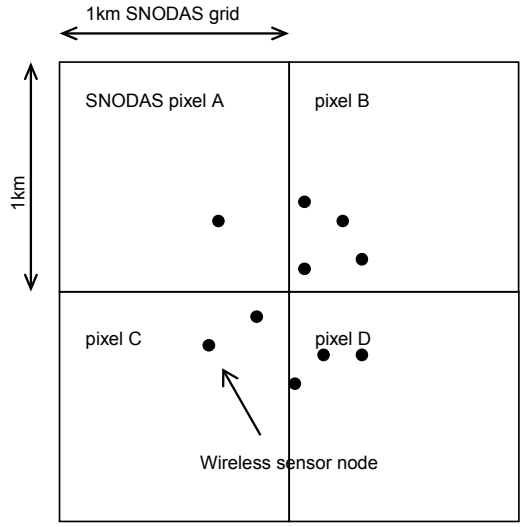


Figure 4.6: Illustration of how a weighted average of SNODAS SWE data were calculated for each local cluster, for comparison with WSN data. For SNODAS data gridded at 1-km spatial resolution, pixels containing nodes for each local cluster were extracted and averaged for that local cluster. In the example shown, the SNODAS mean of this site was calculated as (A+4B+2C+3D)/10.

4.5 Results

WSN performance

The wireless-network links formed a redundant multi-hopped mesh network of sensors and repeaters for data transport. Figure 4.7 shows the stable layout of sensor nodes for the Alpha cluster (ALP), and illustrates how repeaters were non-uniformly distributed to connect the sensor nodes via at least two independent paths to the base station (see Figure 4.7b for photographs of base station, nodes and repeater). A relatively large number of repeaters were installed to provide redundant paths to sensor nodes 6, 8, and 9, where a steep change in slope produced a radio path kink and reliable network links were challenging to establish. Surprisingly, it was found that lower-gain antennas worked better in steep terrain. During 213 days of consecutive recording only 662 out of over 56 million packets were lost in transmission. The average number of hops for packets to transmit from a node to the base station was 3.6 and the maximum seven. The average latency of the network, the time it takes from the packet being sent until it arrived at the base station, was 1.01 second. On average, each node received 181,000 packets over the period when network statistics were gathered.

Two measures indicate the reliability and performance of the network: i) the number of other sensor or repeater nodes connected to each node and ii) the average received signal strength indicator (RSSI). RSSI is closely associated with an important network-performance

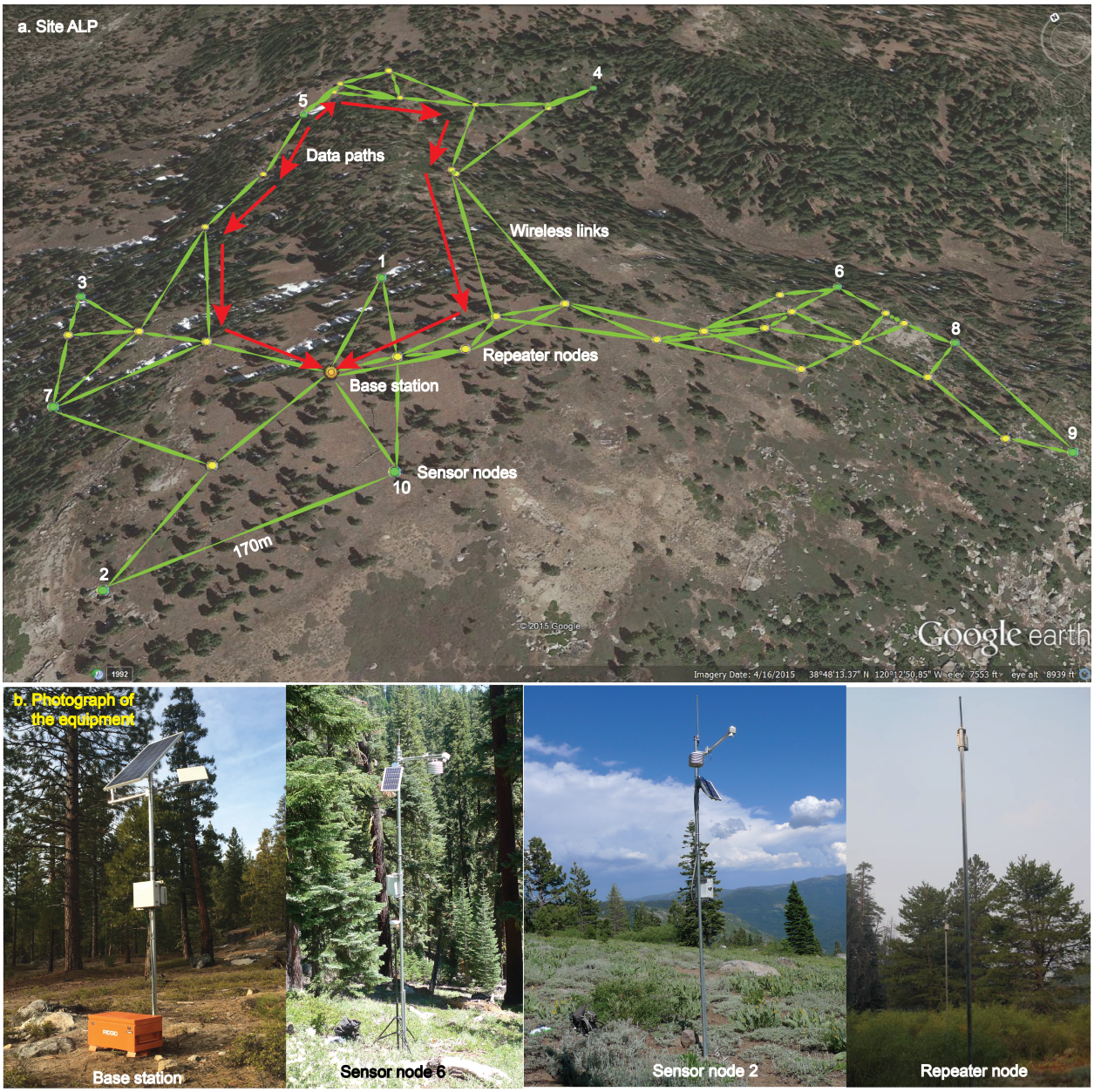


Figure 4.7: Node layout and steady-state network connections (green lines) at ALP, overlain on Google Map. Sensor nodes are numbered. Two possible paths of data out from sensor node 5 to the base station are marked with red arrows.

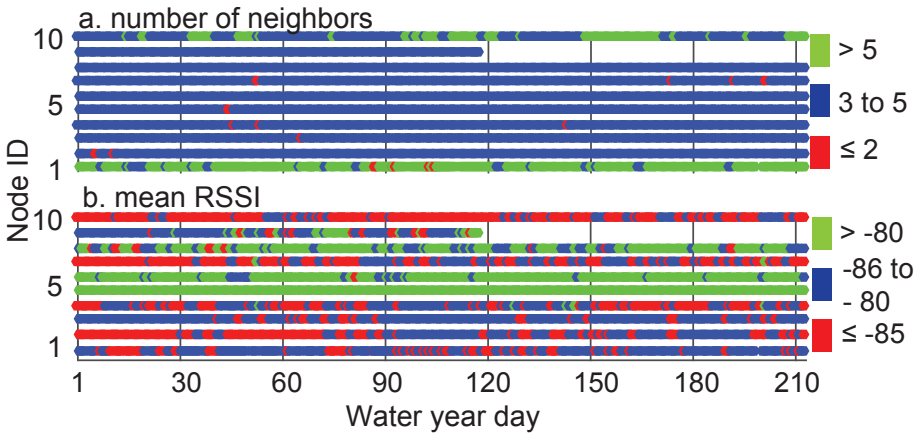


Figure 4.8: Average daily network performance of sensor nodes at ALP for seven-month period. Top panel shows number of network neighbors for each of the 10 sensor nodes, and bottom panel is the average received signal strength indicator (RSSI) for each sensor node. A white gap indicates no communication. The data-stream gap for node 9 in January 2015 was due to a non-network related hardware failure.

indicator called packet delivery ratio (PDR). In aggregate, each node was connected to at least two other nodes over 95% of the time, and to three or more nodes 68% of the time (see Figure 4.5). Taking all nodes together, RSSI values were above -85 dBm, the manufacturer-specified threshold for efficient transmission over 54% of the time, with values above -80 dBm 33% of the time.

There was no clear influence of environmental factors, e.g., temperature, humidity and snow-induced topographic changes, on network performance (Figure 4.9). Each node was connected to one to five other nodes at each time step (Figure 4.9a). RSSI values at each node typically fluctuated +5 dBm, and the average RSSI (Figure 4.9b) depended on node location rather than temperature (Figure 4.9c), humidity (Figure 4.9d) or topographic changes due to snow accumulation (see water-year days 72 and 80, Figure 4.9e).

Temperature, humidity and snow patterns

Daily air and dew-point temperatures from the 10 wireless-sensor clusters that were installed prior to the 2014 water year showed very similar temporal patterns (Figure 4.10a), with average temperature differences reflecting elevation differences between clusters. Temperatures for all pairs of clusters were highly correlated, $r > 0.91$ for air temperature and $r > 0.86$ for dew-point temperature, $p < 0.05$. Daily temperatures were used to derive surface-level lapse rates, which over the eight-month period varied from close to zero to -12^{o}C/km for both air and dew-point temperatures (Figure 4.10b). The respective average lapse rates for the months before snow accumulation (Oct-Dec) were -4.6 and -5.7 ^{o}C/km, increasing to -5.5 ^{o}C/km for air temperature

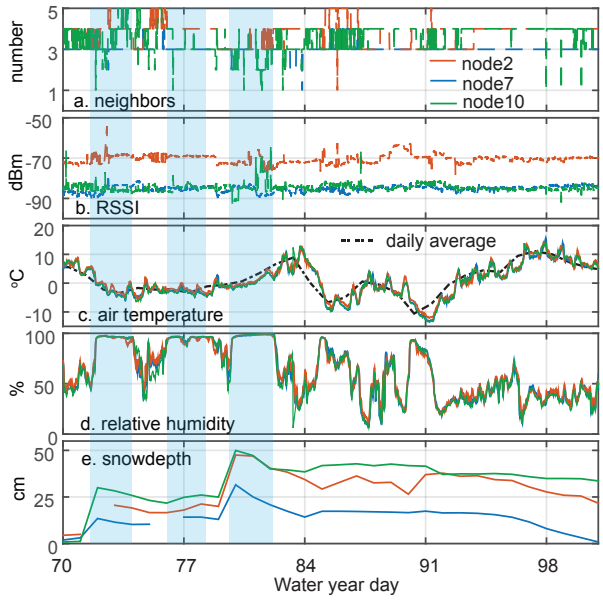


Figure 4.9: Network performance of sensor nodes 2,7 and 10 at ALP: a) hourly data of network neighbors number, b) the corresponding average RSSI, c) average air temperature, d) hourly average humidity, and e) daily average snow depth. Shaded periods represent precipitation events. For clarity, data from three sensor nodes are presented.

and decreasing to -4.7 ^{o}C/km for dew-point temperature during the snow season. The day-to-day variability in lapse rates during the snow-covered period was also lower than earlier in the water year. The transition to a period with less variability in lapse rate is also illustrated by the higher R^2 values starting on water-year day 121, when snow started accumulating in the basin (Fig 4.10c). Note that less-negative air-temperature lapse rates, associated with lower R^2 values, were associated with temperature inversions.

Daily mean air and dew-point temperatures taken across the ten clusters were adjusted to 2100 m using the mean daily lapse rates (Figure 4.10d). The average standard deviation is 3.3 ^{o}C for air temperature and 3.5 ^{o}C for dew-point temperature, a variability equivalent to the average difference over about 600 m and 545 m elevation based on the eight-month average lapse rate of -5.5 ^{o}C/km and -5.0 ^{o}C/km, respectively. While any index elevation could be used for this comparison, 2100 m is generally representative of the upper part of the rain-snow-transition elevation zone.

Mean relative humidity across WSN clusters varied from 15 to 100%, with similar patterns across all 10 clusters (Figure 4.10e). The relations were strong, $R^2 = 0.83$, $p < 0.05$, for all pairs of clusters. While differences in average relative humidity values between clusters were small, absolute humidity and vapor-pressure deficit values were larger. The mean water vapor-pressure deficit for each cluster ranged from zero to 1.5 kPa (Figure 4.10f), with daily inter-cluster difference between the lowest and highest values as much as 55%. The highest variability in vapor-

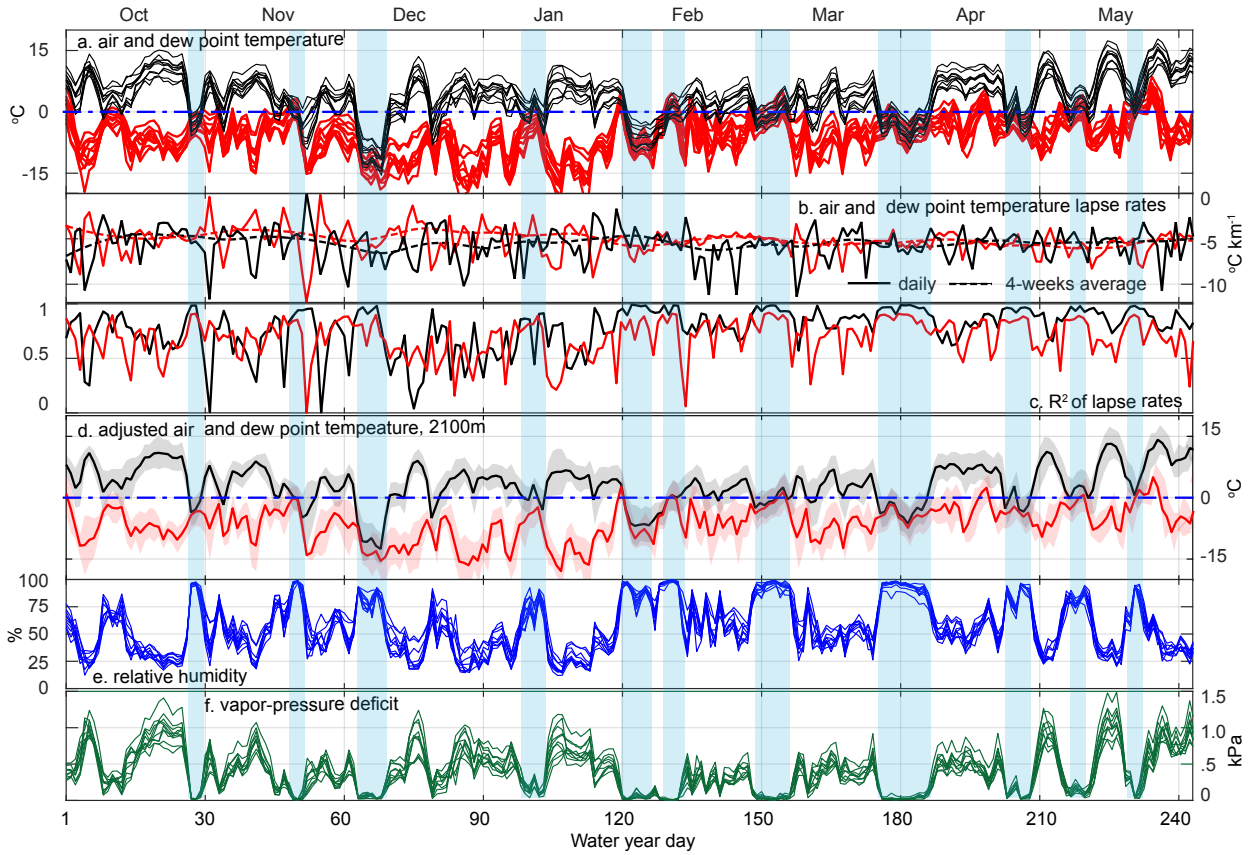


Figure 4.10: Network performance of sensor nodes 2,7 and 10 at ALP: a) hourly data of network neighbors number, b) the corresponding average RSSI, c) average air temperature, d) hourly average humidity, and e) daily average snow depth. Shaded periods represent precipitation events. For clarity, data from three sensor nodes are presented.

pressure deficit was associated with periods of higher temperature and lower relative humidity, indicating a warmer and drier condition. Periods with lower variability of inter-site vapor-pressure deficit were closely associated with sub-zero temperatures in the basin, typically triggered by precipitation events.

Snow-depth data (Figure 4.11) show a clear elevation trend, with variability also increasing with elevation. One exception was SCN, which has a tighter grouping of measured snow depths as compared to lower-elevation sites. During the very warm and dry WY-2014 snow season, sustained snow cover accumulated mainly at elevations above 2100 m.

Snow depths were also compared with co-located or nearby snow-course measurements (Figure 4.11). At lower-elevation clusters, due to the timing of the snow-course measurements, most surveys missed the snow-cover peak accumulation. At ONN, snow-course data showed a small amount of snow throughout the season, missing the few individual peaks. Snow-course values at ECP were generally lower than the mean cluster value across the season.

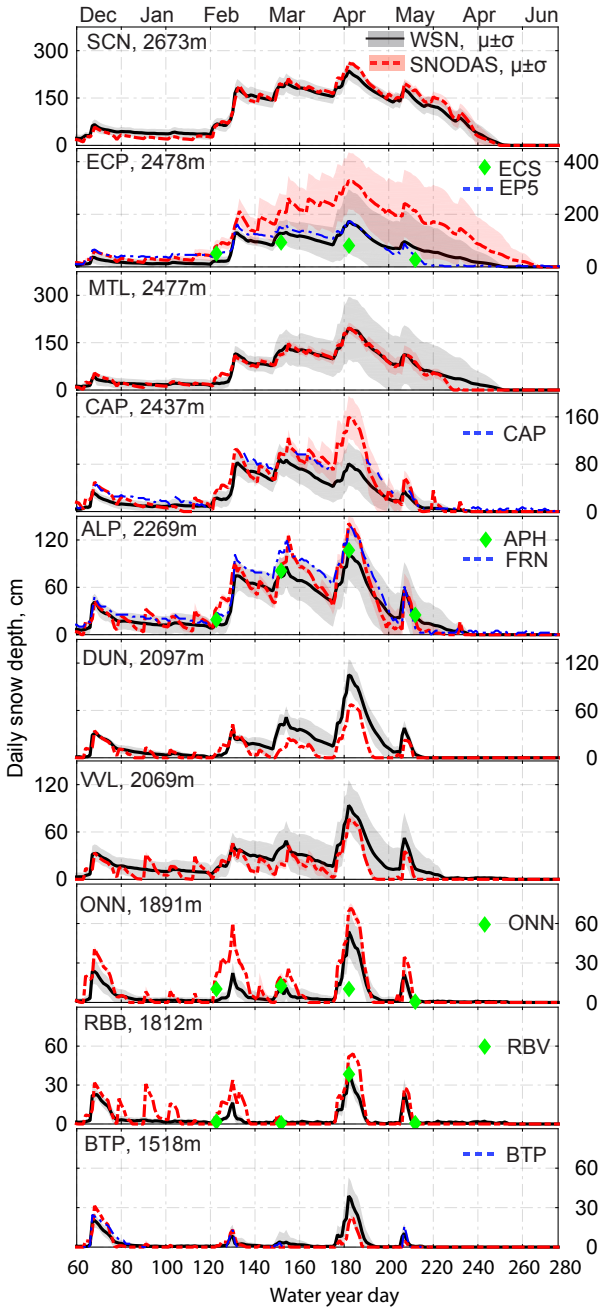


Figure 4.11: Daily mean and standard deviation of snow depth from the WSN clusters and SNODAS, plus available operational snow-depth-sensor data (blue dashed), and snow courses (green diamond).

There are substantial differences between the WSN, nearby operational snow-depth sensors, and SNODAS snow depth at most clusters. Compared to WSN means, nearby operational sensors tended to overestimate snow depth during early season (e.g. at ECP, CAP, and ALP), and matched the WSN mean better at peak accumulation. Near-by operational sensors also showed faster melt than indicated by cluster means for the same sites. The time series of SNODAS values is comparable to the WSN data at MTL and SCN for much of the season, with similar magnitude and high correlation. SNODAS data generally fall within one standard deviation of WSN nodes at these sites. At lower-elevation sites, such as BTP, VAN and DUN, SNODAS underestimated snow depth at peak accumulation by as much as 50% compared to the WSN. At all other sites, SNODAS overestimated peak-accumulation snow depth by as much as 80% compared to the WSN mean.

4.6 Discussion

WSN design and performance

With 945 sensors across 14 clusters, the WSN offers representative, real-time monitoring of the meteorological and hydrologic conditions of the basin. Our analysis of the variability of temperature, humidity, snow and derived quantities shows the importance of multiple landscape attributes in determining their variability. This WSN is also arguably the largest long-term, remote wireless-sensor platform deployed for environmental monitoring. Even though some aspects of the networks in ARHO share similar properties with the prototype installation at the Southern California Critical Zone Observatory [56], the more-recent network statistics help to resolve several previously unanswered questions. The longer-term performance of the networks, subjected to the test of a full snow season, showed that WSNs can be a viable solution for distributed sensing in the at this scale. ARHO networks showed resilience to factors such as humidity and snow-induced topographic changes across different part of the basin. The positive result is likely due to the combination of the Dust Network's radio technologies such as time-synchronized channel-hopping, time-slotted mesh protocol (see section 2.3 for details of the technology), effective network topology, and the use of lower-gain antennas.

A stringent criterion of design was low power consumption, requiring the sensor node to be powered with a 6-amp-hour battery recharged by a 10-watt solar panel. The low-power requirement constrains radio-power output, so the range of the radio limits the size and performance of the network. The WSN uses a multi-hopped network to overcome the range limitations. However, a network that requires more hops to deliver the same number of packets ultimately consumes more power than higher-power, more-energy-demanding radios. Placing the network manager close to the geographic center of the network limited the number of hops. The overall reliability of the network is protected from isolated failures of nodes by radio-link redundancy, provided by the network's mesh topology. This property allows nodes to be sited indiscriminately to potential natural hazards, e.g. treefall.

Topographic relief is one of the more serious challenges to overcome for good system perfor-

mance. Different from earlier installations, the networks in ARHO encountered more-challenging, steep forested terrain. A lower-gain 4-dBm omni-directional antenna provided improved network connectivity, especially in steep terrain, compared to the 12-dBm antennas used by Kerkez et al.[56] on more-even terrain. Even with the improvement, the capability of the network to communicate over steep slopes is limited by the antenna. The ALP site is a good example, where some radio links operated at the edge of the acceptable RSSI level. The network performance was stable but less efficient, indicated by the lower PDR values, compared to Kerkez et al.[56], who had shorter data hops.

Spatial pattern and variability of hydrologic attributes

The following three examples illustrate how our spatially distributed, daily data over complex terrain set provides better estimates of important hydrologic attributes. A more-detailed analysis is the subject of a subsequent chapter 5.

Dew-point temperature

A widely accepted model of near-surface air temperature in mountains is the ground-level lapse rate [22, 96, 51, 58]. Scientists and modelers use lapse-rate-derived temperature to evaluate model responses due to temperature perturbations [32, 7]. In those applications the lapse rate, often averaged over a monthly to annual period, is used to approximate input temperature for models with a much shorter (daily) time increment. This approach, however, does not account for short-term variability. WSN data show that the day-to-day lapse rate was highly variable, particularly before snow accumulation (Figure 4.10b). Not only does the array of sensors provide a more temporally resolved lapse-rate estimate, we also found that redundancy of instruments provides a more-robust estimate of the quantity. Linear models of daily air temperature were constructed with a training set and a cross-validation set of 60 randomly selected nodes. The results were compared with models computed using seven nearby met stations. On average, the cross-validation root-mean-square error was reduced from 1.41 to 1.18 ^{o}C using random sets of 60 measurements versus data from seven nearby met stations. The uncertainty in air temperature was reduced by 16%.

Dew-point temperature complements air temperature in providing a reliable estimate of the timing and phase of precipitation. The reduction of uncertainty in temperature and humidity patterns helps to better determine the elevation range of the rain/snow transition. Air temperature is approximately equal to dew-point temperature, indicating saturated air, when precipitation occurs (Figure 4.10). The phase change from rain to snow usually occurs around the 0 ^{o}C dew-point [72]. Compared to air-temperature-based methods, dew-point temperature is a less geographically dependent variable to determine the solid or liquid precipitation [119]. Due to lack of relative-humidity measurements for most met stations, calculation of dew-point temperature cannot be performed from met-station data alone.

Evaporative potential

Direct measures of vapor-pressure-deficit patterns from a dense array of ground-based sensors can be important for scaling evapotranspiration and assessing forest health [81, 82, 15]. Accurately estimating vapor-pressure deficit is crucial as the saturation-pressure deficit becomes relatively more important in the Penman-Monteith equation [121]. Despite the importance of the variable, reliable field-based estimates of vapor-pressure deficit in mountains are rare. The performance of satellite-based estimates varies, with RMSE values from upwards of 0.3 kPa to 1.1 kPa, limiting their accuracy as estimates of vapor-pressure deficit across steep terrain [88, 43]. A WSN with relative-humidity measurement at every sensor node fills this gap.

Snow depth

The differences in snow depth between WSN and near-by operational sensors can be explained by the patterns of snow accumulation. Operational snow-depth sensors are typically placed near flat meadows or ridge tops free of overhead obstructions or hazards, which produce known biases [78, 3, 94]. We placed our nodes in both forested and non-forested area to produce a more spatially representative measurement. Figure 4.11 indicates that operational snow-depth sensors data had a systematic positive bias in snow depth in the early season. During the melting season, the canopy acts as a shield, limiting energy input to the snowpack [73, 100, 87]. The canopy also shelters the snow surface from wind, reducing turbulent heat transfer. The net result is an extended melt season recorded by sensor nodes in the forested area compared to the operational snow-depth sensors.

Due to local redundancy of the WSN, the data stream is more complete than operational snow-depth sensors at CAP and BTP. Large sections of data were missing from the operational-snow-depth sensors from those two sites during the storm around water-year day 180 (Figure 4.11). This reflects a reality of operational water-resources networks, namely the ability to respond in a timely manner to problems in remote sensors. The redundancy provided by our WSNs helps to address this constraint.

The differences in snow depth between SNODAS and the WSN were less systematic, as there is no apparent trend in the bias across different sites. One pronounced difference between WSN and SNODAS snow depth was at the steep ECP site, where the 1-km^2 SNODAS product overestimated snow depth (Figure 4.11). This follows previous reports that without sufficient data, estimates of snow depth under these conditions can be difficult and error prone due to the underlying variance in elevation within grid boundaries [45]. Clow et al.[18] showed that while over forested regions of the Colorado Rockies, SNODAS estimates of snow depth accounted for as much as 72% of the variance line (1-km resolution) in forested areas, SNODAS was able to account for only 16% of snow-depth variance in areas above the tree.

4.7 Conclusion

A wireless-sensor network distributed over the 2154 km^2 snow-dominated portion of a mountain basin provided good coverage of watershed attributes. With ten measurement nodes per each of fourteen clusters, the WSNs reliably provided spatially distributed measurements of temperature, relative humidity and snow depth every 15 minutes over the basin. The WSN also provided measurements of the significant within-cluster spatial variability of these attributes, which were influenced by local topography, primarily through cold-air drainage.

Compared to existing operational sensors, the wireless-sensor network reduces uncertainty in water-balance measurements in at least three distinct ways. Redundant measurements in temperature improved the robustness of temperature lapse-rate estimation, reducing cross-validation error compared to that of using met-station data alone. Second, distributed measurements capture local variability and constrain uncertainty, compared to point measures, in attributes important for hydrologic modeling, such as air and dew-point temperature and snow precipitation. Third, the distributed relative-humidity measurements offer a unique capability to monitor upper-basin patterns in dew-point temperature and better characterize precipitation phase and the elevation of the rain/snow transition.

Chapter 5

Insights into mountain precipitation and snowpack from a basin-scale wireless-sensor network

5.1 Introduction

At the basin scale, measurements of mountain water cycles currently are limited in both spatial coverage and temporal resolution, with data largely provided by a few operational precipitation, snowpack, climate and stream-gauging stations [8, 24]. In the Sierra Nevada, measurement sites tend to be limited to middle and lower elevations and flat terrain in forest clearings[77]. Research networks include a few selected headwater basins where a more-complete set of meteorological and hydrologic attributes are accurately measured[56]. While these catchments offer some detailed information on mountain hydrology, they provide a limited understanding of the hydrology of larger mountain river basins that can be characterized by steep gradients in temperature, precipitation, and rain versus snow fraction. In mountainous environments, the interaction between soil, vegetation, and existing snowpack and precipitation depends strongly on the precipitation phase[60]. One of the common ways to determine the phase of precipitation is through a calibrated model of air temperature. However, the precipitation phase and air temperature relationship is higher variable across different seasons, sites, and storm tracks, while dew-point temperature shows a strong relationship to precipitation phase[72].

Operational forecasts of runoff are sensitive to estimation of rain versus snow, as illustrated by a tripling of storm runoff for a 600-m change in the estimated melting-level elevation in one simulated 24-hour precipitation event in the American River basin[114]. Freezing levels are of particular interest for warm events, and it has been observed that forecast biases associated with freezing levels above 2300 m in the basin have been under-forecasted by as much as 900 m[115]. It has been noted that the rain-snow transition region is of great scientific as well as practical interest, affecting both current forecasting and potential improvements in predictive tools[91].

Mesoscale differences between the atmospheric 0^o C elevation and the mountainside snow line in the Sierra Nevada make both characterizing and predicting the snow line particularly challenging. Data from three years of storms show that the mesoscale lowering of the snow line is a feature common to nearly all major storms, with an average snow-line drop of 170 m[76]. While radar can be an effective tool for detecting rain versus snow, and other attributes of a precipitation event, its use is limited in the complex topography of the upper American River basin[74].

Recently developed and deployed WSN clusters in the American River basin provide a rich dataset with denser spatial sampling as compared to operational sites[122]. This richer dataset allows us to understand and characterize the critical gradients in temperature, humidity, and precipitation that help define the dynamics of mountain water balance. It can enable use of new classes of spatially explicit hydrologic-modeling tools to produce quantitative assessments, influence hydrologic forecasting, probe system response to climate and land-cover perturbations, increase process understanding of basin-scale water cycles, and provide defensible scenarios for infrastructure planning over a scale currently not possible. These WSN clusters complement deployments by others monitor extreme weather events for flood forecasting[113].

The specific aims of the research reported in this paper were to assess how spatially distributed sensor-network data can improve estimates of: i) the basin's average rain-snow transition elevation during precipitation events, ii) the amount of rain versus snow during mixed-precipitation events across the basin, and iii) the amount and timing of snowmelt across the basin. Together, these affect the soil-moisture and runoff patterns across a basin.

5.2 Methods

Data from 140 wireless-sensor nodes clustered in 10 strategically chosen locations in the American River basin were used together with operational data to estimate precipitation as rain versus snow and snow ablation across the upper part of the basin for the 2014 water year. The sensor-network nodes provided temperature, relative humidity and snow-depth data across the range of topographic and vegetation characteristics in the portion of the basin where snow is currently an important part of the water balance 3.

Study area

The study area was the ARHO, a spatially distributed water-balance sensor in the upper, snow-dominated part of the American River basin on the western slope of the Sierra Nevada in California (36.069 N, -120.583 W). The basin is incised with steep river canyons and is comprised of three sub-basins: the North, Middle, and South forks, which combine to form a drainage basin of 5311 km^2. Basin elevations range from 200 m at Folsom Reservoir to 3100 m at the Sierra crest, with precipitation transitioning from rain to snow at about 1400-1600 m[89, 59]. Sixty percent or about 2154 km^2 of the basin is above 1500 m: the location of the WSNs. The basin supports diverse vegetation types ranging from grasslands, oak woodland, chaparral, and oak savannas

at the lower elevations, mixed conifers and montane hardwoods at the mid to upper elevations, and above the montane forest is the sub-alpine, alpine meadows, and shrub land[105]. The canopy structure exhibits high heterogeneity in both percent coverage and vegetation type, as indicated by National Land Cover Database[55]. The forest landscape is subjected to land-cover perturbations such as forest thinning and fire. Locations for both the ten sensor network clusters, each distributed over approximately 1.5-km^2 area, and sensor nodes within each cluster, were placed to physiographically represent the variability of the upper basin (Figure 4.1a). See Table S5.4 for locations and other information on the ten clusters of nodes used.

Data

Data were from two sources. First, time-series data were taken from 80 sensor nodes of the sensor-network clusters for the first eight months (Oct-May) of the 2014 water year (WY). Each node provided 15-minute data for snow depth, temperature and relative humidity. Hourly and daily products for each attribute were developed from the 15-minute data. Second, precipitation, temperature, snow depth and snow water equivalent (SWE) from across the basin were acquired from operational stations. All data were subjected to quality control to remove noise, following the protocols described in Daly et al[19]. Hourly and daily products were developed for periods where no less than 75% of data, were valid within the averaging window.

In order to estimate SWE from sensor-network measurements of snow depth, we used a basin-averaged snow density derived from ten snow-telemetry sites where snow depth and SWE were both measured. A density time series was developed using the ratio of daily SWE and snow depth (See Figure 5.9). Density values from all sites were averaged, and this mean-density time series used for SWE calculations at all sensor-network nodes, with mean values ranging from about 130 kg m^{-3} in Jan to 420 kg m^{-3} in May. There was no apparent elevation pattern to the density record. Density data at the beginning and the end of the season when the snow pillow may not have been completely covered were omitted. Snow density during this period was assumed to be 330 kg m^{-3}, the seasonal average. Near the end of the season, the last valid snow-density value from each site was extended to calculate basin-average snow density.

Precipitation phase and rain-snow transition elevation

The total amount of precipitation for each of the 10 main precipitation events was estimated for each sensor-network node using the changes in accumulated snow and dew-point temperature. Solid precipitation at each node was computed as the sum of the daily increase in SWE over the duration of the storm event. Dew-point temperatures were calculated at each node based on an empirical formula[63], using air temperature and relative humidity. The duration of each precipitation event was determined based on the mean difference of hourly air and dew-point temperatures. We assumed that dew-point and air temperatures were approximately equal during a precipitation event. For each day with solid precipitation, the mean daily dew-point temperatures were used to determine the precipitation phase. The proportion of liquid and solid precipitation was computed from the daily dew-point temperature. With a

dual-threshold temperature of +1^{o}C, the proportion of snow and rain was allowed to vary linearly between the thresholds[90]. We thus defined the rain/snow transition zone as the band of elevations between +1^{o}C dew-point temperature. Precipitation was considered as 100% solid and 100% liquid if computed mean dew-point temperature were below and above the threshold temperatures. Results were evaluated using measurements of SWE from snow pillows and precipitation from rain gauges at two sites (ALP and ECP).

Basin mean SWE and melt

The amount of accumulated snow and melt-out dates provide an additional check on the partitioning of rain versus snow precipitation. Spatially averaged sensor-network SWE estimates were compared to operational snow data to assess differences in melt patterns and melt-out dates across elevation. The portion of the basin above 1500 m was divided to three elevation zones of equal area, with cutoff elevations at 1752 m and 2041 m. The basin's area/elevation relationship was derived from a 30-m DEM. SWE estimates from sensor nodes in each zone were averaged to zone and basin means. The snow-disappearance day was determined as the day snow depth fell below 1 cm.

5.3 Results

Temperature, humidity and precipitation

Data from the 80 measurement nodes illustrate average patterns across elevation as well as within-cluster heterogeneity. For example, over a typical 14-day period, the daily cycle of air temperature for the 10 nodes at ALP (Figure 5.1a) showed a 5-10^{o}C difference between hourly maximum and minimum values on a given day, with temperatures below 0^{o}C during snow accumulation (Figure 5.1c). The smallest snow accumulation during the event on water-year day (WYD) 207 was 22 cm for a heavily forested location, with two other nodes (in the forest clearing) receiving 31 cm of snow. Relative humidity peaked at 100% from WYD 206 to 208 (Figure 5.1b), and showed little variability across the site. Most of the snow disappeared within three days after the event due to warm temperatures.

Across the basin, eight of the 10 clusters of nodes were co-located with a met station (Table 5.4), allowing comparison between the sensor-network spatial mean versus single met-station temperature. At four of these, the met-station average temperature was within 1^{o}C of the 10-node average. Over an 8 month-period, at four of the clusters (BTP, VVL, CAP, ECP) the average daily sensor-network temperature was 1.5, 1.1, 1.1 and 1.8 ^{o}C, respectively, below that for the nearby met station (see Figure 5.2). For these sites, 80%, 2%, 58% and 77%, respectively, of days had a difference greater than 1^{o}C. The intersections of air and dew-point temperature coincided well with precipitation events at all sites. Four of the ten sites also have a met station with a rain gauge to measure precipitation, plotted as daily values on Figure 5.2. Note that precipitation was recorded for most, but not all events at these sites.

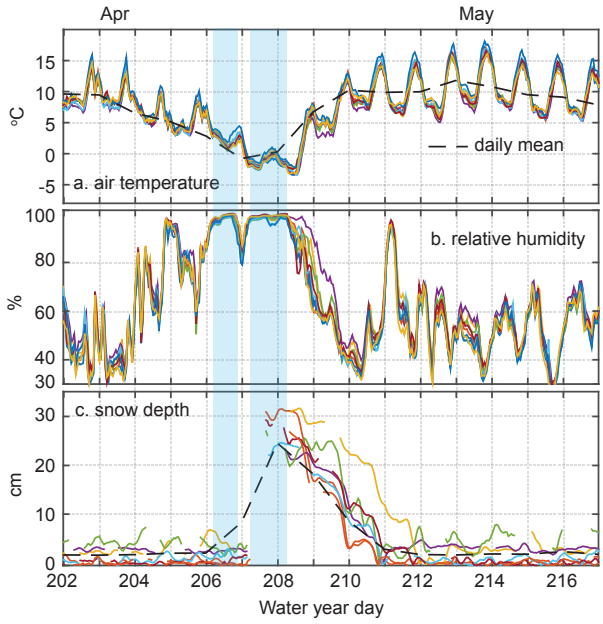


Figure 5.1: Measurements from the 10 nodes at ALP over 2 weeks: a) hourly and daily mean air temperature b) relative humidity, and c) snow depth. Shaded bands indicate rain (WYD 206-207) or snow (WYD 207-208) events.

Rain-snow transition zone

The elevations of the rain-snow transition zone were apparent in nine out of 10 precipitation events (Table 5.1). The coldest period was WYD 67 (Event 3), with an average daily dew-point temperature of -13.5 ^{o}C, and maximum of -9.1 ^{o}C. During this event, the rain-snow transition was at approximately 600-m elevation, well below all of our sensors. The rest of the events all had 0 ^{o}C dew-point temperature elevations between the highest and lowest of our sensors.

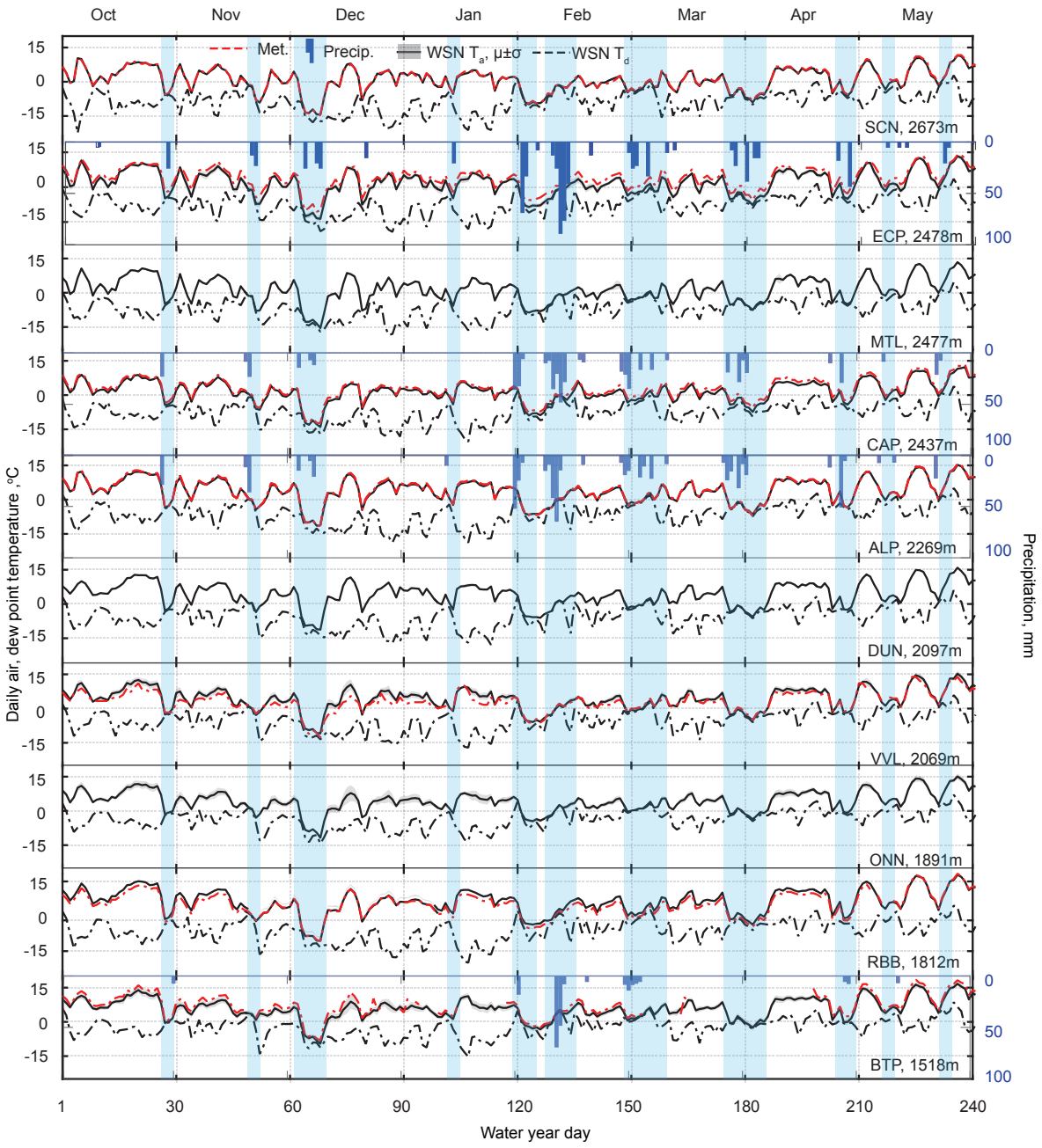


Figure 5.2: Daily mean air temperature (T_a), and dew-point (T_d) from sensor nodes at each WSN cluster, along with temperature and precipitation data from nearby operational met stations. Shaded areas represent periods of precipitation.

Table 5.1: Rain/snow transition characteristics of the ten precipitation events[a]

Event	Duration, WYD	Mean elev. of,1 ^{o}C T_d, m	Δ elevation to -1 ^{o}C T_d, m	Mean T_d lapse rate, ^{o}C km^{-1}
E1	27-29	1463	328	-5.4
E2	49-52	1721	328	-4.7
E3	67-69	627	217	-6.5
E4	102-103	1691	336	-4.7
E5	121-122	1720	242	-4.8
E6	128-132	1681	343	-5.0
E7	148-157	1889	298	-5.7
E8	176-183	1485	298	-5.9
E9	206-208	1509	301	-5.7
E10	231-232	2011	330	-5.1
Mean	-	1580	302	-5.0
St.Dev	-	376	42	0.6

[a]Values are averages of hourly data over the entire event

The event E6 is used as an illustration of our analysis. The average lower boundary of the rain-snow transition zone for the five-day period of E6 was 1681 m (Figure 5.3a). During this event, the rain/snow transition zone gradually moved up in elevation. The hourly dew-point lapse rates during precipitation were relatively stable (mean -5.0 ^{o}C km^{-}1) compared to pre- and post-event values, with R2 values significantly higher and RMSE lower during precipitation (Figure 5.3b-d). Early in the event, snow depths at all clusters tracked each other, with a divergence of trends observed on WYD 131, when only half of the clusters at higher elevation recorded increases in snow depth (Figure 5.3e). At FRN (2269 m elevation), the precipitation gauge recorded 5 cm on WYD 130, with no SWE increase recorded on the snow pillow. On the other hand, ECP, roughly 200 m higher in elevation, recorded a 15-cm increase in SWE and 13-cm increase in precipitation during the same day (Figure 5.3f). Both sites received solid precipitation on WYD 132 as the storm intensified (Figure 5.3f).

Upslope migration of the rain/snow transition during event E6 was also visible in the daily increase of snow-depth measurements across the basin. The lower-elevation nodes gradually lost recordable snow starting from WYD 130 (Figure 5.3g). Migration of the snow line coincided with the upward movement of the zero dew-point temperature (Figure 5.3a). The mean hourly difference between air temperature and dew-point temperature was 0.4 ^{o}C during the four-day event (Figure 5.3h). The air remained relatively saturated across elevations and time span of the entire event. Similar records for other events are shown in supplemental Figures 5.10 to 5.18. The mean R^2 of the dew-point-temperature lapse rate for all 10 events was 0.78 and the RMSE was 149 m, indicating an overall good fit of the dew-point data during the precipitation events.

The amount of solid and total precipitation by event is shown in Figure 5.4a for each node and in Table 5.2 for the basin average above 1500 m. The rain-snow mix was more significant

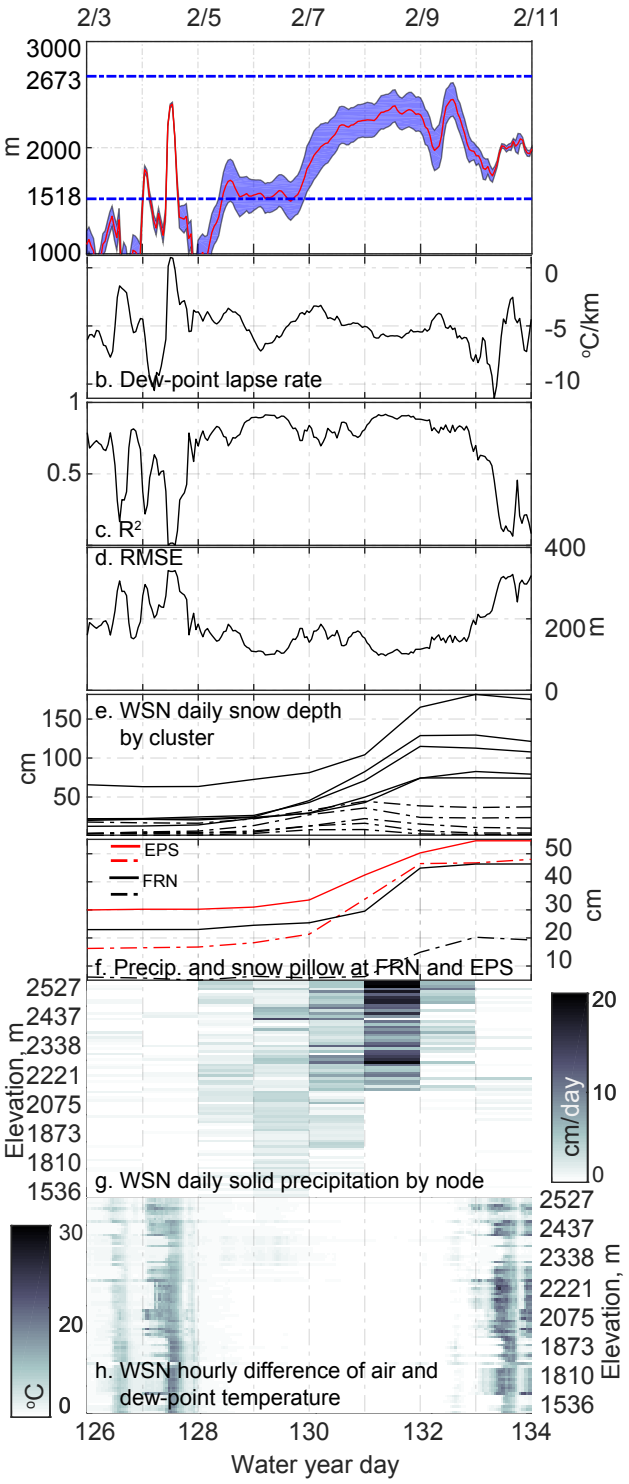


Figure 5.3: Characteristics of event 6: a) elevation of 0 ^{o}C and $\pm$1 ^{o}C window (shaded) of dew-point temperature, b) dew-point temperature lapse rate, c) R^2 value of the lapse-rate fit, d) RMSE of the fit, e) daily average snow depth from the 10 WSN clusters (solid lines are for the five highest-elevation sites and dashed lines the lower five sites), f) snow-pillow and precipitation-gauge data from co-located operational sites, g) daily solid precipitation captured by WSN sensor nodes, and h) hourly differences between air and dew-point temperature.

at lower elevations, with considerable variability across the nodes of each local cluster owing to differences in dew-point temperature. A precipitation lapse rate for each event was calculated as the slope of the linear best-fit line. The values for solid precipitation varied from 1-3 cm km^{-1} in E3, E4 and E10, all relatively small events, to over 30 cm km^{-1} in the largest event E6 (Table 5.6). All slopes in Figure 5.4a were statistically significant ($p < 0.05$), except total precipitation for E8 (p=0.27). For binned data, E3 and E8 had $p > 0.05$ (0.08, 0.11, respectively), with the other eight having statistically significant fitted lines. Slopes for solid precipitation for the binned data were on average about 0.6 cm km^{-1} higher than for all node data. Binning provides a similar number of values across the basin as are currently available from operational sensors. The analysis shown on Figure 5.4b uses 11 bins of 100-m in elevation; results were essentially the same for fewer bins (data not shown). Doing the same analysis of total and liquid precipitation using snow-pillow data, together with dew-point temperature from the wireless-sensor nodes (Figure 5.4b), gives slopes that are about 2.3 cm km^{-1} lower than those for the binned node data, and on average differ from the slopes for the node data by about 25%. Differences between binned-node and snow-pillow data were relatively large for the two events with the steepest precipitation lapse rates, E6 and E7, for about 3.7 and 4.6 cm km^{-1}, respectively. Using rain-gauge data gave precipitation laps rates that were near zero, and in some cases negative for the 10 events (Table 5.6). However, only one event (E5) had a statistically significant trend. The p-values for the other nine events were in the range of 0.16-1.00.

Table 5.2: Partitioning of rain and snow, and area-weighted means, of ten precipitation events (cm)[a]

	WSN, All 80 nodes			WSN, 11 bins			Snow pillow		Rain gauge	
Event	ΔSWE	Total	Mean daily ΔSWE	ΔSWE	Total	Mean Mean ΔSWE	ΔSWE	Total	Gauge only	Blended with WSN
E1	4.0	4.2	2.0	3.8	4.1	1.9	2.2	2.7	2.2	4.2
E2	1.9	2.1	0.6	1.9	2.1	0.6	1.5	1.6	3.5	3.6
E3	3.5	3.5	1.8	3.3	3.3	1.6	4.0	4.0	3.2	3.5
E4	0.4	0.4	0.4	0.6	0.6	0.6	0.9	0.9	0.8	0.8
E5	1.4	1.5	1.4	1.4	1.4	1.4	2.2	2.2	6.5	7.5
E6	9.3	12.1	2.3	9.0	11.9	2.3	11.5	15.3	24.5	25.3
E7	7.2	8.4	0.8	6.8	7.9	0.8	6.5	8.6	15.7	14.7
E8	17.9	19.1	2.6	16.1	17.0	2.3	10.0	11.0	15.4	19.1
E9	9.1	9.3	4.5	8.0	8.5	4.0	4.6	5.1	5.7	9.3
E10	0.5	0.8	0.5	0.4	0.7	0.4	0.4	0.8	1.7	2.2
Mean	5.5	6.2	1.7	5.1	5.7	1.6	4.4	5.2	7.9	9.0
St.Dev	5.5	5.7	1.2	4.9	5.2	1.0	3.8	4.9	7.9	8.6

[a]See Table 5.5 for statistics

Normalizing precipitation lapse rates by the mean precipitation amounts in Table 5.2 gives

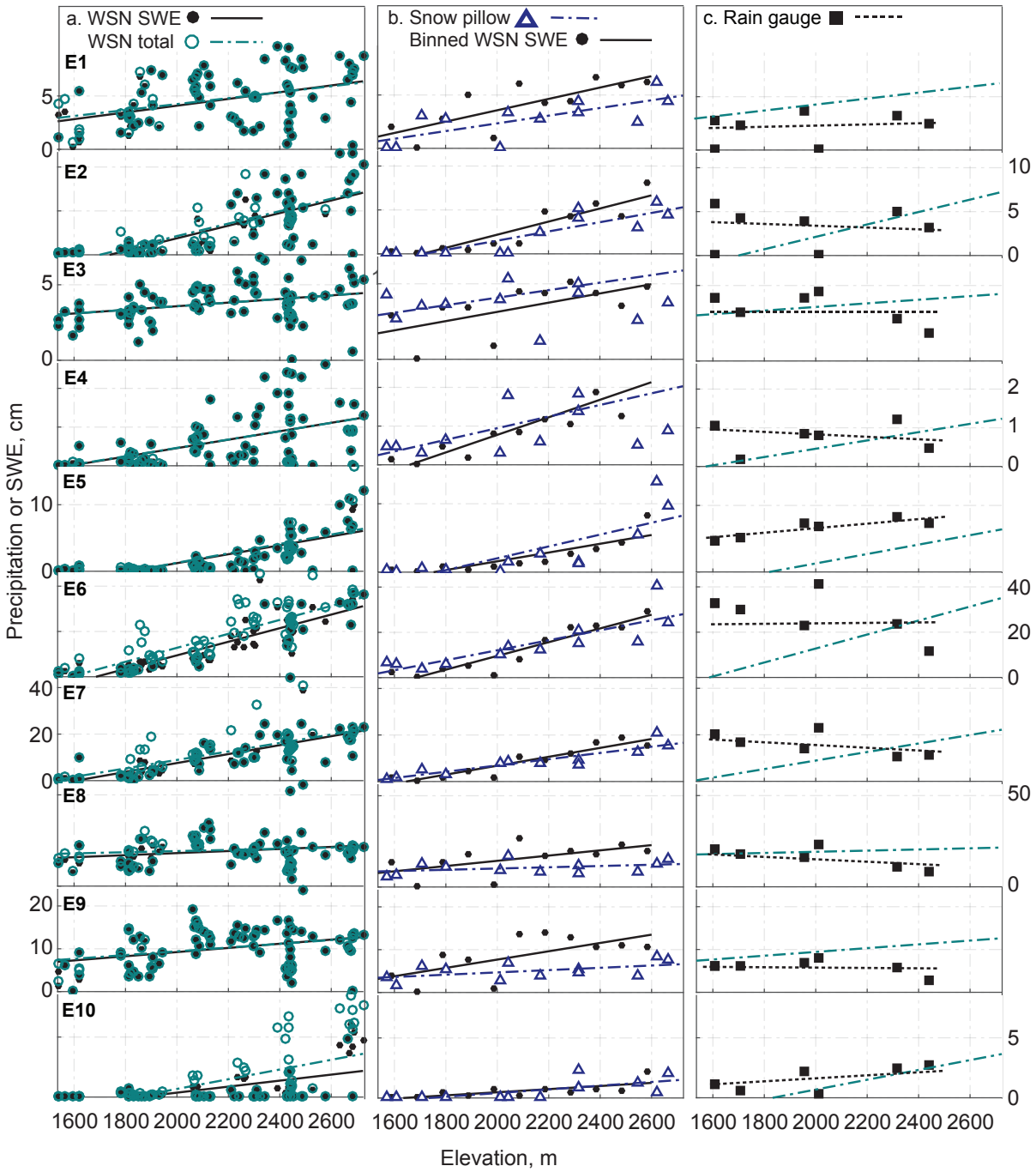


Figure 5.4: Comparison of SWE and total precipitation estimates: a) for 80 sensor nodes within the 10 clusters, b) binned sensor nodes and snow pillows, and c) sensor nodes and rain gauges. The solid and dashed lines are the best linear fit of the data. See Table 5.6 for statistics and slopes of lines. See Table 5.8 and Table 5.9 for a list of snow pillows and rain gauges in the American River above 1500 m.

values of about 2.4±1.8 km^{-1} (mean ± standard deviation) for all nodes, versus 2.2±1.0 km^{-1} for binned data, and 2.0±1.3 km^{-1} for the snow-pillow data. On average, normalized precipitation lapse rates for each event differed by about +33% for the sensor-node versus snow-pillow data. Normalized total precipitation lapse-rate values for sensor-node data range from 0.3-0.8 for E1, E3, E8 and E9 to 2.5-3.2 for E4, E6 and E7 to 3.7-5.1 for E2, E5 and E10.

The contribution of liquid precipitation for mixed rain-snow events can be significant. For example, during event E6, 2.8 cm liquid precipitation, averaged across elevations, was recorded, with 9.3 cm solid precipitation by the nodes (Table 5.2). Using the node data, it is estimated that across the 10 events, about 11% of the estimated total 62 cm of precipitation above elevation 1500 m fell as liquid. However, this likely underestimates the rain contribution to total precipitation at some of the lower sensors. Rain-gauge data, averaged over the basin and summed for all 10 events, give 79 cm total. Using binned node data gives only 57 cm total precipitation (11% liquid), and snow-pillow data give 52 cm total (15% liquid).

SWE and snowmelt

SWE data show a clear elevation trend, with variability also increasing with elevation (Figure 5.5). Maximum SWE occurred around April 1. During the warm and dry 2014 snow season, snow cover accumulated mainly at elevations above 2100 m. At elevations below that, snow melted soon after a precipitation event. At peak accumulation, average SWE measured by the WSN for the 2154 km^2 above 1500 m was 24.7 cm. Thus the estimated water stored in the snowpack at the American River basin above 1500 m was 532 million m^3 (0.43 million acre feet), or 10 cm averaged across the 5311 km^2 basin. The averages of snow-pillow SWE values showed positive biases when compared to the WSN. However, basin-mean SWE on April 1st from snow courses was 15 cm lower than sensor-network average SWE (Figure 5.5).

Sites at higher elevations generally had a longer melt season compared to lower elevations. For sensor-network sites above 2000 m, snow melt out progressed upslope an average of 13 m day^{-1} (R^2 = 0.68) (Figure 5.6). This is comparable to the 14 m day^{-1} observed by Rice et al. [2011] for a dry year (2004) using satellite snow-cover data. The mean melt-out progressed upslope about 25 m day^{-1} (R^2 = 0.67) if lower-elevation sites were considered. The entire melt season lasted roughly 65 days after April 1 (WYD 183), indicated by the first node at BTP to melt out versus the last node at ECP. The error bars in Figure 5.6 indicate a 3-22 day variability in the progress of snowmelt among nodes within each cluster. ECP experienced the longest period (61 days) between the first and the last nodes to melt out. The progression of snowmelt within each site was also recorded. Differences in the timing of melt-out between the sensor-network nodes vs. snow pillows were also apparent (Figure 5.6). Snow-pillow data show an earlier melt out as compared to the cluster means, with 18-, 22- and 30-day differences at VVL, ALP and ECP, respectively.

Related to snowmelt timing, there were also significant differences in cumulative temperature between sensor-network nodes and met-station sensors. For the main snowmelt season, April 4 to June 27, the differences between sensor-network nodes and met-station cumulative values were +24 ^{o}C-day at VVL, and -68 ^{o}C-day at ECP. Using an average degree-day factor of 0.4

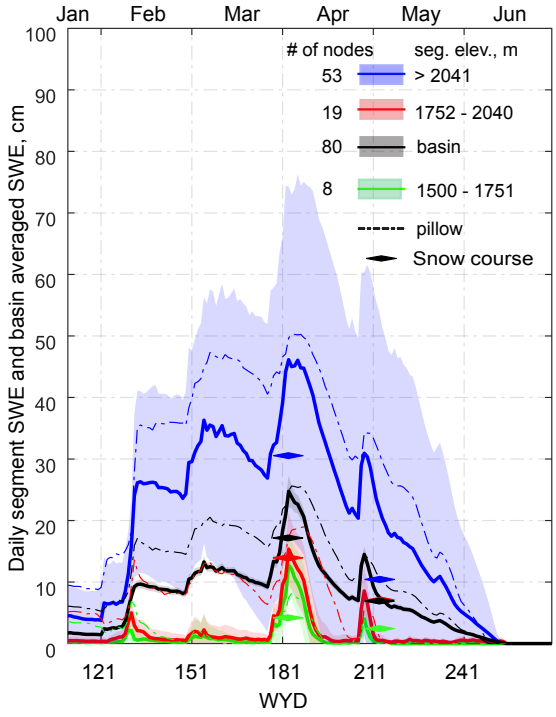


Figure 5.5: Basin mean SWE estimated by averaging SWE from equal-area segments above 1500 m. The shaded area represents the standard deviation of SWE estimated by WSN nodes. Also shown are basin mean SWE values calculated by snow-pillow and snow-course data, using the same method (dashed lines). See Table 5.10 for a list of snow courses in the American River.

cm per $^{o}C^{-}1day^{-1}$[99], the resulting difference in potential snowmelt would amount to about +9.70 cm at VVL and -28 cm SWE at ECP. In contrast, temperature data between the sensor nodes and met-stations from SCN and ALP showed much less difference, +9.7 ^{o}C-day and +4.5 ^{o}C-day (Figure 5.6). The differences in potential snowmelt would be +3.9 and +1.8 cm of SWE for SCN and ALP.

5.4 Discussion

Dew-point temperature and rain-snow transition

Combining dew-point temperature with air temperature has previously been shown to provide a reliable estimate of the timing and phase of precipitation[72]. Precipitation occurs when temperature approximately equals dew-point temperature, indicating saturation of an air parcel. The critical temperature range of separating solid versus liquid is usually wider and more variable for a method based on air temperature compared to one based on dew-point temperature. Therefore, dew-point temperature-based method to determine the phase of precipitation

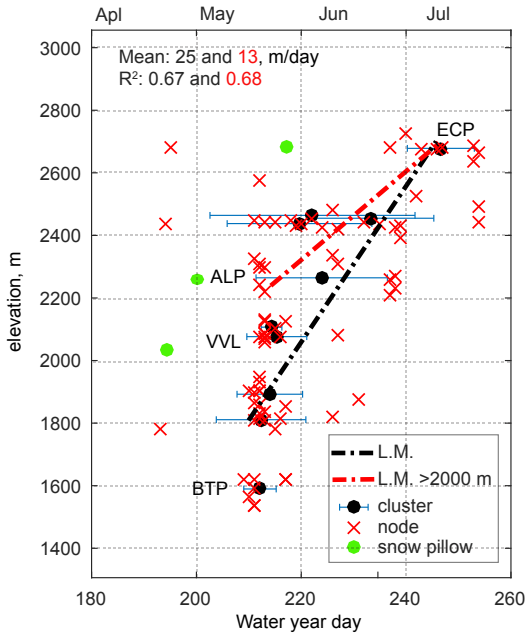


Figure 5.6: Basin mean SWE estimated by averaging SWE from equal-area segments above 1500 m. The shaded area represents the standard deviation of SWE estimated by WSN nodes. Also shown are basin mean SWE values calculated by snow-pillow and snow-course data, using the same method (dashed lines). See Table 5.10 for a list of snow courses in the American River.

is generally less geographically dependent[119]. It has also been observed that using ground-based dew-point temperatures to determine the phase of precipitation is potentially more accurate than radar-based methods due to reduction of error associate to interpreting the radar measurement[68]. Precipitation data from the nearby rain gauges at ALP and EP5 showed that the dual-temperature method reliably mapped the timing of precipitation (Figure 5.3g). The better characterization of the timing of an event provided by the sensor network helps to more-accurately estimate event-based indexes such as lapse rate and RMSE. That is, the reduction of uncertainty in temperature and humidity elevation patterns helps to determine the elevation range associated with the rain-snow transition.

Due to limited relative-humidity measurements at met stations in the basin, dew-point temperature is not routinely available. Previous studies in the American River basin demonstrated the inaccuracy of estimating daily dew-point temperature patterns using empirical method-based spatial-projections algorithms, radiosonde data, or PRISM lapse rates[28]. They suggest that dense field measurements or down-scaled atmosphere-model data are two viable solutions to more-accurately estimate daily dew point. Our telemetered sensor-network data are of the density and quality that can support the dual-temperature method.

The bottom of the rain/snow transition elevation, represented as 1^oC dew-point temperature, ranged from 627 to 2011 m for the ten events (Table 5.1), a wider range than observed at Reynolds Creek, Idaho[79]. Marks et al.[72] did a detailed analysis of the rain-snow transition by putting in place a transect of seven measurement stations providing temperature, humidity and snow depth every 50 m, including precipitation and wind at the bottom (1500 m), top (1800 m) of the catchment, and in a sheltered site just below the top (1750 m) in a 1.8 km^2 subdrainage within the Reynolds Creek Experimental Watershed. Their work showed that the rain - snow transition is very dynamic moving up and sown during mountain storms, and that rain - snow transition level can be reliably determined from site humidity. However, in the absence of such an extensive measurement network, methods to reliably estimate the melting level are still an open question in mountain hydrology[68]. Widely used snow models such as SNOW-17 are very sensitive to the melting-level parameter input[75]. A melting-level error of 500 m can sometimes result in a 200% difference in peak flow prediction[114]. Previous studies interpolated ground melting elevation from atmosphere hydrometeor measurements using Doppler-profiling radar[68, 76]. However, the uncertainty in estimates made from these methods were at best about 300 m in the American River basin. The mean error of estimating dew-point temperature was about 150 m using our WSNs (Table 5.1).

Rain-snow partitioning results are sensitive to the choice of the dew-point temperature thresholds. Marks et al.[72] proposed a 1^oC window, $\pm 0.5\ ^oC$, as the boundary of the transition zone. While they used hourly data for calculation, our study used daily averaged dew-point temperature and SWE in order to mitigate noise in snow depth and possible lags in density data. We also use wider bounds, $\pm 1\ ^oC$, to accommodate uncertainties from measurements and method. The wider bounds allow inclusion of more daily events in the rain/snow transition zone. Specifically, the 100% widening of bounds resulted in an increase in the mean predicted liquid precipitation from 6.2% to 9.9% seasonally and from 14.2% to 21.1% for E6.

Partitioning rain and snow

The ability to estimate liquid-precipitation amounts from solid precipitation could substantially extend the current capability of precipitation observations. Those observations are important, as they are the basis of interpolation for some gridded-precipitation products such as PRISM, WprldClim, and Climate Research Unit CL 2.0[80, 93, 46]. In a study by Lundquist et al.[67], total ground precipitation was assumed to be the increase SWE at elevation above 2500 m when compared to gridded datasets. Seasonally, the assumption holds for American River site for WY 2014. However, the rain/snow transition zone can reach elevations above 2500 m for warmer event such as E6 (Figure 5.4). To extent the analysis to lower elevation and higher temporal resolution, accurate temperature or dew-point temperature data are needed to help reduce the uncertainty in predicting the phase of the precipitation.

Besides local spatial variability of dew-point temperature, the temporal variability of dew point temperature can also produce suitable temperature conditions for snowfall. Event 6 around WYD 131 was an example of such event (Figure 5.2). On WYD 131, 7 cm of precipitation was recorded at the BTP rain gauge (Figure 5.3). With a daily averaged dew-point temperature at

3.1^{o}C (Figure 5.3a), almost all of the precipitation was rain. There was little to no SWE recorded by the five lower WSN clusters (Figure 5.3). At higher elevation clusters, almost all of the precipitation was snow. Despite the high mean daily dew-point temperature, there was still a small but measurable amount of snow deposited at BTP during the event. This was likely due to both the temporal and spatial variability of dew-point temperature, which is not completely captured by a daily, cluster-averaged lapse rate and temperature. An hourly analysis of dew point and precipitation as rain versus snow is feasible using sensor-node data; however, melting during the warm part of the day would need to be accounted for.

On average, the data indicate that up to 89% of the WY 2014 precipitation in these 10 events fell as snow at the sensor nodes. The challenges in predicting the amount of liquid precipitation was smaller compared to previous studies[72]. This is due to the relatively larger elevation differences among sensor nodes, where some higher-elevation nodes receive mostly solid precipitation from all events. This difference could be even smaller if events with 100% liquid precipitation event were accounted for. In this paper, the derived total precipitation relies on some portion of precipitation to be snow in order to calculate the liquid portion of the precipitation. Due to lack of snow precipitation in some event at lower elevation, this method could result in under estimate total precipitation.

Across all events, only 10% of precipitation occurred when daily averaged dew-point temperature was above 0 ^{o}C. The general trend of increasing precipitation followed a weaker orographic effect during some events such as E1, E3, E8 and E9, i.e. higher elevations generally received similar amount of precipitation compared to lower elevations. During those events, elevations account for no more than 27% of the variance in snow deposition.

Inspection of moisture-transport suggests that a steeper precipitation lapse rate was associated with atmospheric-river events, which made landfall, e.g. events E2, E6 and E7 (M. Dettinger 2016, personal communication). Events E1, E3, E4 and E9 showed no effect of atmospheric river moisture. The outliers are E8, the largest event of the season, which had a weak precipitation lapse rate but was weakly associated with atmospheric-river transport; and E5 and E10, smaller events that had steeper lapse rates, but were not associated with land-falling atmospheric rivers.

For 1998-2010, the overall contribution of atmospheric rivers to Sierra Nevada SWE was about 35%, with the rest contributed by less-intense but more-frequent precipitation events citeGuan2012. The majority (80%) of the total SWE was contributed by days with ΔSWE between 0.5-4.5 cm day-1[35]. The contribution of atmospheric rivers for events in that range (30%) doubles for ΔSWE >4.5 cm day^{-}1. For ΔSWE <0.5 cm day^{-}1, atmospheric rivers account for only 2-3% of the total number of days, but about 40% of the cumulative SWE. Eight of the 10 events that we analyzed were in the range of 0.5-4.5 cm day^{-}1. The Sierra Nevada averages about nine days per year with atmospheric rivers contributing precipitation[36].

Comparison of basin-wide precipitation estimates

The rain-from-snow reconstruction result was compared to measurements of two co-located sites at EP5 and ALP. Both sites simultaneously recorded SWE on a snow pillow and precipita-

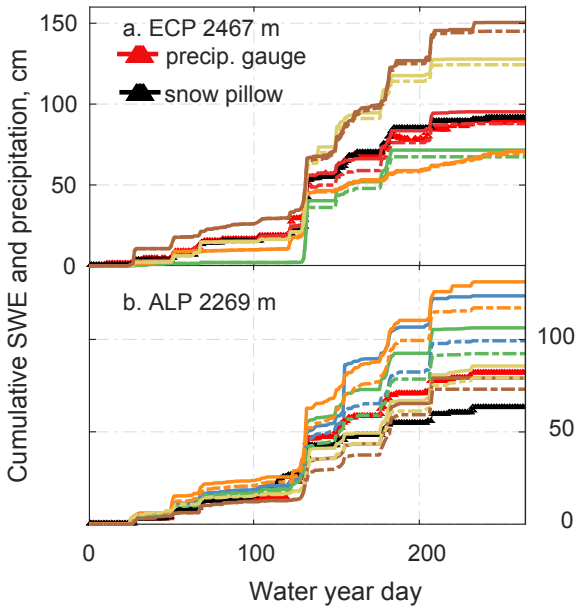


Figure 5.7: Cumulative solid (dashed lines) and total precipitation (solid lines) from the sensor-network nodes at ALP and ECP compared to co-located snow-pillow and precipitation-gauge data, For clarity, data from only five nodes are shown.

tion by a rain gauge. Seasonally, most of the precipitation was predicted as solid at EP5 from the sensor-network data (Figure 5.7a). Good agreements were observed between the WSN and operational data. On the other hand, the snow pillow and precipitation gauge at ALP measured about 25% liquid precipitation versus 12-42% liquid predicted by the sensor-network nodes (Figure 5.7b). At ALP, there were larger uncertainties associated with predicted liquid water content in snow. Compared to ECP, ALP is more susceptive to the effects of rain/snow transition due to moderate elevation. Besides the natural heterogeneity, canopy interception and wind redistribution of snowfall to the wireless-sensor nodes at ALP could add sources of uncertainties in predicting liquid water content. Although, these uncertainties may be resolved using subsequent-year data from the WSN, gridded and then evaluated and using the spatial snow measurements.

It is useful to compare precipitation estimates from operational versus sensor-node data across the basin above 1500 m elevation. First, it is apparent from Figure 5.3b that the 80 sensor nodes capture landscape variability that is not apparent in snow-pillow data. While the two data sets give broadly similar precipitation lapse rates, the WSN data potentially offer four types of added value. First, the WSN data better capture spatial variability, and thus should give improved estimates of gridded SWE (paper in preparation), essential for spatially explicit modeling and forecasting tools. Second, compared to snow-pillow data, the WSN provides redundancy in measurements that can bridge data gaps and uncertainties in snow-pillow measurements. Third, compared to snow-course data, the WSN provides temporally continuous

data over the same or larger spatial domain. This is especially important given the shifting rain-snow mix and earlier peak in SWE at snow courses as climate warms. Fourth, these differences together reduce uncertainty in precipitation, snowpack storage and snowmelt.

It is apparent from Figure 5.4c that total precipitation estimated using snow depth, and dew-point temperature from the WSN show differences in precipitation across elevation and other landscape attributes that are not captured by the rain-gauge data. However, the relationship between precipitation estimates from the two data sets differs for warmer versus colder events. For the colder events E1, E3, E8 and E9, total precipitation estimated by the WSN is higher than that measured in the precipitation gauges. For these four events the rain-snow transition elevation was at or near the 1500-m lower limit of the WSN, and little liquid precipitation contributed to the total. For the other six events, the 0 ^{o}C dew-point elevation was well above 1500 m, and liquid precipitation was more important. A second estimate of total event precipitation can come from combining the two types of measurements, i.e. using precipitation-gauge data at lower elevations and WSN data at higher elevations. Given the few precipitation gauges available in the basin, we used an average value for lower elevations, combining that with WSN data for elevations above that value. The result is higher precipitation estimates for the mixed rain-snow events (Table 5.2). Over all 10 events, this results in a season total of 90 cm, versus 79 cm for gauges alone and 62 cm for WSN alone. These represent respective total precipitation amounts of 1.9, 1.7 and 1.3 billion cubic meters (1.6, 1.4, 1.1 million acre-ft). A potential value added from this combined estimate is to provide a more-accurate elevation dependence of precipitation.

An additional comparison comes from summing precipitation by elevation across all events, by elevation. Here, we compare two estimates (Figure 5.8). Note the strong elevation dependence of the WSN estimates versus PRISM data, which is gridded based on operational data. PRISM reflects the mean of precipitation-gauge data in the basin. Here, the point of intersection is about 2400 m, which is still in the snow-dominated part of the basin. Rain-gauge data are shown for reference, and represent part of the data used in the PRISM interpolation. Going forward, improvements in precipitation estimates across the basin should focus more on snow in this region, and rain below. In the mixed rain-snow zone, blending the two should be pursued.

Basin SWE and snow melt

Similar to other studies, the variability in SWE increases with elevation, but the coefficient of variation showed no distinct trend in elevation at the spatial scale of the WSNs [83]. Seasonally, the coefficient of variation for the lower one third of the basin is 1.04, similar to the upper one third at 1.05 (Table 5.3). The differences in variability at similar elevations can in large part be accounted by differences in forest canopy coverage[17]. At a few higher-elevation sites (e.g. ECP and MTL), the high variability in SWE was caused by high SWE values recorded by a small subset of nodes.

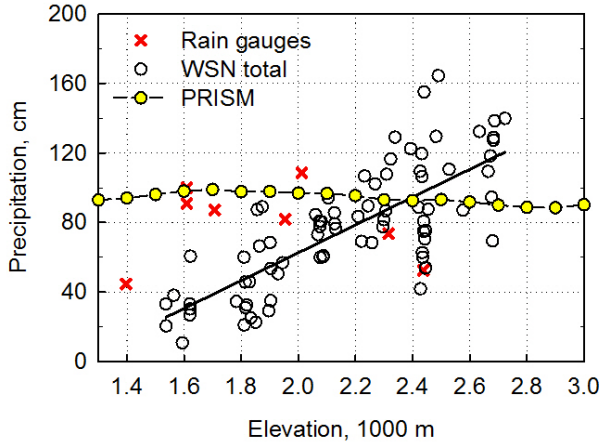


Figure 5.8: Seasonal total comparison of precipitation estimates with PRISM.

Table 5.3: Daily average of μ, σ and coefficient of variation (CV) of elevation segments

	Lower 1/3		Middle 1/3		Upper 1/3		Basin >1500 m	
	WSN	Pillow	WSN	Pillow	WSN	Pillow	WSN	Pillow
Ave. daily SWE, cm	0.53	1.03	0.9	3.05	9.15	12.89	3.5	5.7
Ave. daily σ, cm	0.49	0.64	0.78	2.14	8.61	10.64	3.3	4.5
Ave. daily CV	1.04	0.8	0.92	0.66	1.05	0.99	1.0	0.8

The study characterized the basin's mean SWE by three zones similar to Welch et al.[111], which showed that strategically placed sensors in three to eight 'clusters' could efficiently characterize SWE in the basin. They recommended measurements taken from clusters with strong elevation differences. Our SWE measurements followed this approach, aimed at producing a representative basin-wide SWE. Continuous measurements at higher temporal resolution can aid in accurately monitoring a basin's hydrologic condition. The combination of dry and warm conditions during WY 2014 places it within the most severe drought periods (WY2012-2014) in the last 1200 years[34]. There were no large storms, and of the few events that occurred, only five deposited snow at all ten sensor-network cluster sites. Snow deposited at lower elevations started to melt as soon as the precipitation event ended. Monthly snow courses at lower elevations missed the timing of those small 'peaks' in SWE for the season (Figure 5.3).

Reconstruction of SWE backwards from the last day of snow towards the peak accumulation using energy-based methods is of interest to hydrologists for its higher accuracy and robustness to climate change[36]. One of the major uncertainties affecting the performance of the method is the knowledge of the snow-disappearance date[89]. To determine snow cover, Satellite remote-sensing observes snow-covered area, however obstruction of view by cloud cover and canopy are significant challenges for this method. Ground-based sensors can measure snow-covered versus snow-free conditions without that uncertainty. On-the-ground snow

measurements can also be used to verify satellite and LiDAR remote-sensing datasets, and account for their limited under-canopy measurements.

5.5 Conclusion

A spatially distributed WSN, installed across the 2154 km^2 portion of the American River basin above 1500 m elevation, reliably provided spatial measurements of temperature, relative humidity and snow depth. Mixed rain/snow events were common at elevations below 2100 m; and distributed-sensor data showed significant heterogeneity in rain versus snow precipitation that was not apparent in more-limited operational data. Distributed dew-point temperature measurements provided estimates of ground melting levels that were consistent with distributed observations of snow accumulation. Using daily dew-point temperature and the amount of snow accumulation at each node to estimate the fraction of rain versus snow resulted in an underestimate of total precipitation below the 0 ^{o}C dew-point elevation, which averaged 1730 m across 10 precipitation events, but was as high as 2170 m during one warm event. Rain-gauge measurements failed to capture the elevation and other topographic variability of precipitation. However, blending lower-elevation rain-gauge data with higher-elevation sensor-node data for each event provided precipitation estimates that were on average 15-30% higher than using either set of measurements alone. Given the increasing importance of liquid precipitation in a warming climate, a strategy that blends distributed measurements of both liquid and solid precipitation will provide the most-accurate basin-wide precipitation estimates. However, blending data from the current operational rain-gauge and snow-pillow measurements underestimates basin-wide precipitation and snowpack storage. Distributed, representative measurements also improve upon operational estimates of snowpack water storage and snowmelt amount and snowmelt timing across the basin.

5.6 Supporting information

Introduction

This document includes supporting tables and figures referenced in the main text.

Supplemental tables

Table 5.4: list of WSN sites and existing co-located instruments in the American River basin

WSN Site name	Abbr.	Co-located oper. site abbr. on CDEC	Elev., m	Lat,Lon	# sensor nodes with data in WY 2014
Schneiders	SCN	SCN[b]	2673	38.745, -120.067	8
Echo Peak	ECP	EP5[b,c,d] , ECSe	2478	38.848, -120.079	7
Mt. Lincoln	MTL		2477	39.287, -120.328	8
Caples Lake	CAP	CAP[b,c,d,e]	2437	38.711, -120.042	9
Alpha	ALP	FRN[d], APH[b,c,d,e]	2269	38.804, -120.216	10
Duncan Peak	DUN		2097	39.154, -120.510	6
Van Vleck	VVL	VVL[b,d]	2069	38.944, -120.306	6
Onion Creek	ONN	ONN[e]	1891	39.274, -120.356	10
Robbs Saddle	RBB	RBBb,[d], RBV[e]	1812	38.912, -120.379	9
Bear Trap	BTP	BTP[b,c]	1518	39.095, -120.577	8

[a]Out of 10 nodes installed at each local cluster
[b]Rain gauge. Note that SCN an VVL did not have data in WY 2014
[c]Snow depth
[d]Snow pillow. CAP had no data in WY 2014
[e]Snow course

Table 5.5: SNOTEL sites used in deriving snow density

Site name	Elevation, m
Fallen Leaf	1901
Truckee #2	1984
Ward Creek #3	2028
CSS Lab	2089
Echo Peak	2338
Rubicon #2	2344
Forestdale Creek	2444
Squaw Valley G.C.	2447
Independence Lake	2546
Carson Pass	2546

Table 5.6: Statistical results from fitting SWE and precipitation data (main text, Figure 5.4)

	WSN all nodes		11 bins WSN		Pillow	Gauge
Event	ΔSWE	Total	ΔSWE	Total	ΔSWE	precip.
	P-value					
E1	0.00	0.00	0.01	0.05	0.01	0.80
E2	0.00	0.00	0.00	0.00	0.00	0.72
E3	0.00	0.00	0.08	0.08	0.24	1.00
E4	0.00	0.00	0.00	0.00	0.07	0.49
E5	0.00	0.00	0.00	0.00	0.00	0.00
E6	0.00	0.00	0.00	0.00	0.00	0.95
E7	0.00	0.00	0.00	0.00	0.00	0.30
E8	0.05	0.27	0.11	0.19	0.24	0.35
E9	0.00	0.00	0.02	0.04	0.08	0.74
E10	0.00	0.00	0.01	0.01	0.01	0.16
	R2					
E1	0.18	0.15	0.45	0.31	0.52	0.01
E2	0.56	0.54	0.72	0.71	0.76	0.02
E3	0.10	0.10	0.25	0.25	0.13	0.00
E4	0.27	0.27	0.78	0.78	0.28	0.08
E5	0.56	0.52	0.72	0.67	0.63	0.76
E6	0.77	0.70	0.79	0.77	0.70	0.00
E7	0.56	0.48	0.77	0.76	0.82	0.18
E8	0.05	0.02	0.22	0.15	0.13	0.15
E9	0.12	0.10	0.42	0.34	0.28	0.02
E10	0.32	0.35	0.49	0.53	0.49	0.30
	Slope, cm km^{-1}					
E1	3.16	2.79	4.90	3.89	3.61	0.43
E2	7.18	7.09	5.79	6.34	5.27	-0.99
E3	1.20	1.20	2.60	2.60	2.49	0.00
E4	1.09	1.08	1.61	1.59	1.50	-0.30
E5	6.71	7.06	4.33	4.40	8.90	3.29
E6	30.01	30.84	25.77	30.71	22.07	0.93
E7	19.23	18.23	18.04	20.31	13.47	-5.39
E8	5.65	3.17	12.99	10.89	3.44	-5.96
E9	4.86	4.37	10.43	8.99	2.68	-0.63
E10	2.50	4.05	1.16	2.31	1.54	1.32

Table 5.7: Table 5.6. (cont.)

	WSN all nodes		11 bins		WSN	Gauge
Event	ΔSWE	Total	ΔSWE	Total	ΔSWE	Precip.
	Normalized slope, km^{-1}					
E1	0.79	0.67	1.28	0.96	1.61	0.19
E2	3.71	3.36	3.12	3.02	3.50	-0.29
E3	0.34	0.34	0.79	0.79	0.62	0.00
E4	2.53	2.47	2.66	2.58	1.66	-0.35
E5	4.64	4.64	3.19	3.06	4.00	0.50
E6	3.24	2.55	2.86	2.58	1.92	0.04
E7	2.68	2.17	2.64	2.57	2.06	-0.34
E8	0.31	0.17	0.81	0.64	0.34	-0.39
E9	0.53	0.47	1.31	1.06	0.58	-0.11
E10	5.13	4.83	3.06	3.28	3.67	0.79
	RMSE, cm km^{-1}					
E1	2.26	2.26	1.77	1.89	1.40	1.52
E2	2.15	2.19	1.18	1.31	1.21	2.53
E3	1.22	1.22	1.47	1.47	2.58	1.47
E4	0.60	0.60	0.28	0.27	0.97	0.39
E5	1.98	2.28	0.87	1.00	2.77	0.72
E6	5.59	6.81	4.28	5.49	5.89	14.21
E7	5.69	6.43	3.19	3.72	2.54	4.60
E8	8.37	8.46	8.00	8.37	3.58	5.68
E9	4.51	4.45	3.99	4.05	1.77	1.76
E10	1.23	1.86	0.38	0.71	0.64	0.80

Table 5.8: Snow pillow sites in the American River basin

Site name	Abbrev.	Elevation, m
Schneiders	SCN	2667
Lake Lois	LOS	2621
Carson Pass	CXS	2546
Caples Lake[a]	CAP	2438
Alpha	ALP	2316
Forni Ridge	FRN	2316
Silver Lake	SIL	2164
Van Vleck	VVL	2042
Huysink	HYS	2012
Robbs Saddle	RBB	1798
Greek Store	GKS	1707
Blue Canyon	BLC	1609
Robbs Powerhouse	RBP	1570

[a]Insufficient data for WY 2014

Table 5.9: Precipitation gauges in the American River basin above 1500 m

Site Name	Abbrev.	Elevation, m
Schneiders[a]	SCN	2667
Caples Lake	CAP	2438
Caples (Twin) Lake[a]	CPT	2438
Forni Ridge	FRN	2316
Alpha (SMUD)[a]	ALP	2316
Silver Lake	SIL	2164
Van Vleck[a]	VVL	2042
Huysink	HYS	2012
Loon lake (SMUD)	LON	1954
Robbs Saddle[a]	RBB	1798
Greek Store	GKS	1707
Blue Canyon #2 (ETI)[a]	BL2	1609
Blue Canyon (DWR-2)	BYM	1609
Blue Canyon	BLC	1609
Robbs Powerhouse[a]	RBP	1570

[a]Insufficient data for WY 2014

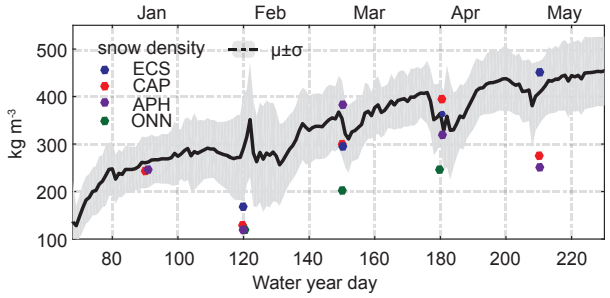


Figure 5.9: Basin-wide daily mean (μ) and standard deviation (σ) of snow density calculated from 9 snow-pillow sites in and around the American River basin 5.5. We used SNOTEL sites adjacent to the American R. basin for this analysis, as the snow-pillow data in the basin had significant gaps. Densities measured by four-snow courses sites within American River basin are shown as discrete points.

Table 5.10: List of snow course sites in the American River basin above 1500 m

Snow course site ID	Snow course site name	Elev., m	Survey date	SWE, cm
106	Upper Carson Pass	2591	25-Mar	17.5
331	Lower Carson Pass	2560	25-Mar	17.0
107	Caples Lake	2438	25-Mar	11.5
365	Alpha	2316	2-Apr	13.5
338	Lost Corner Mt.	2286	3-Apr	11.0
108	Echo Summit	2271	28-Mar	12.0
110	Lake Audrain	2225	28-Mar	12.0
109	Silver Lake	2164	27-Mar	5.0
316	Wrights Lake	2103	31-Mar	11.5
113	Phillips	2073	1-Apr	8.0
320	Lyons Creek	2042	2-Apr	10.5
289	Tamarack Flat	1996	1-Apr	14.5
114	Wabena Meadows	1920	28-Mar	6.5
369	Miranda Cabin	1890	28-Mar	3.0
120	Onion Creek	1859	26-Mar	1.0
371	Diamond Crossing	1844	28-Mar	1.0
122	Talbot Camp	1753	28-Mar	1.5
322	Robbs Valley	1707	31-Mar	3.0
127	Ice House	1615	31-Mar	1.5

Suplemental figures

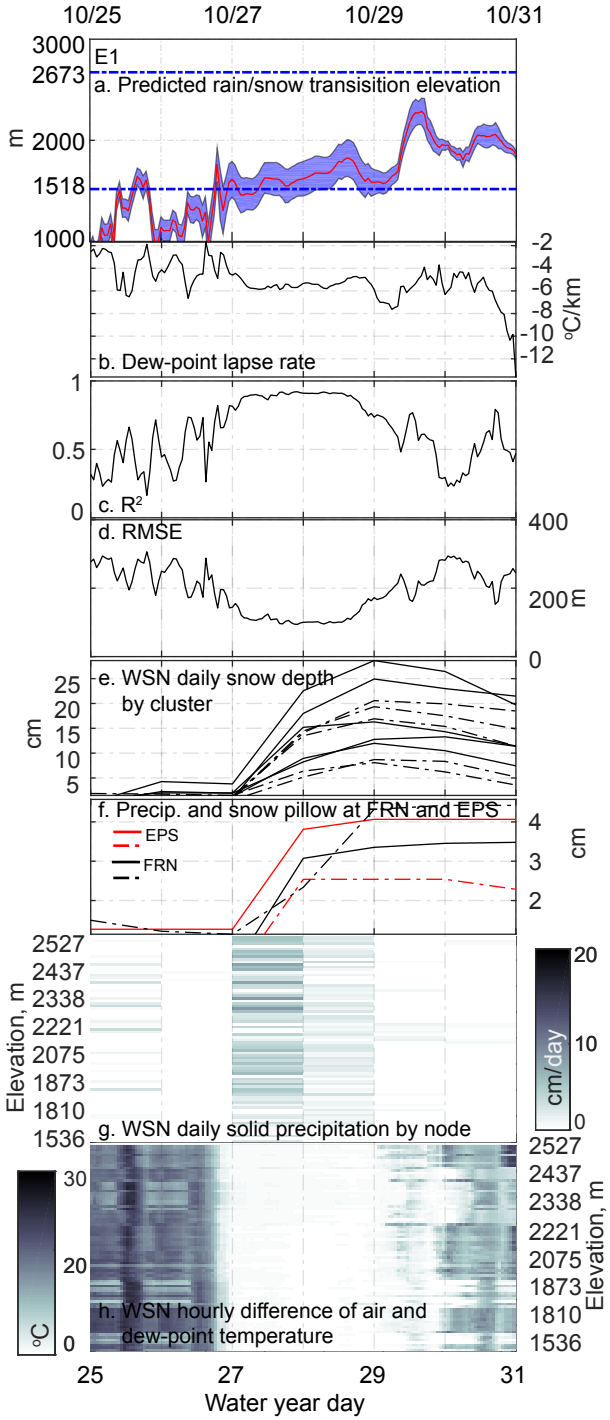


Figure 5.10: Characteristics of event 1 to 10 in Figure 5.10 - 5.18: a) elevation of 0^oC and ± 1 ^{o}C window (shaded) of dew-point temperature, b) dew-point temperature lapse rate, c) R^2 value of the lapse-rate fit, d) RMSE of the fit, e) daily average snow depth from the 10 WSN clusters (solid lines are for the 5 highest-elevation sites and dashed lines the lower five sites), f) snow-pillow and precipitation-gauge data from co-located operational sites, g) daily solid precipitation captured by WSN sensor nodes, and h) hourly differences between air and dew-point temperature.

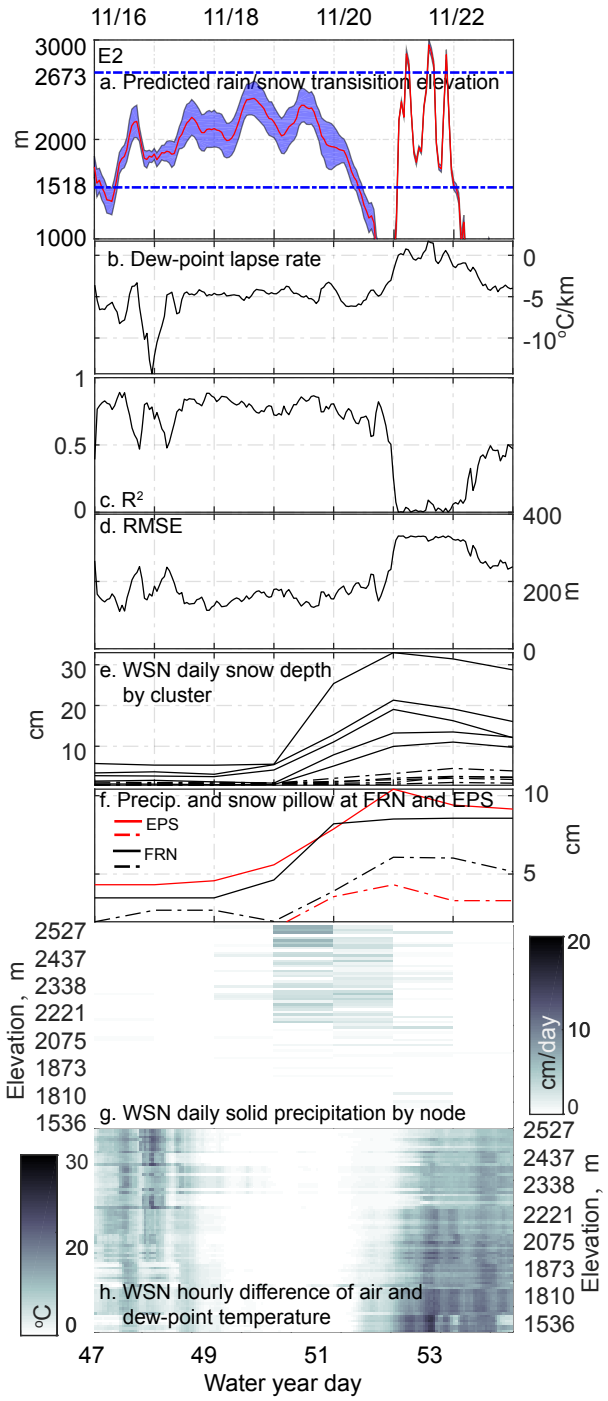


Figure 5.11: E2

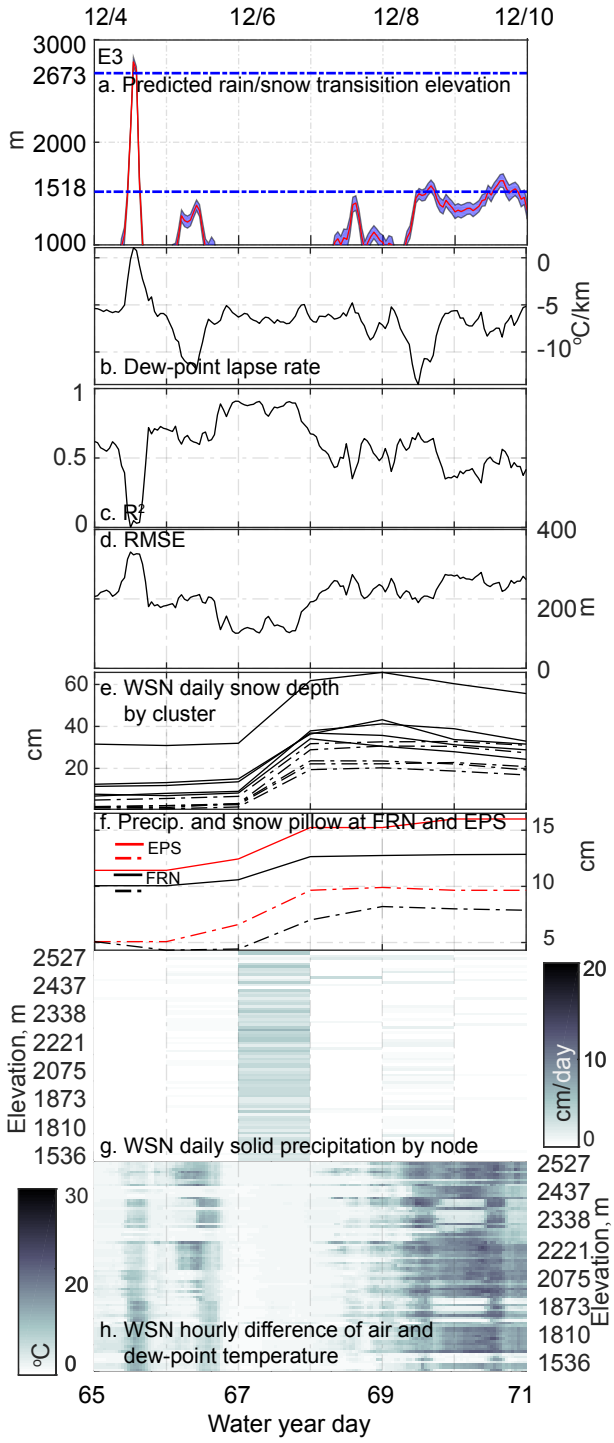


Figure 5.12: E3

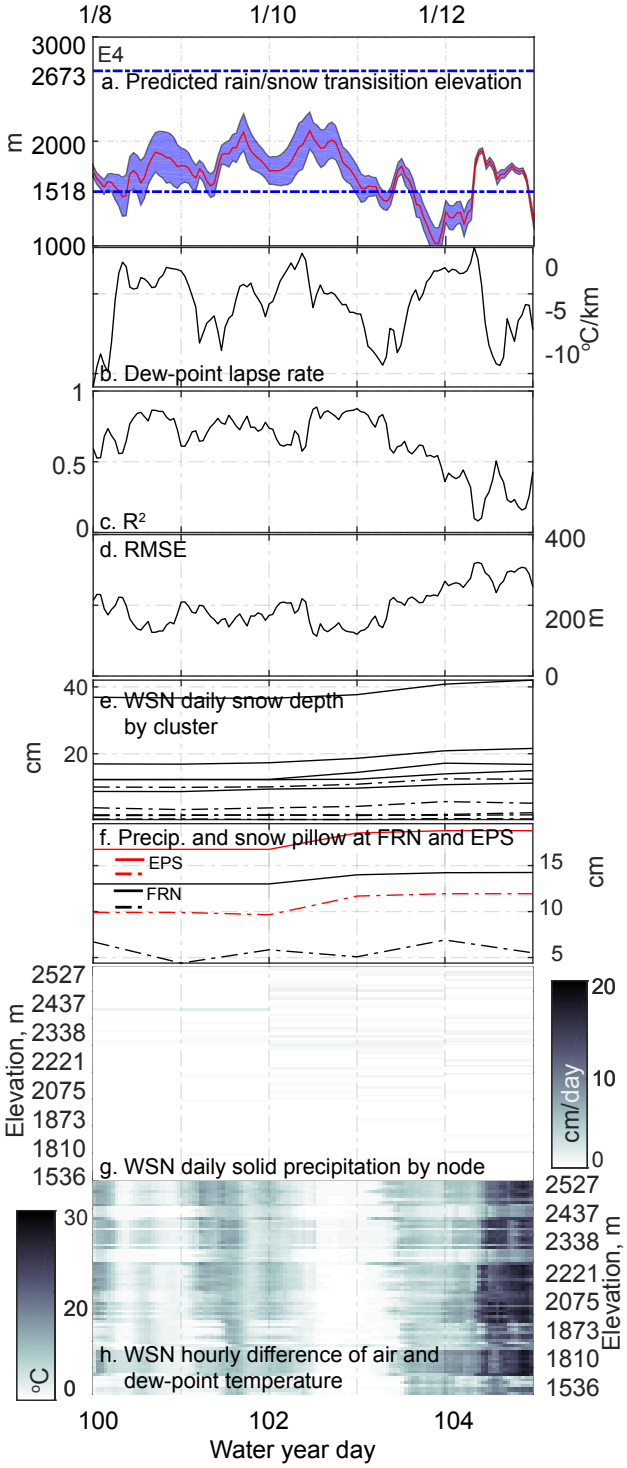


Figure 5.13: E4

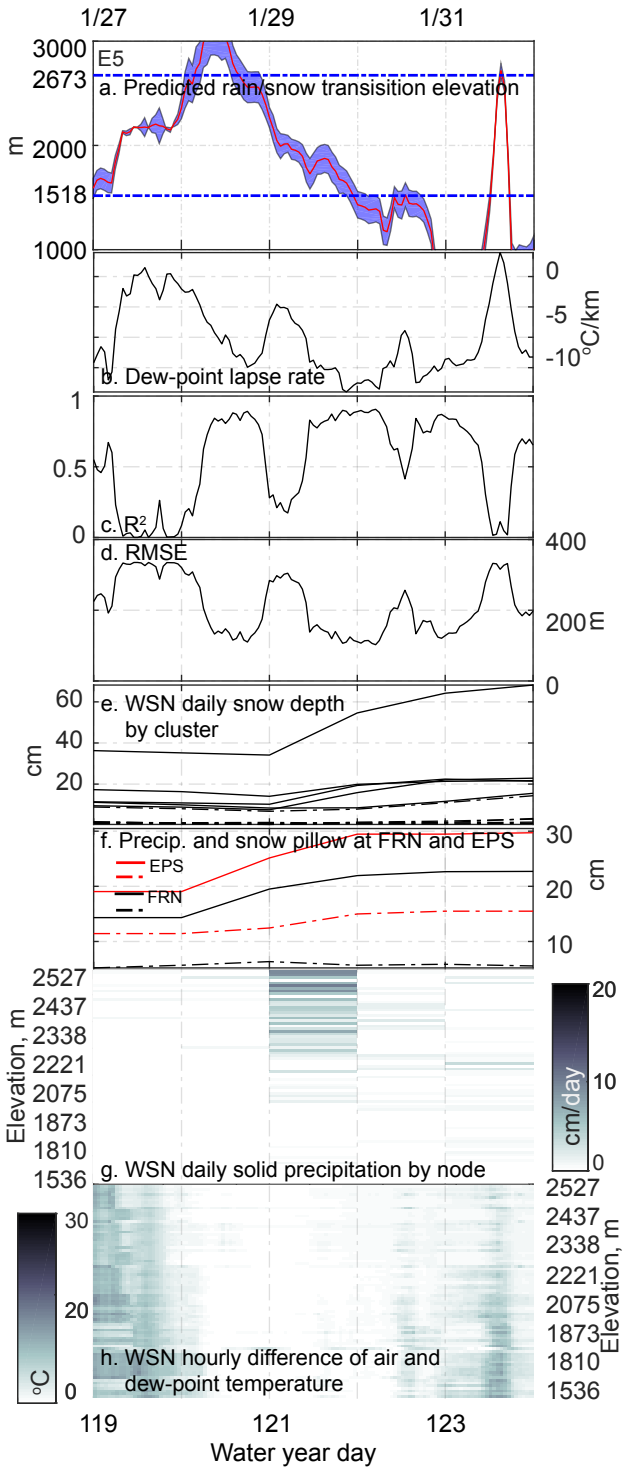


Figure 5.14: E5

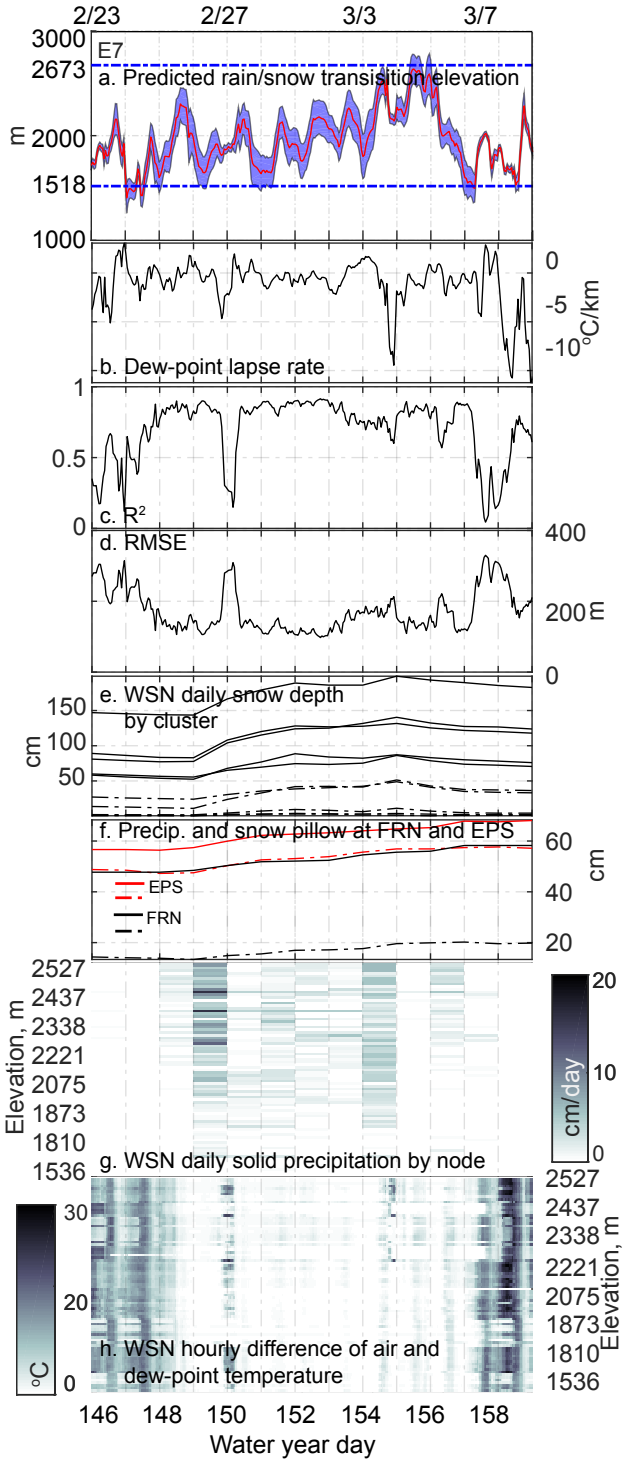


Figure 5.15: E7

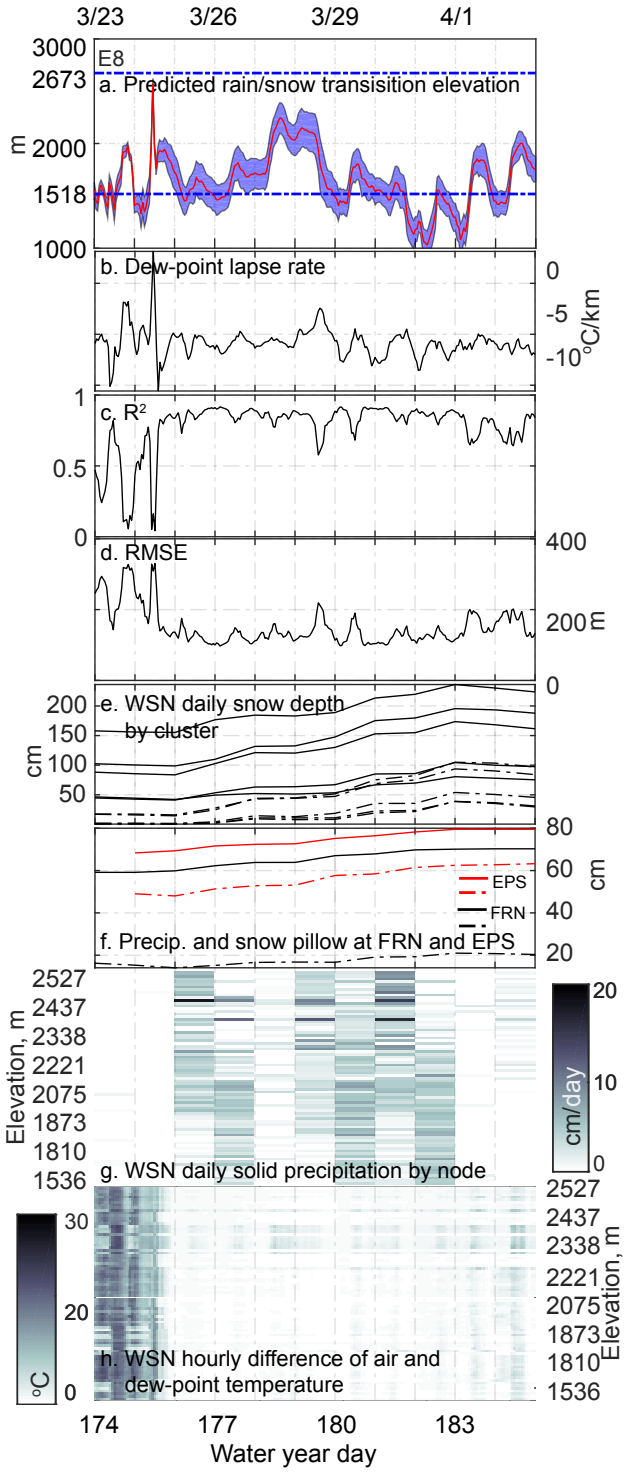


Figure 5.16: E8

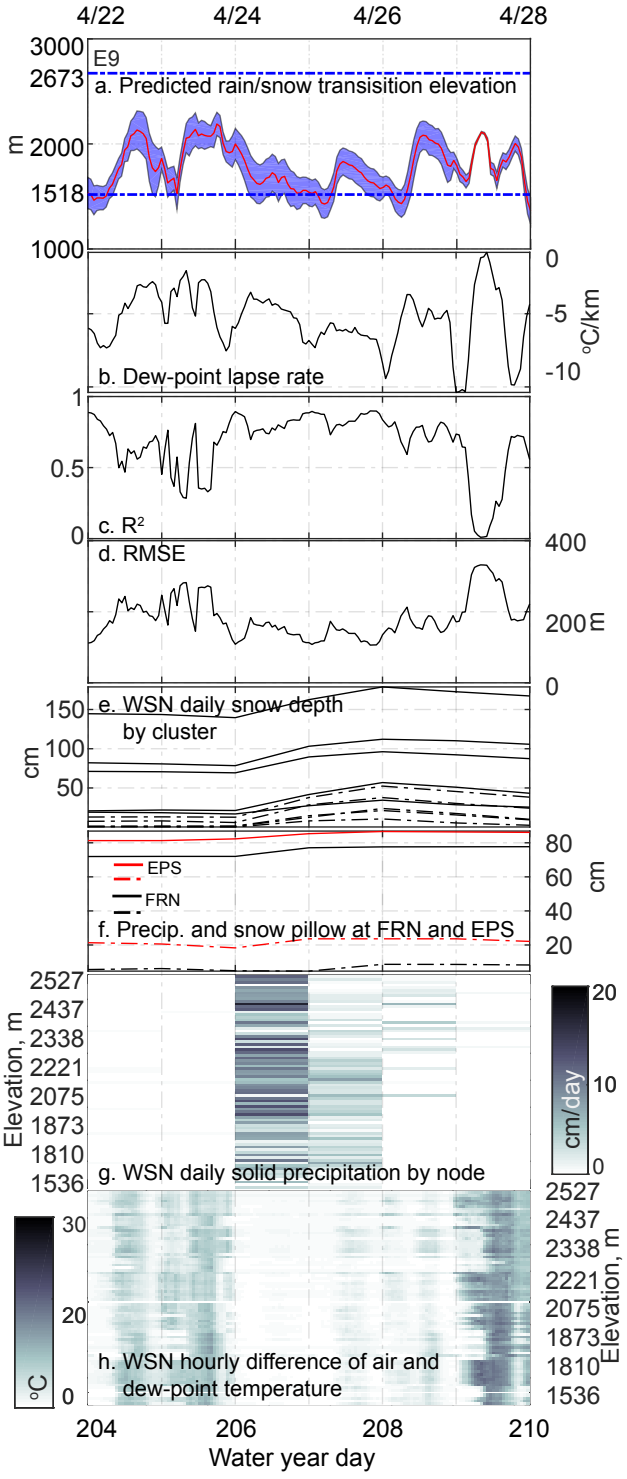


Figure 5.17: E9

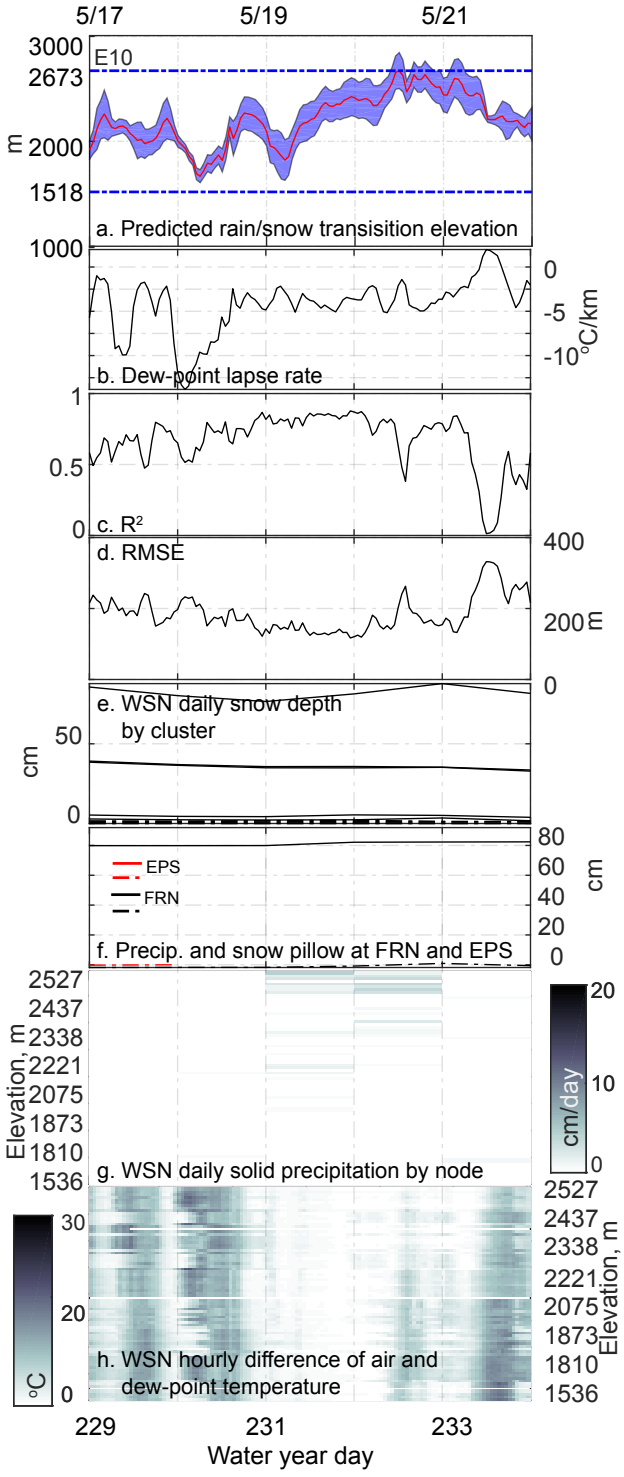


Figure 5.18: E10

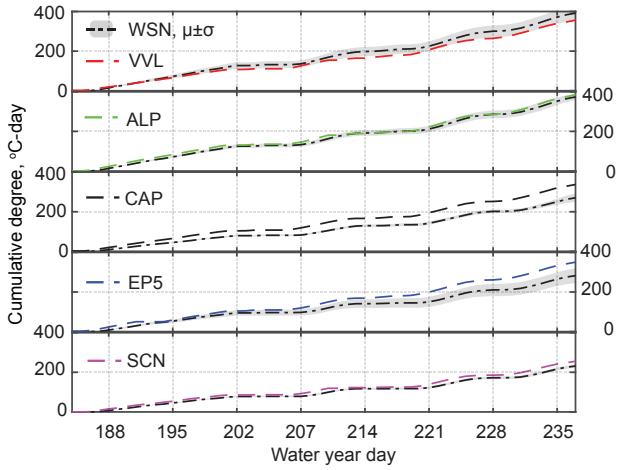


Figure 5.19: Cumulative degree day for WSN and the operational network at five clusters. The mean (μ) plus standard deviation (σ) are shown for the WSN sites. Data from operational-network sensors are shown in dashed lines.

Chapter 6

Conclusion

My research has sought to design and deploy a ground-based system of WSNs to the Mountains of Sierra Nevada, so as to monitor snowpacks and variables related to hydrologic changes. Taking the series of instruments as a whole, the system provides more comprehensive and representative snapshots of the important hydrologic variables, allowing new questions to be answered. While comparing measurement results with data from the existing system, models, and remote sensing product, we recognize that no single approach can characterize the complexity of the natural phenomenon, and that all the available tools must be combined to address questions in mountain hydrology. The core contributions of the research are two fold. The four chapters of the dissertation have attempted to address some of the early questions.

Chapter 2 explored a series of constraints in using WSN in the remote environment. This chapter discussed the basic requirements for an effective WSN to perform distributed sensing of snow in the remote mountains. The size of the American River Basin and the number of devices deployed as part of the ARHO project confirm the clear advantages of using low-power wireless technology over a wired field solution. The idea of using an off-the-shelf low-power wireless solution has proven a successful one, although the ARHO project has demonstrated that the networking aspect is only a small piece of the challenge.

Following the deployment in the ARHO, Chapter 3 examined a data-driven model to characterize the connectivity between network elements in complex real-world environment using machine-learning-based method. The result showed improvement in error reduction over canonical and empirical propagation models. The methodology is a new way of predicting connectivity in wireless networks. The result of the study can be applied directly to systems alike in the similar environment to predict network connectivity and line quality.

I discussed the methodology of instrumenting the American river basin with a system comprised of clusters of wireless sensor nodes in Chapter 3. Compared to existing operational sensors, the wireless-sensor network reduces uncertainty in water-balance measurements in three distinct ways. Redundant measurements of temperature improved the robustness of temperature lapse-rate estimation, reducing cross-validation error compared to that of using met-station data alone. Second, distributed measurements capture local variability and constrain uncertainty, compared to point measures, in attributes important for hydrologic mod-

eling, such as air and dew-point temperature and snow precipitation. Third, the distributed relative-humidity measurements offer a unique capability to monitor upper-basin patterns in dew-point temperature and better characterize precipitation phase and the elevation of the rain/snow transition.

Chapter 5 identified mixed rain/snow events, which were common at elevations below 2100 m; and distributed-sensor data showed significant heterogeneity in rain versus snow precipitation that was not apparent in more limited operational data. Distributed dew-point temperature measurements provided estimates of ground melting levels that were consistent with distributed observations of snow accumulation. Using daily dew-point temperature and the amount of snow accumulation at each node to estimate the fraction of rain versus snow resulted in an underestimate of total precipitation below the 0 ^{o}C dew-point elevation, which averaged 1730 m across 10 precipitation events, but was as high as 2170 m during one warm event. Rain-gauge measurements failed to capture the elevation and other topographic variability of precipitation. However, blending lower-elevation rain-gauge data with higher-elevation sensor-node data for each event provided precipitation estimates that were on average 15-30% higher than using either set of measurements alone. Given the increasing importance of liquid precipitation in a warming climate, a strategy that blends distributed measurements of both liquid and solid precipitation will provide the most-accurate basin-wide precipitation estimates. However, blending data from the current operational rain-gauge and snow-pillow measurements underestimates basin-wide precipitation and snowpack storage. Distributed, representative measurements also improve upon operational estimates of snowpack water storage and snowmelt amount and snowmelt timing across the basin.

My research has highlighted the importance of network monitoring and management tools. Although the SmartMesh IP manager offers a complete interface to query the state of the network, receive alerts, and verify the performance of the network, a multi-network management interface is missing. Such a solution would complement an existing SmartMesh IP network by offering a unified interface to assist during network deployment, visualize the network, run network health routines, display the sensor data, and log maintenance activity. REALMS associate team (github.com/realms-team/), and plan on deploying it within six months. The connectivity model developed in Chapter 3 is an opportunity to derive methods to improve network robustness and boost the efficiency of deployment by automating the process of repeater placement. Future study of this subject could use the model developed in Chapter 3 to optimize network topology in complex terrain by modeling the connectivity of the network over the set of feasible signal-repeater placements. This would facilitate the automated deployment of new networks to ensure they are robust to path loss, and significantly reduce human cost during deployment.

Taken as a whole, my research sets the stage for more in-depth analyses with the dataset being collected. My research provide a rich dataset for future work in hydrologic modeling. The distributed in-situ dataset opens new ways to interpolate hydrologic variables across the upper American River basin by blending remote and ground sensing dataset using geospatial analysis. The dataset can also improve our knowledge on how uncertainty propagates in snow/hydrologic models. Snow accumulation and melt models require calibration to make accurate estimates, and uncertainties can exist in the forcing data, model parameters, and model

structures. The model error is often related to the inability to characterize the uncertainty in various model components. The distributed WSN measurements contain critical model parameters, allowing calculation of the uncertainty of the forcing . It provides ground truth for the model response, i.e. changes in snowdepth. Given a fixed model structure, future effort can be spent on examining the efficacy of the calibrated model parameters across different regions of the basin in hoping of developing a better strategy to calibrate snow/hydrologic models.

Bibliography

[1] Mohammad Abu Alsheikh et al. “Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications”. In: *IEEE Communication Surveys & Tutorials* 16.4 (2014), pp. 1996–2018.

[2] Nicola Accettura and Giuseppe Piro. “Optimal and Secure Protocols in the IETF 6TiSCH Communication Stack”. In: *International Symposium on Industrial Electronics (ISIE)*. IEEE. 2014, pp. 1469–1474.

[3] B. Ainslie and P. L. Jackson. “Downscaling and Bias Correcting a Cold Season Precipitation Climatology over Coastal Southern British Columbia Using the Regional Atmospheric Modeling System (RAMS)”. In: *Journal of Applied Meteorology and Climatology* 49.5 (2010), pp. 937–953. ISSN: 1558-8424. DOI: 10.1175/2010JAMC2315.1.

[4] I Akyildiz et al. “Wireless sensor networks: a survey”. In: *Computer Networks* 38.4 (2002), pp. 393–422. ISSN: 13891286. DOI: 10.1016/S1389-1286(01)00302-4. arXiv: 1004.3164.

[5] S. P. Anderton, S. M. White, and B. Alvera. “Evaluation of spatial variability in snow water equivalent for a high mountain catchment”. In: *Hydrological Processes* 18.3 (2004), pp. 435–453. ISSN: 08856087. DOI: 10.1002/hyp.1319.

[6] Joaquim A. R. Azevedo and Filipe E. S. Santos. “An Empirical Propagation Model for Forest Environments at Tree Trunk Level”. In: *IEEE Transactions on Antennas and Propagation* 59.6 (2011), pp. 2357–2367.

[7] Roger C. Bales, Robert Rice, and Sujoy B. Roy. “Estimated Loss of Snowpack Storage in the Eastern Sierra Nevada with Climate Warming”. In: *Journal of Water Resources Planning and Management* 141.2 (2015), p. 04014055. ISSN: 0733-9496. DOI: 10.1061/(ASCE)WR.1943-5452.0000453.

[8] Roger C. Bales et al. “Mountain Hydrology of the Western United States”. In: *Water Resources Research* 42.8 (2006), pp. 1–13.

[9] Roger C. Bales et al. “Soil Moisture Response to Snowmelt and Rainfall in a Sierra Nevada Mixed-Conifer Forest”. In: *Vadose Zone Journal* 10 (2011), pp. 786–799.

[10] Benjamin Balk and Kelly Elder. “Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed”. In: *Water Resources Research* 36.1 (2000), pp. 13–26. ISSN: 00431397. DOI: 10.1029/1999WR900251.

[11] R Beckwith, D Teibel, and P Bowen. “Report from the Field: Results from an Agricultural Wireless Sensor Network”. In: *International Conference on Local Computer Networks (LCN)*. IEEE. Tampa, FL, USA, 2004, pp. 471–478.

[12] Achim Berger et al. “Energy Efficient and Reliable Wireless Sensor Networks - An Extension to IEEE 802.15.4e”. In: *EURASIP Journal on Wireless Communications and Networking* 126.1 (2014), pp. 1–12.

[13] H. R. Bogena et al. “Evaluation of a low-cost soil water content sensor for wireless network applications”. In: *Journal of Hydrology* 344.1-2 (2007), pp. 32–42. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2007.06.032.

[14] H. R. Bogena et al. “Potential of Wireless Sensor Networks for Measuring Soil Water Content Variability”. In: *Vadose Zone Journal* 9.4 (2010), pp. 1002–1013.

[15] David R. Bowling et al. “13C content of ecosystem respiration is linked to precipitation and vapor pressure deficit”. In: *Oecologia* 131.1 (2002), pp. 113–124. ISSN: 00298549. DOI: 10.1007/s00442-001-0851-y.

[16] Leo Breiman. “Random Forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.

[17] Martyn P. Clark et al. “Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review”. In: *Water Resources Research* 47.7 (2011), pp. 1–23. ISSN: 00431397. DOI: 10.1029/2011WR010745.

[18] David W. Clow et al. “Evaluation of SNODAS snow depth and snow water equivalent estimates for the Colorado Rocky Mountains, USA”. In: *Hydrological Processes* 26.17 (2012), pp. 2583–2591. ISSN: 08856087. DOI: 10.1002/hyp.9385.

[19] Christopher Daly et al. “Physiographically Sensitive Mapping of Climatological Temperature and Precipitation across the Conterminous United States”. In: *International Journal of Climatology* 28.15 (2008), pp. 2031–2064.

[20] Silvia Demetri, Gian Pietro Picco, and Lorenzo Bruzzone. “Estimating Low-power Radio Signal Attenuation in Forests: A LiDAR-based Approach”. In: *International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE. Fortaleza, Brazil, 2015.

[21] Zigbee Document. “ZigBee RF4CE : ZRC Profile Specification”. In: (2009).

[22] Rusty Dodson and Danny Marks. “Daily air temperature interpolated at high spatial resolution over a large mountainous region”. In: *Climate Research* 8.1 (1997), pp. 1–20. ISSN: 0936577X. DOI: 10.3354/cr008001.

[23] Lance Doherty, Jonathan Simon, and Thomas Watteyne. “Wireless Sensor Networks: Challenges and Solutions”. In: *Microwave Journal* 2012.1 (2012), pp. 22–34.

[24] Jeff Dozier. “Mountain hydrology, snow color, and the fourth paradigm”. In: *Eos* 92.43 (2011), pp. 373–374. ISSN: 00963941. DOI: 10.1029/2011EO430001.

[25] Tyler a. Erickson, Mark W. Williams, and Adam Winstral. "Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States". In: *Water Resources Research* 41.4 (2005), pp. 1–17. ISSN: 00431397. DOI: 10.1029/2003WR002973.

[26] Jennifer Erxleben, Kelly Elder, and Robert Davis. "Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains". In: *Hydrological Processes* 16.18 (2002), pp. 3627–3649. ISSN: 08856087. DOI: 10.1002/hyp.1239.

[27] Richard Essery and John Pomeroy. "Vegetation and Topographic Control of Wind-Blown Snow Distributions in Distributed and Aggregated Simulations for an Arctic Tundra Basin". In: *Journal of Hydrometeorology* 5.5 (2004), pp. 735–744. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2004)005<0735:VATCOW>2.0.CO;2.

[28] Shara I. Feld, Nicoleta C. Cristea, and Jessica D. Lundquist. "Representing atmospheric moisture content along mountain slopes: Examination using distributed sensors in the Sierra Nevada, California". In: *Water Resources Research* 49.7 (2013), pp. 4424–4441. ISSN: 00431397. DOI: 10.1002/wrcr.20318.

[29] Fabrizio Fenicia, Jeffrey J. McDonnell, and Hubert H G Savenije. "Learning from model improvement: On the contribution of complementary data to process understanding". In: *Water Resources Research* 44.6 (2008). ISSN: 00431397. DOI: 10.1029/2007WR006386.

[30] Yoav Freund and Robert E Schapire. "A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139.

[31] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The Elements of Statistical Learning*. Vol. 1. Berlin, Germany: Springer Series in Statistics, 2001.

[32] A Gardner and M Sharp. "Sensitivity of net mass-balance estimates to near-surface temperature lapse rates when employing the degree-day method to estimate glacier melt". In: *Annals of Glaciology* 50.1 (2009), pp. 80–86. ISSN: 02603055. DOI: 10.3189/172756409787769663.

[33] Edwin Gilbert. "Research Issues in Wireless Sensor NetworkApplications: A Survey". In: *International Journal of Information and Electronics Engineering* 2.5 (2012), pp. 702–706. ISSN: 20103719. DOI: 10.7763/IJIEE.2012.V2.191.

[34] Daniel Griffin and Kevin J Anchukaitis. "How unusual is the 2012 - 2014 California drought ?" In: *Geophysical Research Letters* 41 (2014), pp. 9017–9023. ISSN: 00948276. DOI: 10.1002/2014GL062433.1..

[35] Bin Guan et al. "Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements". In: *Geophysical Research Letters* 37.20 (2010). ISSN: 00948276. DOI: 10.1029/2010GL044696.

[36] Bin Guan et al. "Snow water equivalent in the Sierra Nevada: Blending snow sensor observations with snowmelt model simulations". In: *Water Resources Research* 49.August (2013), pp. 5029–5046. ISSN: 00431397. DOI: 10.1002/wrcr.20387.

[37] V.C. Gungor and G.P. Hancke. “Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches”. In: *IEEE Transactions on Industrial Electronics* 56.10 (2009), pp. 4258–4265.

[38] V.C. Gungor, Bin Lu, and G.P. Hancke. “Opportunities and Challenges of Wireless Sensor Networks in Smart Grid”. In: *IEEE Transactions on Industrial Electronics* 57.10 (2010), pp. 3557–3564.

[39] Xiu-Ming Guo et al. “A Model with Leaf Area Index and Apple Size Parameters for 2.4 GHz Radio Propagation in Apple Orchards”. In: *Precision Agriculture* 16.1 (2015), pp. 180–200.

[40] Joaquin Gutierrez et al. “Automated Irrigation System Using a Wireless Sensor Network and GPRS Module”. In: *IEEE Transactions on Instrumentation and Measurement* 63.1 (2014), pp. 166–176.

[41] M.P.M. Hall. *Radiowave Propagation Effects on Next-generation Fixed-services Terrestrial Telecommunication Systems*. Tech. rep. ICT COST Action 235. European Cooperation in Science and Technology (COST), 1996.

[42] Carl Hartung et al. “FireWxNet: A Multi-Tiered Portable Wireless System for Monitoring Weather Conditions in Wildland Fire Environments”. In: *International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM. Uppsala, Sweden, 2006.

[43] Hirofumi HASHIMOTO et al. “Satellite-based estimation of surface vapor pressure deficits using MODIS land surface temperature data”. In: *Remote Sensing of Environment* 112.1 (2008), pp. 142–155. ISSN: 00344257. DOI: 10.1016/j.rse.2007.04.016.

[44] Andreas Hasler et al. “Wireless Sensor Networks in Permafrost Research: Concept, Requirements, Implementation and Challenges”. In: *International Conference on Permafrost (NICOP)*. 2008, pp. 669–674.

[45] Andrew Hedrick et al. “Independent evaluation of the SNODAS snow depth product using regional-scale lidar-derived measurements”. In: *The Cryosphere* 9.1 (2015), pp. 13–23. ISSN: 1994-0424. DOI: 10.5194/tc-9-13-2015.

[46] Robert J. Hijmans et al. “Very high resolution interpolated climate surfaces for global land areas”. In: *International Journal of Climatology* 25.15 (2005), pp. 1965–1978. ISSN: 08998418. DOI: 10.1002/joc.1276.

[47] Jason L. Hill and David E. Culler. “Mica: A wireless platform for deeply embedded networks”. In: *IEEE Micro* 22.6 (2002), pp. 12–24. ISSN: 02721732. DOI: 10.1109/MM.2002.1134340.

[48] C. G. Homer et al. “Completion of the 2011 National Land Cover Database for the Conterminous United States-Representing a Decade of Land Cover Change Information”. In: *Photogrammetric Engineering and Remote Sensing* 81.5 (2015), pp. 345–354.

[49] Goran Horvat. “Power Consumption and RF Propagation Analysis on”. In: (2012), pp. 222–226.

[50] Pei Huang et al. “The Evolution of MAC Protocols in Wireless Sensor Networks: A Survey”. In: *IEEE Communications Surveys and Tutorials* 15.1 (2012), pp. 101–120.

[51] Shengli Huang et al. “Modeling Monthly Near-Surface Air Temperature from Solar Radiation and Lapse Rate: Application over Complex Terrain in Yellowstone National Park”. In: *Physical Geography* 29.2 (2008), pp. 158–178. ISSN: 0272-3646. DOI: 10.2747/0272-3646.29.2.158.

[52] *IEEE802.15.4-2011: Low-Rate Wireless Personal Area Networks (LR-WPANs)*. IEEE Computer Society, 2011.

[53] Digi International. “XBee ® /XBee-PRO ® RF Modules”. In: *Product Manual v1.xEx-802.15.4 Protocol* (2009), pp. 1–69.

[54] International Telecommunication Union (ITU). “Influence of Terrain Irregularities and Vegetation on Tropospheric Propagation”. In: *CCIR XVth Plenary Assembly*. Vol. V: Propagation in Non-Ionised Media. ITU-R Report 236-6. Dubrovnik, Croatia, 1986.

[55] Suming Jin et al. “A comprehensive change detection method for updating the National Land Cover Database to circa 2011”. In: *Remote Sensing of Environment* 132 (2013), pp. 159–175. ISSN: 00344257. DOI: 10.1016/j.rse.2013.01.012.

[56] Branko Kerkez et al. “Design and Performance of a Wireless Sensor Network for Catchment-Scale Snow and Soil Moisture Measurements”. In: *Water Resources Research* 48.July 2011 (2012), pp. 1–18.

[57] Yunseop Kim, Robert G. Evans, and William M. Iversen. “Remote Sensing and Control of an Irrigation System Using a Distributed Wireless Sensor Network”. In: *IEEE Transactions on Instrumentation and Measurement* 57.7 (2008), pp. 1379–1387.

[58] Manfred Kirchner et al. “Altitudinal temperature lapse rates in an Alpine valley: trends and the influence of season and weather patterns”. In: *International Journal of Climatology* 33.3 (2013), pp. 539–555. ISSN: 08998418. DOI: 10.1002/joc.3444.

[59] P. Zion Klos, Timothy E. Link, and John T. Abatzoglou. “Extent of the rain-snow transition zone in the western U.S. under historic and projected climate”. In: *Geophysical Research Letters* (2014), pp. 4560–4568. ISSN: 1944-8007. DOI: 10.1002/2014GL060500.

[60] P. R. Kormos et al. “Soil, snow, weather, and sub-surface storage data from a mountain catchment in the rain-snow transition zone”. In: *Earth System Science Data* 6.1 (2014), pp. 165–173. ISSN: 18663516. DOI: 10.5194/essd-6-165-2014.

[61] George Kuczera et al. “There are no hydrological monsters, just models and observations with large uncertainties!” In: *Hydrological Sciences Journal* 55.6 (2010), pp. 980–991. ISSN: 0262-6667. DOI: 10.1080/02626667.2010.504677.

[62] Osman Kurnaz and Selcuk Helhela. “Near Ground Propagation Model for Pine Tree Forest Environment”. In: *International Journal of Electronics and Communications* 68.1 (2014), pp. 944–950.

[63] Mark G. Lawrence. “The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications”. In: *Bulletin of the American Meteorological Society* 86.2 (2005), pp. 225–233. ISSN: 00030007. DOI: 10.1175/BAMS-86-2-225.

[64] Michael Lehning et al. “Instrumenting the Earth: Next-Generation Sensor Networks and Environmental Science”. In: *The Fourth Paradigm: Data-Intensive Scientific Discovery*. 2009, pp. 45–51. ISBN: 9780982544204.

[65] Z. Li et al. “Practical Deployment of an In-Field Soil Property Wireless Sensor Network”. In: *Computer Standards & Interfaces* 36.2 (2014), pp. 278–287.

[66] Jessica D. Lundquist and Daniel R. Cayan. “Surface Temperature Patterns in Complex Terrain: Daily Variations and Long-term Change in the Central Sierra Nevada, California”. In: *Journal of Geophysical Research: Atmospheres* 112.11 (2007), pp. 1–15.

[67] Jessica D. Lundquist et al. “High-Elevation Precipitation Patterns: Using Snow Measurements to Assess Daily Gridded Datasets across the Sierra Nevada, California*”. In: *Journal of Hydrometeorology* 16.4 (2015), pp. 1773–1792. ISSN: 1525-755X. DOI: 10.1175/JHM-D-15-0019.1.

[68] Jessica D. Lundquist et al. “Rain versus Snow in the Sierra Nevada, California: Comparing Doppler Profiling Radar and Surface Observations of Melting Level”. In: *Journal of Hydrometeorology* 9.2 (2008), pp. 194–211. ISSN: 1525-755X. DOI: 10.1175/2007JHM853.1.

[69] Muhammad Adeel Mahmood, Winston K.G. Seah, and Ian Welch. “Reliability in Wireless Sensor Networks: A Survey and Challenges Ahead”. In: *Computer Networks* 79 (2015), pp. 166–187.

[70] Alan Mainwaring et al. “Wireless Sensor Networks for Habitat Monitoring”. In: *Proceedings of the 1st {ACM} International Workshop on Wireless Sensor Networks and Applications*. 2002, p. 88. ISBN: 1-58113-589-0. DOI: 10.1145/570748.570751.

[71] Wolf Dietrich Marchand and Killingtveit. “Statistical probability distribution of snow depth at the model sub-grid cell spatial scale”. In: *Hydrological Processes* 19.2 (2005), pp. 355–369. ISSN: 08856087. DOI: 10.1002/hyp.5543.

[72] D. Marks et al. “An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin”. In: *Advances in Water Resources* 55 (2013), pp. 98–110. ISSN: 03091708. DOI: 10.1016/j.advwatres.2012.11.012.

[73] Danny Marks et al. “The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: a case study of the 1996 Pacific Northwest flood”. In: *Hydrological Processes* 12.10-11 (1998), pp. 1569–1587. ISSN: 08856087. DOI: 10.1002/(SICI)1099-1085(199808/09)12:10/11<1569::AID-HYP682>3.0.CO;2-L.

[74] Sergey Y. Matrosov, Kurt A. Clark, and David E. Kingsmill. "A polarimetric radar approach to identify rain, melting-layer, and snow regions for applying corrections to vertical profiles of reflectivity". In: *Journal of Applied Meteorology and Climatology* 46.2 (2007), pp. 154–166. ISSN: 15588424. DOI: 10.1175/JAM2508.1.

[75] Edwin P. Maurer and Clifford Mass. "Using Radar Data to Partition Precipitation into Rain and Snow in a Hydrologic Model". In: *Journal of Hydrologic Engineering* 11.3 (2006), pp. 214–221. ISSN: 1084-0699. DOI: 10.1061/(ASCE)1084-0699(2006)11:3(214).

[76] Justin R Minder and David E Kingsmill. "Mesoscale Variations of the Atmospheric Snow Line over the Northern Sierra Nevada: Multiyear Statistics, Case Study, and Mechanisms". In: *Journal of the Atmospheric Sciences* 70.3 (2012), pp. 916–938. ISSN: 0022-4928. DOI: 10.1175/JAS-D-12-0194.1.

[77] Noah P. Molotch and Roger C. Bales. "Scaling snow observations from the point to the grid element: Implications for observation network design". In: *Water Resources Research* 41.11 (2005), pp. 1–16. ISSN: 00431397. DOI: 10.1029/2005WR004229.

[78] Noah P. Molotch and Roger C. Bales. "SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence". In: *Hydrological Processes* 20.September 2005 (2006), pp. 723–739. ISSN: 08856087. DOI: 10.1002/hyp.6128.

[79] A. Nayak et al. "Long-term snow, climate, and streamflow trends at the reynolds creek experimental watershed, Owyhee Mountains, Idaho, United States". In: *Water Resources Research* 46.6 (2010). ISSN: 00431397. DOI: 10.1029/2008WR007525.

[80] Mark New et al. "A high-resolution data set of surface climate over global land areas". In: *Climate Research* 21.1 (2002), pp. 1–25. ISSN: 0936577X. DOI: 10.3354/cr021001.

[81] R. Oren et al. "Sap-flux-scaled transpiration responses to light, vapor pressure deficit, and leaf area reduction in a flooded Taxodium distichum forest." In: *Tree physiology* 19.6 (1999), pp. 337–347. ISSN: 1758-4469. DOI: 10.1093/treephys/19.6.337.

[82] Ram Oren et al. "Sensitivity of mean canopy stomatal conductance to vapor pressure deficit in a flooded<SMALL> Taxodium distichum</SMALL> L. forest: hydraulic and non-hydraulic effects". In: *Oecologia* 126.1 (2001), pp. 21–29. ISSN: 0029-8549. DOI: 10.1007/s004420000497.

[83] L. Baker Perry et al. "Synoptic Classification of Snowfall Events in the Great Smoky Mountains, USA". In: *Physical Geography* 31.2 (2010), pp. 156–171. ISSN: 0272-3646. DOI: 10.2747/0272-3646.31.2.156.

[84] Kieu-Ha Phung et al. "Schedule-Based Multi-Channel Communication in Wireless Sensor Networks: A Complete Design and Performance Evaluation". In: *Ad Hoc Networks* 26 (2015), pp. 88–102.

[85] Kristofer S. J. Pister and Lance Doherty. "TSMP: Time synchronized mesh protocol". In: *Proc.Parallel and Distributed Computing and Systems (PDCS 08)* (2008), pp. 391–398.

[86] Stefan Pohl et al. “Potential of a Low-cost Sensor Network to Understand the Spatial and Temporal Dynamics of a Mountain Snow Cover”. In: *Water Resources Research* 50.3 (2014), pp. 2533–2550.

[87] John Pomeroy, Xing Fang, and Chad Ellis. “Sensitivity of snowmelt hydrology in Marmot Creek, Alberta, to forest cover disturbance”. In: *Hydrological Processes* 26.12 (2012), pp. 1891–1904. ISSN: 08856087. DOI: 10.1002/hyp.9248.

[88] S. D. Prince et al. “Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using advanced very high-resolution radiometer satellite observations: Comparison with field observations”. In: *Journal of Hydrology* 212-213.1-4 (1998), pp. 230–249. ISSN: 00221694. DOI: 10.1016/S0022-1694(98)00210-8.

[89] Mark S. Raleigh and Jessica D. Lundquist. “Comparing and Combining SWE Estimates from the SNOW-17 Model using PRISM and SWE Reconstruction”. In: *Water Resources Research* 48.1 (2012), pp. 1–16.

[90] Mark S. Raleigh et al. “Approximating snow surface temperature from standard temperature and humidity data: New possibilities for snow model and remote sensing evaluation”. In: *Water Resources Research* 49 (2013), pp. 8053–8069. ISSN: 00431397. DOI: 10.1002/2013WR013958.

[91] F. Martin Ralph, Paul J. Neiman, and Richard Rotunno. “Dropsonde Observations in Low-Level Jets over the Northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean Vertical-Profile and Atmospheric-River Characteristics”. In: *Monthly Weather Review* 133.4 (2005), pp. 889–910. ISSN: 0027-0644. DOI: 10.1175/MWR2896.1.

[92] Jens Christian Refsgaard. “Parameterisation, calibration and validation of distributed hydrological models”. In: *Journal of Hydrology* 198.1-4 (1997), pp. 69–97. ISSN: 00221694. DOI: 10.1016/S0022-1694(96)03329-X.

[93] Clim Res et al. “A Knowledge-based Approach to the Statistical Mapping of Climate”. In: *Climate Research* 22.2 (2002), pp. 99–113.

[94] Robert Rice and Roger C. Bales. “Embedded-Sensor Network Design for Snow Cover Measurements around Snow Pillow and Snow Course Sites in the Sierra Nevada of California”. In: *Water Resources Research* 46 (2010), pp. 1–13.

[95] Coen J. Ritsema et al. “A new wireless underground network system for continuous monitoring of soil water contents”. In: *Water Resources Research* 46.4 (2010), pp. 1–9. ISSN: 00431397. DOI: 10.1029/2008WR007071.

[96] Christian Rolland. “Spatial and seasonal variations of air temperature lapse rates in alpine regions”. In: *Journal of Climate* 16.7 (2003), pp. 1032–1046. ISSN: 08948755. DOI: 10.1175/1520-0442(2003)016<1032:SASVOA>2.0.CO;2.

[97] Olga Semenova and Keith Beven. “Barriers to progress in distributed hydrological modelling”. In: *Hydrological Processes* 29.8 (2015), pp. 2074–2078. ISSN: 10991085. DOI: 10.1002/hyp.10434.

[98] Jaydip Sen. "A Survey on Wireless Sensor Network Security". In: *International Journal of Communication Networks and Information Security (IJCNIS)* 1.2 (2009), pp. 55–78.

[99] Eylon Shamir and Konstantine P. Georgakakos. "Distributed snow accumulation and ablation modeling in the American River basin". In: *Advances in Water Resources* 29.4 (2006), pp. 558–570. ISSN: 03091708. DOI: 10.1016/j.advwatres.2005.06.010.

[100] Jean Emmanuel Sicart et al. "A Sensitivity Study of Daytime Net Radiation during Snowmelt to Forest Canopy and Atmospheric Conditions". In: *Journal of Hydrometeorology* 5 (2004), pp. 774–784. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2004)005<0774:ASSODN>2.0.CO;2.

[101] S. Simoni et al. "Hydrologic Response of an Alpine Watershed : Application of a Meteorological Wireless Sensor Network to Understand Streamflow Generation". In: *Water Resources Research* 47.10 (2011), pp. 1–16.

[102] Matthew Sturm and Carl Benson. "Scales of spatial heterogeneity for perennial and seasonal snow layers". In: *Annals of Glaciology* 38 (2004), pp. 253–260. ISSN: 02603055. DOI: 10.3189/172756404781815112.

[103] Robert Szewczyk et al. "Lessons from a Sensor Network Expedition". In: *European Conference on Wireless Sensor Networks (EWSN)*. Vol. 2920. Berlin, Germany: Springer Berlin Heidelberg, 2004, pp. 307–322.

[104] A. Terzis et al. "Slip Surface Localization in Wireless Sensor Networks for Landslide Prediction". In: *International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE. Nashville, TN, USA, 2006, pp. 109–116.

[105] Jan W Van Wagtendonk and Jo Ann Fites-Kaufman. "Sierra Nevada Bioregion". In: *Fire in California's Bioregions* (1997), pp. 264–294.

[106] Thomas Watteyne. "Machine-to-machine (M2M) Communications, Architecture, Performance and Applications". In: ed. by Mischa Dohler and Carles Anton-Haro. Woodhead Publishing, 2015. Chap. Lower-power Wireless Mesh Networks for Machine-to-Machine Communications Using the IEEE802.15.4 Standard, pp. 63–79.

[107] Thomas Watteyne et al. "Industrial IEEE802.15.4e Networks: Performance and Trade-offs". In: *IEEE International Conference on Communications (ICC)*. London, UK, 2015.

[108] Thomas Watteyne et al. "Mitigating Multipath Fading Through Channel Hopping in Wireless Sensor Networks". In: *International Conference on Communications (ICC)*. IEEE. Cape Town, South Africa, 2010.

[109] Thomas Watteyne et al. "Technical Overview of SmartMesh IP". In: *International Workshop on Extending Seamlessly to the Internet of Things (esIoT)*. Taiwan, 2013.

[110] Mark A. Weissberger. *An Initial Critical Summary of Models for Predicting the Attenuation of Radio Waves by Trees*. Tech. rep. EST-TR-81-101. Annapolis, Maryland, USA: Department of Defense, Electromagnetic Compatibility Analysis Center, 1982.

[111] Stephen C. Welch et al. “Sensor placement strategies for snow water equivalent (SWE) estimation in the American River basin”. In: *Water Resources Research* 49.February (2013), pp. 891–903. ISSN: 00431397. DOI: 10.1002/wrcr.20100.

[112] Geoffrey Werner-Allen et al. “Deploying a Wireless Sensor Network on an Active Volcano”. In: *IEEE Internet Computing* 10.2 (2006), pp. 18–25.

[113] A. B. White et al. “A twenty-first-century california observing network for monitoring extreme weather events”. In: *Journal of Atmospheric and Oceanic Technology* 30.8 (2013), pp. 1585–1603. ISSN: 07390572. DOI: 10.1175/JTECH-D-12-00217.1.

[114] Allen B. White et al. “An automated brightband height detection algorithm for use with Doppler radar spectral moments”. In: *Journal of Atmospheric and Oceanic Technology* 19.5 (2002), pp. 687–697. ISSN: 07390572. DOI: 10.1175/1520-0426(2002)019<0687:AABHDA>2.0.CO;2.

[115] Allen B White et al. “Developing a Performance Measure for Snow-Level Forecasts.” In: *Journal of Hydrometeorology* 11.3 (2010), pp. 739–753. ISSN: 1525755X. DOI: 10.1175/2009JHM1181.1.

[116] Wiley-Blackwell, ed. *Antennas and Propagation for Wireless Communication Systems*. 2nd ed. Wiley-Blackwell, 2007.

[117] Eric F. Wood et al. “Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water”. In: *Water Resources Research* 47.5 (2011), pp. 1–10. ISSN: 00431397. DOI: 10.1029/2010WR010090.

[118] Xbee Xbee-pro and Digimesh Rf. “XBee/XBee-PRO DigiMesh 2.4 User Guide”. In: ().

[119] Hengchun Ye, Judah Cohen, and Michael Rawlins. “Discrimination of Solid from Liquid Precipitation over Northern Eurasia Using Surface Atmospheric Conditions*”. In: *Journal of Hydrometeorology* 14.4 (2013), pp. 1345–1355.

[120] Jennifer Yick, Biswanath Mukherjee, and Dipak Ghosal. “Wireless Sensor Network Survey”. In: *Elsevier Computer Networks* 52.12 (2008), pp. 2292–2330.

[121] Robert R Ziemer. “Evaporation and transpiration”. In: *Reviews of Geophysics* 17.6 (1979), pp. 1175–1186. ISSN: 1944-9208. DOI: 10.1029/RG017i006p01175.

[122] Roger C. Bales Martha Conklin Robert Rice Ziran Zhang Steven D. Glaser and Danny G. Marks. “Technical report: the design and evaluation of a basin-scale wireless sensor network for mountain hydrology”. In: *Water Resour. Res.* (2016).