2019

Modeling Agricultural Outcomes in a Warmer, Wetter Vermont

Rachel Mason
University of Vermont

Modeling Agricultural Outcomes in a Warmer, Wetter Vermont

A Thesis Presented

by

Rachel Mason

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Plant & Soil Science

May, 2019

Defense Date: March 1st, 2019
Thesis Examination Committee:

Scott Merrill, Ph.D., Advisor
Josef Gorres, Ph.D., Advisor
Joshua Faulkner, Ph.D.
Meredith Niles, Ph.D., Chairperson
Cynthia J. Forehand, Ph.D., Dean of Graduate College

Abstract

This thesis aimed to model agricultural outcomes that are important to Vermont dairy farms and surrounding communities – runoff, erosion, nitrogen and phosphorus losses, crop yields, and timeliness of farm operations – in a set of possible future climates. The Agricultural Policy/Environmental eXtender (APEX) model was used, and the models were calibrated with data from a project that measured most of these outcomes on a set of local farms. The modeling methodology is thoroughly documented and may be a useful starting point for others who are new to agricultural modeling.

Applied to two farms growing continuous corn, the future climate simulations showed that increasing temperatures by 2° C, combined with raising total precipitation or changing the seasonality of precipitation, had little effect on any outcome. Intense rainfall has increased greatly in recent decades, so a combination of higher temperatures and more intense precipitation was also simulated. This led to more runoff, more soil loss, and more nutrient losses. While median values were only modestly increased, the 95%-ile and total losses over the simulation period increased by a larger amount (as much as 53%, depending on the site). Management practices that can reduce runoff and soil/nutrient loss exist, but their effectiveness when a higher fraction of losses occur in large events is not well known.

Crop yields changed by <10% in all simulations, and in some cases increased slightly. Other studies have warned of decreases in yields because of high summer temperatures and droughts. The pilot simulations in this thesis probed only a limited range of climate parameter space, so running the models for a wider range of scenarios may illuminate the circumstances in which particularly harmful and beneficial outcomes occur.

Finally, APEX can in principle calculate the delays to corn planting that are expected if climate change leads to wetter conditions in the spring. However, the models consistently predicted that only harvest operations will be delayed. The reasons for this are not well understood, and it may be a useful avenue for future work.

The present work is limited in a number of ways. Chief among these are somewhat mediocre model performance, and the narrow range of farming systems and climates investigated. Statistics describing the performance of the calibrated models were poorer than anticipated, and satisfactory results could not be obtained for some nutrient loss pathways. Only two farms were modeled, in just four hypothetical climates; results for other relevant farming systems and climates may be quite different. Nonetheless, it is hoped that this thesis serves to illustrate the potential and limitations of utilizing APEX in this context, and that it lays the groundwork for a more extensive investigation of agricultural outcomes under climate change in Vermont.

Acknowledgements

Many people generously gave their time to answer my questions as I was starting to learn about agriculture in Vermont. They include: Juan Alvez, Brian Beckage, Gabriela Bucini, John Hayden, Heather Darby, Anne Lazor, Jack Lazor, Fred Magdoff, Bob Parsons, Ellie Searles, Mike Winchell, and Sara Ziegler. It makes me happy that people are so willing to share what they know. Specifically relating to this thesis, Dave Braun at Stone Environmental promptly responded to my numerous requests for data and clarification, and Luca Doro patiently dealt with my posts on the APEX user forum.

I'd also like to thank my advisors and committee for guidance and input. They are: Scott Merrill, Josef Görres, Meredith Niles, and Joshua Faulkner. In particular, I'm grateful to Scott and Josef for entertaining the possibility that someone could leave a perfectly good career in astronomy and start again from the beginning in agriculture. It's been quite the journey.

At the risk of this turning into an Oscar acceptance speech: I'll always be grateful that Mum, Dad, and Ossie have accepted and supported the unusual life decisions I've made. And of course, I wouldn't be where I am now without Inger's unwavering encouragement and support.

This project was supported by a research assistantship funded by the University of Vermont Agriculture Research Station Hatch grant program, and by the USDA National Institute of Food and Agriculture AFRI Foundational Program, NIFA Award #: 2015-67020-23180.

Table of Contents

List of Figures

List of Tables

1 Introduction

A very basic purpose of any food system is to provide the calories and nutrients that people require to stay alive. Beyond that, though, food systems are multifunctional, multifaceted, and have countless direct and indirect effects on people, animals, and the environment. Agriculture provides livelihoods for farmers, farm workers, and people employed elsewhere in the system. Choices made on the farm, yet strongly shaped by outside factors, affect whether agriculture maintains or erodes the natural resources on which it depends. Food can contribute to vitality, health, and thriving; or to sickness and premature death. The system, and its constituent parts, may or may not have the capacity to bear up under ongoing stresses and bounce back from sudden shocks.

This thesis will explore the effects on dairy farms of some of the stresses and shocks that climate change is likely to bring. In terms of production, dairy farming is Vermont's largest-scale entanglement with the food system. According to a 2015 report by the Vermont Dairy Promotion Council, there are almost 900 dairy farms in Vermont, located in every county in the state. Although they account for only 12% of Vermont's farms, more than 80% of the farmland (approximately 15% of the total area) in the state is devoted to pasture and feed crops for dairy farms, and dairy products make up roughly 70% of the state's agricultural sales by market value. The dairy industry supports numerous farm service and equipment providers, and it is estimated that it generates \$2.2 bn in economic activity in the state each year. Approximately 2.5% of the state's workforce is employed, directly or indirectly, by the dairy sector (Vermont Dairy Promotion Council, 2014).

Beyond their contribution to livelihoods, dairy farms in Vermont appear to be important to a wide cross-section of local society. In a survey of 271 Vermonters, >90% agreed that dairy farms add to the beauty of the state and that they contribute to Vermont's quality of life (Vermont Dairy Promotion Council, 2014). 84% believed that dairy is important to tourism in the state, consistent with the popularity of events and attractions like the Vermont Cheesemakers Festival and the Ben & Jerry's ice cream factory. Vermont's dairy farms are an integral part of the state's "iconic patchwork of forests and open fields" (Radel et al., 2010), and many Vermont cheeses are recognized as being of very high quality, winning numerous awards (Sakovitz-Dale, 2006).

Dairy farming may be an integral part of the history and economy of Vermont, but making a living as a dairy farmer is not easy. The story of dairy in the state, and in the US generally, in the last few decades has been one of continually increasing production and decreasing, but volatile, prices. Improvements in feed, genetics, technology, and management mean that in 2010 the average cow produced three times more milk than in 1965 (Parsons, 2010). This high supply rate has driven milk prices ever lower, to the point that today's prices are comparable to those in the late 1970s – without adjusting for inflation. Furthermore, exposure to global markets and geopolitical events means that milk prices fluctuate wildly (Charles, 2016; Macdonald et al., 2016), with swings of as much as 30% in a single month on record (Jesse and Cropp, 2008; Parsons, 2010).

At the same time, dairies are vulnerable to changes in the costs of inputs like grains, which are also highly variable (Macdonald et al., 2016). The price of milk regularly drops below the cost of production in Vermont and elsewhere, putting tremendous strain on farmers (Parsons, 2010; Kardashian, 2012). At the time of writing, milk prices are below the break-even point, and the local Agri-Mark co-operative has been making their farmers aware of mental health and suicide prevention services (Hudzik, 2018).

From the farmer's point of view, one potential solution to extremely narrow per-cow profit margins is to milk more cows, thereby providing more total income. With more revenue, the largest producers may also be better placed to weather economic storms. Indeed, as farm after farm has gone out of business in Vermont, the average farm size has greatly increased. In 1965 there were roughly 6000 dairy farms in Vermont, whereas now there are fewer than 900 (Winsten et al., 2010). In the same time period, the average herd size increased from $\sim$40 to $\sim$125 cows. Farmers' strategies are diverse, though, and a class of relatively thriving small farms has also emerged, often producing organic milk and/or "grass milk" that commands higher and historically more stable prices (Winsten et al., 2010).

Alongside the necessity of making a living, farmers must minimize negative effects on the land, air, and water around them. Dairy farms are an important part of Vermont's economy and landscape, but they also contribute to water pollution and other environmental problems. Nitrogen (N) and phosphorus (P), applied to agricultural fields as manure and fertilizer, are lost in water that runs off and through the land,

and attached to particles that are eroded from fields. In the watersheds surrounding Lake Champlain, these nutrients are transported to the lake, where phosphorus fuels the growth of cyanobacteria. As well as discouraging recreational use of the lake, toxic algal blooms have killed animals, caused human illnesses, and contaminated the source of the drinking water that is supplied to 145,000 people from the lake (Lake Champlain Basin Program, 2015). Agricultural activity is estimated to account for about 38% of the phosphorus in the lake (Lake Champlain Basin Program, 2015).

The existing nutrient loads in Lake Champlain are now high enough that reducing them to acceptable levels will be a very long-term project (Winslow, 2016). However, there are additional, immediate reasons for trying to prevent losses of agricultural inputs, soil, and water from farms. Agricultural runoff has impaired the ability of Vermont's streams and rivers to support aquatic life, and bacterial contamination means that some waterways are no longer suitable for recreational activities like fishing and boating (Vermont Department of Environmental Conservation, 2006). Recent severe storms, combined with prior wet weather, have led to episodes of extreme soil erosion on some farms (Yellen et al., 2016).

Issues such as these are likely to become more pressing as Vermont's climate changes. The region is expected to experience higher temperatures and increased rainfall, a combination that will have a slew of implications for agriculture (Frumhoff et al., 2007; Galford, 2014; Horton et al., 2014; Markowitz, 2017; Tobin et al., 2015). Intense rain can increase erosion and runoff, so one consequence may be that soil and nutrients are lost at a faster rate. At the same time, more frequent summer droughts are projected to increase the need for water-conserving practices. Physical losses of soil, water, and nutrients can translate to economic losses for farms, and farmers are also under significant pressure from lawmakers and clean water advocates to reduce their impact on the environment (Dolan, 2016).

Farmers can reduce the environmental impacts of their operations in many ways. Practices such as decreasing tillage in annual crop fields, using cover crops to avoid bare ground, replacing annual crops with hay or pasture, and not spreading manure on frozen or saturated ground, can all help to keep soil and nutrients on the farm. Some of these "best management practices" (BMPs) are now mandatory on Vermont farms, as part of the Required Agricultural Practices (RAPs) introduced by the Vermont State Legislature (General Assembly of the State of Vermont, 2015; VAAFM, 2016). Depending on the circumstances, these resource-saving practices can be a good economic deal for farmers, and they have the potential to aid in adapting to climate change as well (e.g. Kaye and Quemada, 2017). On the other hand, implementing BMPs can impose new economic stresses on farmers that are already under financial pressure. And it is not yet clear how effective the various BMPs will continue to be as Vermont's climate changes over the next few decades.

Climate change may affect aspects of dairy farms beyond their environmental impact. The unusually wet spring of 2017 prevented many farmers planting silage corn until well past the optimal dates (UVM Extension, 2017). Fine late summer weather allowed crop growth to catch up in some cases, but farmers expecting poor yields may not have invested in maintaining their crop, and purchased feed for their cows instead (J. Faulkner, personal communication). Wet weather can also present land management and cow health problems for farmers whose cows spend time on pasture (UVM Extension, 2017).

This kind of situation may become more common in a future climate with increased precipitation. Conversely, unusually dry summers like that of 2016 could reduce yields by subjecting crops to drought stress. Opportunities may also arise, such as an extended freeze-free season providing the occasion to relax restrictions on winter manure spreading, or a longer window for fall cover crop establishment. On the other hand, if frozen ground is replaced by saturated ground, winter manure spreading will remain undesirable. In short, interactions between climate, management practices, and operation schedules are likely to be complex and possibly counterintuitive.

Understanding the possible outcomes from a complex system like this requires numerical models. Hydrological watershed models can simulate crop growth and soil/nutrient transport processes under current conditions and for possible future climates. They can also be used to assess whether the timeliness of farm operations is likely to be affected by climate change, and whether that in turn affects crop yields and the implementation and effectiveness of best management practices. At the same time, the models are necessarily simplified representations of complicated natural processes. They must be informed by, and tested against, measurements from the field.

The aims of this thesis are to (1) investigate the utility of the Agricultural Policy and Environmental eXtender model (APEX; Gassman et al., 2010) for simulating farm operations, yields, and soil and nutrient losses from dairy farms in Vermont, and (2) perform a pilot study of a handful of climate scenarios on two farms growing continuous corn. Between 2012 and 2017, several farms around the state participated in a project that gathered runoff, water quality, and agronomic data from corn and hay fields on the farms (Braun et al., 2016). The measurements were originally used to examine the effectiveness of BMPs on the fields, but they also provide a valuable source of input and calibration data for the APEX model. Once calibrated using the field data, the models can then be used to simulate farm operations under a range of possible future climates.

The purpose of this work is to build a foundation for identifying and giving warning of future problems to farmers, service providers, policymakers and other interested parties. It is hoped that the methodology and lessons learned can be used to expand the modeling beyond this initial pilot project to examine a wider range of farming systems and practices – while illustrating the potential limitations of the modeling approach. The wider economic system will most likely continue to "hollow out the middle" of the Vermont dairy industry for the foreseeable future, driving towards fewer, larger farms alongside a handful of small, niche producers. Nonetheless, it may ultimately be possible to find some options that may benefit farms and their surroundings, and to give early warning of climate change effects that will be bad for both.

The remainder of this Introduction goes into more detail about dairy farming and climate change in Vermont, giving the background and context to the modeling presented in this thesis. §1.1 briefly outlines how the state's dairy farms work, with particular reference to the crops required to feed the herd. §1.2 then summarizes the effect of weather on the operations necessary to produce those crops. Next, §1.3 takes a closer look at the possible environmental impacts of dairy farming and the mechanisms behind them. Finally, §1.4 reviews the climatic changes that have been observed and are projected to take place in this region. The information in each section is used to derive the specific issues and scenarios that will be addressed.

At that point, §1.5 outlines how APEX models are typically constructed, and highlights research that argues for the importance of proper model calibration. The Agricultural Practice Modeling and Evaluation (APME) project (Braun et al., 2016), whose data will be used for parameterizing the models in this thesis, is also introduced. In addition, §1.5 briefly discusses issues of transparency and reproducibility in research and how they will be addressed in this thesis. An overview of the whole project is given in §1.6.

1.1 How Vermont's Dairy Farms Work

Dairy farming in the US takes many different forms. Commonly, milking cows are confined to barns and fed farm-grown crops and/or a commercial "total mixed ration" (TMR). High grain intake, genetic improvements, and economies of scale mean that per-cow production from these systems tends to be high, and herd sizes are large: in 2009, 30% of US milk cows were in herds of at least 2000 head, producing an average of 22,000 lbs of milk per cow each year (National Agricultural Statistics Service, 2010). Milk from this type of farm is generally sold to processors as a bulk commodity.

In contrast to the confinement model, some dairies still pasture their milking cows. Management-intensive grazing, in which short grazing and long rest periods are used to improve plant and animal productivity, is a fairly popular technique among these farms, which tend to be smaller than their confined-cow counterparts (Hanson et al., 2013). Most farms supplement their cows' pasture intake with home-grown and/or purchased grain-based feed, although a few do not. Per-cow milk production is usually lower than in confined cow systems, but pasture-based operations may still be able to profit from reduced production costs (Hanson et al., 2013; Winsten et al., 2000). They are often able to take advantage of the higher prices commanded by organic milk, which requires milking cows to spend time on pasture, and they may also produce cheese, yoghurt, and other value-added dairy products.

Vermont's dairy farms span a range of sizes and operating models and include elements of both of the above systems. Winsten et al. (2010) surveyed farms in the northeast US in 2006, and classified 10% of Vermont farms (39 respondents) as "large, modern, confinement" (LMC) facilities, and 20% (74 respondents) as "management intensive grazing" (MIG) farms. (In their scheme, an LMC farm uses a confinement-feeding system and contains at least 300 cows, while MIG is characterized by frequent pasture moves and cows obtaining most of their forage intake from pasture.) The remainder either had smaller confinement herds or were using less intensively-managed pasture. 83% of the certified organic farms in the study area were MIG systems.

Most of Vermont's dairy farms, even the LMC operations, grow at least some of the feed required by their cows. Corn is the preferred crop for many farmers, as silage made from corn provides the energy needed to support high levels of milk production. The corn may be grown in rotation with grass and legumes which are made into hay and haylage (hay crop silage). Alternatively, corn and hay crops may be grown continuously on the same piece of ground. In Vermont in 2016, hay and haylage were harvested from 310,000 acres and corn silage from 85,000 acres (USDA/NASS, 2016).

While silage corn and hay are the main feed crops grown by Vermont's dairy farmers, the number of acres planted to each, and the number of years each one occupies in the rotation, depend on many different factors. For example, a farmer may wish to grow as much corn as possible to maximize their milk production, or prefer to emphasize hay for environmental reasons. Corn may be productive on some soils, but less so on others. A farmer's options are also likely to be constrained by regulatory requirements, such as erosion prevention guidelines and the nutrient management plan that certain farms must follow. Considerations like these also influence the management methods adopted by farmers, such as the kind of tillage to use.

The farms that follow the MIG model also often grow or purchase feed derived from annual grain crops, but some do not. A minority of farms use permanent pasture in place of corn fields, feeding their cows almost entirely through grazing and hay production (e.g. Lazor, 2016). The herd grazes from late spring into autumn, and surplus spring pasture growth is cut and stored for winter feed.

Within this considerable variety of farming systems, this thesis will examine a single case: silage corn grown continuously on the same piece of land for an indefinite period. This system is chosen for two reasons. First, in the Missisquoi Bay Basin region of Vermont, it is estimated that 49% of cropland is used for continuous hay production, 41% for corn and hay in rotation, and only 10% for continuous corn (Winchell et al., 2011). However, continuous corn land appears to contribute disproportionately to phosphorus pollution (Winchell et al., 2011), so it is of interest for this study. In addition, farms with continuous corn production participated in the APME project that will be used to calibrate the APEX models (§2.1). Other considerations relating to data quality etc. are explained in §2.1.1.

The specific management practices and detailed operations schedules that will be modeled are discussed in later sections. For farms growing much of their own feed, corn yield is an important economic outcome. As a rough proxy for economic effects of climate change, then, one of the model outputs that will be examined is the annual forage yield.

1.2 Dairy Farming and the Weather

Farming, of course, is profoundly affected by the weather. Temperature and precipitation may influence crop yields and nutrient losses directly, through effects such as drought stress or soil erosion, and also indirectly, by affecting the timeliness of farm operations. The direct effects of weather and climate on the environmental impacts of agriculture are addressed in §1.3 and §1.4. To illustrate how the weather affects farm operations and crop yields, this section gives a brief overview of a year in the life of the farming systems listed in §1.1.

At the beginning of each year, state regulations prohibit the spreading of manure until April 1st (April 14th on frequently flooded soils; VAAFM, 2016). This is intended to minimize the likelihood of manure/nutrient runoff from frozen or saturated ground (§1.3). Farms' manure pits will be filling up after the long winter, so many farmers will want to spread manure soon after the ban has been lifted. Ideally, the manure will be incorporated into the soil, either via tillage (in conventionally tilled corn) or a lower-impact technique such as injection or aeration (in reduced- or no-till corn and on hay fields) very soon after spreading. This reduces the loss of valuable nitrogen in the form of gaseous ammonia, benefitting both the farmer and the environment.

If a cover crop has been established, it will be terminated before the season's corn crop is planted (e.g. Darby et al., 2012). Corn seeds require soil temperatures >10°C for germination, and soil that is moist but not too wet. At the same time, it is necessary to plant early enough in the spring for the crop to mature before temperatures start to cool again in the fall. The seedbed may be prepared using tillage, or the seed may be no-till planted directly into residue. Starter fertilizer is usually applied at the time the corn is planted, and pesticide and supplemental fertilizer may be applied in the following days or weeks.

The highest quality hay is obtained by cutting early in the season, and this will also allow the collection of several subsequent cuts. Manure spread after each cut of hay (and before the first spring cut, if conditions permit) provides a source of nutrients for the next harvest. Pastured milking cows go out in May, weather permitting, and return to the barn in October.

Corn growth benefits from warm temperatures and adequate (but not excessive) rainfall through the summer. After the corn harvest in mid-September – early October, many farmers will spread and incorporate manure again, to empty the pits for winter. If they intend to plant a cover crop, the cover must be seeded in time for it to establish well before it goes dormant for the winter; poor establishment will negate the benefits of using the cover crop (§1.3). The funding that the Natural Resource Conservation Service (NRCS) provides for cover crops is also contingent on meeting latest acceptable planting dates (NRCS, 2014).

All of these field operations - spreading, planting, fertilizing, haying, grazing, harvesting, and tillage - can cause harm if performed on soils that are excessively moist, and soils in Vermont are often wet in the spring. For example, wet soils can mean losses of fertilizer nitrogen through leaching and denitrification (§1.3), which is both an economic and environmental problem. Allowing cows to graze on wet soil can damage pastures and cause lameness and mastitis in the herd.

Operating heavy machinery and tilling wet ground can also lead to soil compaction (McKenzie, 2010; Wolkowski and Lowery, 2008). Compaction increases the strength of the soil, which reduces the ability of plant roots to explore for nutrients and water. Compacted soil also has less pore space and fewer aggregates, meaning it holds less air and water and is less conducive to water infiltration and percolation. This can cause increased denitrification rates, reduced biological activity (such as nutrient mineralization), and more surface ponding, runoff, and erosion. Taken together, these effects are likely to decrease yields and increase pollution problems.

Wet soils with high clay content are particularly prone to compaction because water forms thin, lubricating films around the clay particles (Winterkorn, 1959). While the relationship between soil moisture content and compaction susceptibility is complex, a common rule-of-thumb is to avoid tillage and other operations when the moisture content is roughly equal to field capacity, particularly on fine-textured soils (Al-Khaisi and Licht, 2005).

Clearly, wet weather can cause many practical problems (and in fact excess moisture accounts for ~60% of crop insurance claims in Vermont; RMA, 2012). In a wet spring, a farmer planting corn may face a choice between planting on time but risking poor germination and soil damage, and waiting for dry soil but risking that the corn will not have time to develop adequately by harvest time. If warm, dry summer weather allows the corn to continue developing into the fall, the farmer may need to choose between maximizing their yield and planting cover crops in time to establish before winter. A wet fall, of course, can interfere with harvest and planting plans as well.

APEX allows the user to perform operations according to a set schedule, or to specify a soil moisture threshold at which operations will be postponed until the soil has dried sufficiently. The corn simulations will therefore be run both ways: operations-as-scheduled, and allowing for the possibility of delays. Days of delay and final corn yields will be recorded in both scenarios. Corn yields are of course also directly affected by temperature and soil moisture. APEX calculates and tracks the temperature, water, and nutrient stresses that the growing crop experiences, and these outputs will be used to understand the reasons behind any changes in yields in the various climate scenarios that will be explored (§4.1.1).

1.3 Dairy Farming and the Environment

All agricultural systems displace "natural" ecosystems and interact, directly and indirectly and in positive and negative ways, with the local and global environment. Dairy farming can affect the environment in numerous ways, from greenhouse gas emission to production of harmful airborne particles. Vermont's most prominent agriculture-related environmental issues are to do with water quality: the state faces a legal requirement to reduce the phosphorus loading in Lake Champlain (Chapman and Duggan, 2016; US EPA, 2015), farmers are under pressure to improve water quality (Dolan, 2016), and local farmers themselves state a desire to prevent nutrient losses into waterways[1]. Therefore the environmental part of this project will focus on losses of soil and nutrients from Vermont's dairy farms. Table 1.1 puts these processes into the context of the wider suite of potential environmental problems associated with dairy farming.

[1] See for example www.champlainvalleyfarmercoalition.com/home.html

In the modern, global food system, biomass and nutrients are continually moved from one place to another. Phosphorus and nitrogen fertilizers are produced in locations with an abundance of minerals rich in P (Rabchevsky, 1997), or where natural gas is available for fixing atmospheric N via the Haber-Bosch process, and transported to areas of crop production. Crops are harvested for human or animal consumption, often far from the farm where they originated, then transformed into human, animal, and food waste that must be disposed of or otherwise dealt with. This distribution creates local excesses and deficiencies of nutrients that can have environmental consequences (Magdoff et al., 1997; Wironen et al., 2018).

In the dairy industry, material is imported to a farm in the form of feed and supplements, fertilizer, and bedding, and exported as milk and other animal products. Feed is processed through the cows and transformed into bovine biomass, milk, manure, and digestive gases. In the case of an operation with many animals and little or no cropland and/or pasture, the result will be a large buildup of nutrients that is likely to cause significant harm to the surrounding environment (Burkholder et al., 2007; Kellogg et al., 2000).

On an integrated crop-livestock farm that grows some of its own feed, like most in Vermont, animal manure will be applied to the crops (often along with some synthetic fertilizer; §1.2). Manure is a valuable source of nutrients and organic matter that may be deficient in the farm's soil, and also replaces material lost in the harvest of crops and milk. However, if the nutrients it contains end up being present in excess in the soil, or in locations or chemical forms that are vulnerable to loss, the manure can still be a source of pollution.

As noted above and in §1, the main nutrient of concern in Vermont is phosphorus.

*Table 1.1: Ways in which dairy farming can impact the environment. Entries in **bold** are those that are evaluated in this thesis.*

Area of Impact	Examples
Water, land, and energy use	Using water to clean milking areas Using fossil fuels to produce fertilizer and power machinery
Greenhouse gas emissions	CH_4 from ruminant digestion CH_4 and N_2O from manure storage and handling N_2O from denitrification of manure, fertilizer, crop residues, etc. CO_2 from general farm operations (e.g. use of fossil fuel, lime, urea)
Other airborne emissions	NH_3 volatilization from manure and fertilizer Particulate matter Odors
Land degradation	Compaction by machinery and livestock **Erosion of cropland** and streambanks Soil acidification through N addition, leaching, and harvest
Water pollution	**Contamination by nutrients and sediment**; also pesticides, pharmaceuticals, pathogens, etc.

P is generally present in manure at a lower N:P ratio than is required by plants. If manure is land applied to meet the nitrogen needs of plants (or simply at whatever rate is needed to dispose of the manure that builds up in the farm's manure pit) P tends to build up. This is a problem because the amount of dissolved P lost in runoff events is correlated with the amount of available P in the soil (Sharpley et al., 1977, 1994), and P sorbed to solids in the soil can also be lost via erosion and overland flow. Nitrogen pollution is also relevant: contamination of drinking water with agricultural nitrates is widespread in the US (Nolan and Ruddy, 2016), and has been the subject of recent reporting in Vermont (Corwin, 2018).

Dairy farms can reduce their impact on water quality by means such as reducing the quantity of material brought in from outside, improving the efficiency with which their animals utilize the nutrients (so less is excreted), using manure to produce biogas, and following a nutrient management plan that specifies appropriate amounts of manure to add to the land (Sharpley et al., 2004). Most relevant to this thesis, though, is what happens once the manure is on the fields. Farmers can implement best management practices that reduce the likelihood of N and P being lost from their fields, and also practices that trap material once it has been displaced.

There are several mechanisms that can transport N and P from farm fields, almost all of which are influenced by the movement of water through the system (Baker et al., 2006). Water falls on the fields as rain (and snow and irrigation water). If plants are present, some of the rainwater will be intercepted by the plant canopy, from where it can evaporate or drip to the ground. The remainder will flow along the soil surface, infiltrate into the soil, and/or pool in surface depressions from where it can either infiltrate or evaporate. Water that falls onto and/or moves along the surface can detach and transport mineral and organic soil particles as well as surface-applied material such as broadcast manure and inorganic fertilizer. Runoff water can also dissolve and transport material from a thin (∼1 cm) mixing zone at the soil surface.

Water that infiltrates into the soil can be stored in soil pores or move horizontally or vertically through the soil profile. Some of the stored water is available to be transpired by plants. Water flowing laterally may re-emerge further down the slope and become surface runoff, or merge with a stream or other surface waterway. Water moving vertically may be driven upwards by evaporative demand at the surface, end up in artificial subsurface drainage ("tile drainage"), or percolate downwards into groundwater. Soil water can carry nutrients and other substances with it, leading to leaching losses of soluble materials.

In summary, nutrients can be lost through direct transport of surface-applied materials; adsorbed to eroded mineral soil particles; as part of eroded soil organic matter; dissolved in runoff water; and dissolved in water that percolates to groundwater and tile drains. P is found in all of these forms: some is found in the soil solution, a larger fraction is fixed to colloids and minerals, it is a constituent of organic matter, and may be surface applied in manure and inorganic fertilizer (Hansen et al., 2002). On conventionally tilled land, erosion and therefore particulate P losses tend to be large (Sharpley et al., 1994). Runoff P losses from grassland and forests tend to be dominated instead by dissolved P. As P is usually strongly fixed by soil colloids, leaching of dissolved P to groundwater is low in most circumstances. However, soils with high levels of dissolved P have shown significant losses in leachate and subsurface drainage (Hansen et al., 2002; VAAFM and VANR, 2017).

The nitrogen cycle is complex and N is found in many forms. Organic N makes up roughly half of the N in liquid dairy manure (Jokela et al., 2004) and is also found in soil organic matter. Soil microorganisms are responsible for mineralizing organic N into plant-available, inorganic forms. The main form of inorganic N in manure (and some fertilizers) is urea (CH_4N_2O), which is rapidly hydrolyzed to ammonium, NH_4^+. If this takes place on the soil surface, N is quickly lost as gaseous NH_3. Within the soil, NH_4^+ is relatively immobile, held on cation exchange sites on clay and organic matter particles.

Nitrifying bacteria transform NH_4^+ into nitrite, NO_2^-, and nitrate, NO_3^-. Nitrate ions are highly susceptible to leaching, and indeed agricultural nitrate leaching is widely implicated in the eutrophication of coastal waters (CERN, 2000; Howarth et al., 1996). In wet, anaerobic soil conditions, denitrifying bacteria reduce nitrates into gaseous N_2O (a potent greenhouse gas), N_2, and various intermediates. The existence of these numerous potential loss pathways means that N can be hard to control (e.g. Mkhabela et al., 2008).

Many factors affect the flow of water and nutrients through agricultural systems. Some of them are fairly fixed; for example, the relative fractions of sand, silt and clay, or the presence of shallow bedrock, exert a strong effect on soil water movement. Other factors, however, can be influenced by farm management. To a large extent this is achieved by influencing the movement of water over and through the soil, and maintaining or improving the structure of the soil itself.

Preserving a cover of living and/or dead plant material reduces the impact force of raindrops on the soil and slows the flow of water over the surface, decreasing the amount of material eroded. Stems and residue can also impede the movement of sediment and runoff water. Slower-moving water, or water that pools within residue cover, has more opportunity to infiltrate into the soil instead of running off, and sediment is more likely to settle out. Well-aggregated soil with many pores and root channels also encourages infiltration and drainage, and is less susceptible to erosion in the first place.

The placement and timing of nutrient applications also affect losses. Manure on the soil surface is vulnerable to NH_3 volatilization and erosion/runoff losses (although manure within the soil may be susceptible to denitrification losses, instead; Duncan et al., 2017). Over time, and if the ground is not frozen or snow-covered, the P in fertilizer and manure becomes fixed to soil particles. If rainfall causes runoff before that happens, then losses of P can be large. In fact, most P loss generally occurs during just a few large rain events each year (Sharpley et al., 1994).

The sensitivity of runoff and sediment/nutrient losses to rainfall volume, intensity, and timing implies that Vermont's environmental problems could be altered – and quite likely exacerbated – by climate change. The ability of best management practices to reduce runoff, erosion, and nutrient loads may also be affected (e.g. Chiang et al., 2012; Jayakody et al., 2014; Woznicki and Nejadhashemi, 2012). Globally, the effects of climate change on erosion, runoff, and nutrient loss depend on a complex set of interactions between temperature, rainfall patterns, plant biomass production, microbial activity, human land use decisions, etc. However, Nearing et al. (2004) suggest that in general erosion will increase by roughly 1.7 times the amount of annual rainfall increase. Closer to home, Marshall and Randhir (2008) predict that sediment loading and winter and spring runoff in the Connecticut River Watershed will increase, leading to greater eutrophication potential. Across the U.S. Northeast, Hayhoe et al. (2007) find that annual total runoff will increase, driven by higher winter runoff volume and earlier peak flows.

In principle, APEX contains the machinery to simulate all of the above soil/water/nutrient processes in arbitrary climates. The model tracks interactions between components including precipitation, temperature, crop growth, and nutrient transformations and translocations, and reports information about all of these items. APEX can also simulate numerous practices intended to reduce the environmental impacts of agriculture.

Several of those practices – reduced tillage, cover cropping, grassed waterways, strip cropping, and others – were implemented on the farms that took part in the APME project. The initial work in this thesis, however, deals with relatively simple cases: two farm sites that broadcast and sometimes incorporated manure, practiced some tillage, and had relatively little soil cover through the winter. APEX's predictions for runoff, sediment, and nutrient losses (as well as crop yields) are calibrated using data from these farms, and the model is then used to make predictions for how these quantities would have changed in different climates. These different climates are described in the following section.

1.4 Vermont's Changing Climate

Increasing atmospheric concentrations of CO_2 and other heat-trapping gases have caused a rise in temperatures and changes in weather patterns around the world (IPCC, 2014) to which global agriculture will have to adapt (e.g. Howden et al., 2007; Smit and Skinner, 2002). Climatic changes in the US Northeast, and in Vermont in particular, have been described in a series of reports (Betts, 2017; Frumhoff et al., 2007; Galford, 2014; Horton et al., 2014; Tobin et al., 2015) based on numerous individual studies. This section takes a closer look at the literature concerning historical and projected temperature and precipitation in the region, with the aim of distilling a small set of plausible and/or interesting possible future scenarios to be used as input to APEX.

1.4.1 Temperature

In Vermont, annual mean temperatures increased by 0.19°C decade^{-1} between 1958 and 2012, with nine of the 10 warmest years on record occurring since 1990 (Guilbert et al., 2014). Lakes are freezing later and thawing earlier, leaves and flowers are appearing earlier, and the frost-free period has increased by about 14 days (Betts, 2011). In the Northeast generally, warming has been more pronounced in winter and spring than in the rest of the year, and the rate of temperature increase has accelerated in the past three decades (Hayhoe et al., 2007; Kunkel et al., 2013b).

These trends are forecast to continue. Climate models are able to reproduce historical temperature data in the Northeast region fairly well, and consistently predict ongoing increases (Guilbert et al., 2014; Hayhoe et al., 2007; Kunkel et al., 2013b). Both Hayhoe et al. (2007) and Guilbert et al. (2014) find that annual temperatures will rise by 2 – 3°C by mid-century. The Guilbert models predict slightly more warming in winter, while the reverse is true for the Hayhoe models. In the Lake Champlain basin, the number of days above 32.2°C is expected to increase from 6 to 24 days yr^{-1} by midcentury, and to 37 days yr^{-1} by the end of the century. The growing season may lengthen by as much as 43 days in that time (Guilbert et al., 2014).

1.4.2 Precipitation

Precipitation patterns in the region have also changed in recent decades. There is broad agreement that the Northeast has been experiencing more total and heavy precipitation (Kunkel et al., 2013a,b; Walsh et al., 2014), which has led to elevated groundwater levels (Weider and Boutt, 2010) and a rise in flood magnitudes (Collins, 2009). However, the characteristics – seasonality, etc. – of the increase in extreme precipitation, and expectations for future behavior, are less clear. There follows, then, a brief review of the literature that addresses these issues. Table 1.2 attempts to synthesize a concise overview from the various different indicators used in the literature, somewhat subjectively dividing results into "primary/significant" and "secondary/non-significant".

Table 1.2: Overview of **observed** *changes in precipitation in the US Northeast, showing "significant/primary" trends in dark gray, and "secondary/non-significant" results in light gray. Both annual and seasonal results are shown where available. See text for full details.*

<table>
<tr><th>Quantity, Reference[a]</th><th>Years</th><th>Jan</th><th>Feb</th><th>Mar</th><th>Apr</th><th>May</th><th>Jun</th><th>Jul</th><th>Aug</th><th>Sep</th><th>Oct</th><th>Nov</th><th>Dec</th></tr>
<tr><td>Total</td><td></td><td colspan="12"></td></tr>
<tr><td>H17</td><td>'79-'14</td><td colspan="12">+4% decade^{-1} (+13% since 2002)</td></tr>
<tr><td></td><td></td><td colspan="5">No significant change</td><td colspan="3">+6.5% decade^{-1}</td><td colspan="3">+2.5% decade^{-1}</td><td></td></tr>
<tr><td>B08[b]</td><td>'63-'03</td><td colspan="3">No significant change</td><td colspan="6">+8%</td><td colspan="3"></td></tr>
<tr><td>Median</td><td></td><td colspan="12"></td></tr>
<tr><td>G15</td><td>N/A[c]</td><td colspan="12">No significant overall trend</td></tr>
<tr><td>Extreme, total</td><td></td><td colspan="12"></td></tr>
<tr><td>H16[d]</td><td>'79-'13</td><td colspan="12">+8-10% decade^{-1}</td></tr>
<tr><td></td><td></td><td colspan="4">+2-4% decade^{-1}</td><td colspan="6">+8-10% decade^{-1}</td><td colspan="2"></td></tr>
<tr><td>H17[d]</td><td>'79-'14</td><td colspan="12">+19% decade^{-1} (+53% since 1996)</td></tr>
<tr><td></td><td></td><td colspan="2">+45% since 1996[e]</td><td colspan="3">+83% since 1996[e]</td><td colspan="3">+27% since 1996[e]</td><td colspan="4">+85% since 1996[e]</td></tr>
<tr><td>Extreme, frequency</td><td></td><td colspan="12"></td></tr>
<tr><td>H16</td><td>'79-'13</td><td colspan="12">+6-8% decade^{-1}</td></tr>
<tr><td>F15[d]</td><td>'01-'12</td><td colspan="12">+15-80% after 2000</td></tr>
<tr><td></td><td></td><td colspan="5">No significant change</td><td colspan="5">"Robust and significant" increase</td><td colspan="2"></td></tr>
<tr><td>G15</td><td>N/A</td><td colspan="5">95%-ile ↑ at $\geq 2/3$ of stations</td><td colspan="4">95%-ile ↑ at $< 2/3$ of stations</td><td colspan="3"></td></tr>
<tr><td>Extreme, magnitude</td><td></td><td colspan="12"></td></tr>
<tr><td>H16</td><td>'79-'13</td><td colspan="12">+1-2% decade^{-1}</td></tr>
<tr><td>F15</td><td>'01-'12</td><td colspan="12">No significant change</td></tr>
</table>

[a] References: B08 – Beckage et al. (2008). F15 – Frei et al. (2015). G15 – Guilbert et al. (2015). H16 – Hoerling et al. (2016). H17 – Huang et al. (2017).

[b] For Vermont only (Burlington International Airport weather station)

[c] Guilbert et al. (2015) calculated the median and 95%-ile of precipitation in a 30-year moving window.

[d] These authors define an “extreme” precipitation percentile (Hoerling: 95%-ile; Frei: 90, 95, 99%-ile; Huang: 95%-ile) and calculate its value for some baseline period. For various periods of interest, they report the total precipitation above this threshold, frequency of events above this threshold, and/or magnitude of events above the threshold.

[e] Huang et al. (2017) report large increases across the 1996 changepoint, but also state that winter (December - February) is the only season which the overall temporal trend is statistically significant.

Using data from Global Historical Climate Network (GHCN) stations, Hoerling et al. (2016) find a 2%-3% decade^{-1} increase in the total amount of heavy (95%-ile) precipitation over the 1901-2013 period in the Northeast. The magnitude of this increase has risen to $\sim$10% decade^{-1} since 1979. This has primarily been a warm season (May – October) phenomenon – the authors point out that "10 of the 12 summers since 2002 experienc[ed] anomalously high rainfall contributions related to extreme daily events". While they find a small (2.5% decade^{-1}) increase in the *magnitude* of extreme events, the rise in the total amount of extreme precipitation has mainly been driven by a higher *frequency* of events.

In rough agreement with Hoerling et al. (2016), Frei et al. (2015) also find that the frequency of warm season (June – October) heavy precipitation has increased in the Northeast. This phenomenon has been particularly pronounced since 2000. Frei et al. do not detect statistically significant increases in cold season extreme event frequency, or in the magnitude of extreme events at any time of year.

Huang et al. (2017) examine total annual precipitation and the total amount of extreme (95%-ile) precipitation in this region. While they observe a 13% increase in total precipitation since 2002, they find that extreme precipitation has increased by a much larger amount, by 53% since 1996. They emphasize the existence of these "changepoints" in 1996 and 2002, as well as the fact that temporal trends in precipitation are very sensitive to the choice of time period used for the analysis. Huang et al. (2017) find that the largest increase in total precipitation has taken place in the summer (+6.5% decade^{-1} in June - August). However, in contrast to Hoerling et al. (2016) and Frei et al. (2015), they find the largest increases in extreme precipitation to have occurred outside the summer months.

Guilbert et al. (2015) also find evidence of increased heavy precipitation outside the summer months. Unlike the studies discussed above, which calculate the 95%-ile of precipitation over some time baseline, and then evaluate precipitation amounts and frequencies above that threshold, Guilbert et al. (2015) calculate the 95%-ile in a moving 30-year window and evaluate how that quantile changes over time. They find that, for the 222 Northeast weather stations in their analysis, more than 2/3 show a positive trend in the daily 95%-ile from October through May. In July and September, fewer than 1/2 of the stations show such a trend. The strongest signal, 0.7 mm d^{-1} $decade^{-1}$, occurs in April, when 148 stations show a significant positive trend, and only 20 a significant negative trend.

As well as extreme events, Guilbert et al. (2015) also examine median daily precipitation and wet and dry persistence (the likelihood of a wet day following a wet day, or a dry day following a dry day). They detect little change in the median and in dry persistence. However, wet persistence increases throughout the year, and most consistently in May and June. Heavy rains in April followed by persistent wetness in May and June is a combination that could have particularly important implications for agriculture.

In summary, studies of historical climate data have shown that: (1) robust, statistically significant increases in temperature and growing season have already occurred; (2) total precipitation in has increased, while median precipitation has stayed fairly constant; (3) heavy/extreme precipitation has increased by more than total precipitation; (4) the increase in total extreme precipitation has been driven more by a higher frequency of extreme events than by a higher magnitude of individual extreme events; and (5) rates of change in all precipitation statistics appear to have increased in the last 2 – 3 decades. However, different analyses have led to different conclusions about the seasonality of extreme precipitation – some analyses have found it to be a warm season phenomenon, others the opposite. It is beyond the scope of this work to track down the origin of the conflicting claims about the seasonality of precipitation, which are likely related to factors such as the weather data sets included in the analysis and the sensitivity of trends to the time periods used for the baseline and comparison statistics (Huang et al., 2017; Keim et al., 2005).

To complete the picture of Vermont's changing climate, I now review model projections of the region's precipitation in decades to come. These simulations suggest that both total precipitation and extreme events will continue to increase, particularly in the winter (Table 1.3).

Table 1.3: Overview of **projected** *changes in precipitation in the US Northeast, showing significant/primary trends in dark gray, and secondary/insignificant results in light gray.*

<table>
<tr><th>Quantity,
Reference</th><th>Years</th><th>Jan</th><th>Feb</th><th>Mar</th><th>Apr</th><th>May</th><th>Jun</th><th>Jul</th><th>Aug</th><th>Sep</th><th>Oct</th><th>Nov</th><th>Dec</th></tr>
<tr><td colspan="14">MID-CENTURY</td></tr>
<tr><td>Total: Daily</td><td colspan="13"></td></tr>
<tr><td>Guilbert et al. (2014)[a]</td><td>2040-2069</td><td colspan="12">+7.1%</td></tr>
<tr><td>Total: Annual</td><td colspan="13"></td></tr>
<tr><td>Hayhoe et al. (2007)[b]</td><td>2035-2064</td><td colspan="12">+5 – 8%[c]</td></tr>
<tr><td></td><td></td><td colspan="2">+6 – 16%[c]</td><td></td><td></td><td></td><td colspan="3">-1 – +3%[c]</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Extreme: 95%-ile</td><td colspan="13"></td></tr>
<tr><td>Guilbert et al. (2014)</td><td>2040-2069</td><td colspan="12">+8.9%</td></tr>
<tr><td>Extreme: 99%-ile</td><td colspan="13"></td></tr>
<tr><td>Guilbert et al. (2014)</td><td>2040-2069</td><td colspan="12">+11.9%</td></tr>
<tr><td colspan="14">LATE CENTURY</td></tr>
<tr><td>Total: Daily</td><td colspan="13"></td></tr>
<tr><td>Guilbert et al. (2014)</td><td>2070-2099</td><td colspan="12">+9.9%</td></tr>
<tr><td>Total: Annual</td><td colspan="13"></td></tr>
<tr><td>Hayhoe et al. (2007)</td><td>2070-2099</td><td colspan="12">+7 – 14%[c]</td></tr>
<tr><td></td><td></td><td colspan="2">+12 – 30%[c]</td><td></td><td></td><td></td><td colspan="3">-2 – +0%[c]</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Extreme: 95%-ile</td><td colspan="13"></td></tr>
<tr><td>Guilbert et al. (2014)</td><td>2070-2099</td><td colspan="12">+12.5%</td></tr>
<tr><td>Extreme: 99%-ile</td><td colspan="13"></td></tr>
<tr><td>Guilbert et al. (2014)</td><td>2070-2099</td><td colspan="12">+16.7%</td></tr>
</table>

[a] Relative to 1961-2000 baseline
[b] Relative to 1961-1990 baseline
[c] Generally identified as significant only for higher emissions scenarios

For the US Northeast, Hayhoe et al. (2007) evaluate nine atmosphere-ocean general circulation models fed by three emissions scenarios. They find that the models are able to match historical temperature patterns quite well. However, they result in a wide range of precipitation trends which do not all have the same sign as the observed trends. Nonetheless, a common feature of the models is that they predict increasing total winter precipitation and no change or a decrease in summer precipitation (Table 1.3). More recent analyses confirm that climate models point to significantly wetter winters, while changes in summer are model-dependent and within the range of natural variability (Fan et al., 2015; Rawlins et al., 2012).

Hayhoe et al. (2007) find that the projected precipitation changes are more pronounced in the north of the region. Combined with warmer temperatures, more winter rain and the same or less summer rain would imply wetter soils and more runoff in winter/early spring, drier soils and less runoff in late spring/summer, and more short-term droughts. The lengthening growing season could further reduce soil water through increased evapotranspiration.

Guilbert et al. (2014) develop projections specifically for the Lake Champlain basin, using two representative concentration pathways (RCPs) and four general circulation models intended to represent particularly warm, cool, wet and dry conditions. They find that the models result in temperature and precipitation changes that are qualitatively consistent with the existing trends in the historical record. According to the models, mean daily precipitation will increase by 7.1% (9.9%) by 2040-2069 (2070-2099), the 95%-ile by 8.9% (12.5%), and the 99%-ile by 11.9% (16.7%). As they point out, "the 99th percentile of daily precipitation is projected to increase by 3.4 cm, which is equivalent to an additional 34,000 m^3 of liquid precipitation per square kilometer". Despite the higher mean daily rainfall, continued increases in temperature lead the authors to predict that the ratio of precipitation to potential evapotranspiration will tend to decrease between April and September.

At this point, we know that both the historical data and the models suggest that Vermont will be a wetter place in the future. However, although more and heavier rain is certainly to be expected from a warming atmosphere, it is not clear that the increased extreme precipitation in the US Northeast in recent decades is the result of human-induced climate change. Guilbert et al. (2014) note that the projected future increases in precipitation are much smaller than historical trends, and Hoerling et al. (2016) find that most of the observed increase so far can in fact be explained by decade-scale variability in ocean temperatures. It is possible that the expected signal of climate change is currently masked by other sources of variation.

At the same time, the model results are also uncertain. They are unable to accurately reproduce precipitation changes that have already been observed. In the case of future summer precipitation, different models give different results. They consistently predict more total winter rainfall, but the observed increases in total rainfall have so far been more of a warm season phenomenon (Table 1.2). In these circumstances deciding on future climates to use as input for agricultural models is not a trivial exercise.

Those issues, and the final selection of climates to be used in this thesis, are discussed in detail in Ch. 4. Briefly, the following broad scenarios will be simulated:

1980 – 2009: Hypothetical scenarios are defined relative to this baseline period.

Warmer: Temperatures increased by 2°C year-round.

Wetter: Temperatures increased by 2°C and total precipitation uniformly increased by 20%

Wet Spring: Temperatures increased by 2°C and 25% of June, July, August precipitation shifted to March, April, May.

Intense Rain: Temperatures increased by 2°C and the 95%-ile of precipitation increased by 30%, with total precipitation held constant.

1.5 Simulating Vermont Farms

To evaluate the effect of climate on agricultural outcomes, we would ideally set up a dense and extensive network of monitoring stations to collect data over many years of farm operations. Valuable edge-of-field studies have indeed been carried out on dairy farms (e.g. Bishop et al., 2005; Braun et al., 2016; Gilker, 2005; Kleinman et al., 2009), but because of the realities of funding, staffing, weather, technology, etc., they tend to be short in duration and concentrated at agricultural research sites and a fairly small number of commercial farms. Computational models that calculate flows of water, sediment, and nutrients into and out of the soil are a cost-effective alternative that can in principle be deployed at any location and over any timescale. This project will employ the widely-used Agricultural Policy/Environmental eXtender (APEX; Gassman et al., 2010) model. This section outlines some considerations about model calibration and accuracy that will constrain the farming and climate change scenarios that can feasibly be simulated. APEX is described in more detail in §2.

APEX has several hundred parameters that can be set by the user. These parameters define the physical and chemical characteristics of the site (e.g. slope, soil properties, weather), management information, and the methods, rates and thresholds used in calculating various processes (Baffaut et al., 2017). Management and weather records, soil test results, and other information is often available to set or constrain some of the parameters. Others may require adjusting until satisfactory outputs are obtained, a process for which several strategies are available (Daggupati et al., 2015; Wang et al., 2012).

If suitable comparison data exist, the user can refine the model parameters until they produce an acceptable match to observed amounts of runoff, erosion, nutrient loss, etc. from the area of interest. Typically, the model is calibrated using a subset of these measurements, then validated using the remainder. When few or no data are available, an alternative is to use judgment and experience to come up with a set of parameters that give results that are consistent with broad expectations, such as county-level historical crop yields and basic erosion calculations (Daggupati et al., 2015; Williams et al., 2010).

APEX, the EPIC model upon which it is built (Williams et al., 1989), and the very similar SWAT model (Arnold et al., 1998), have been used in numerous situations in which detailed empirical data for calibration and validation are not available. The most extensive of these is probably the Conservation Effects Assessment Project (CEAP; Duriancik et al., 2008; Johnson et al., 2015), in which APEX and SWAT have been used for a nationwide study of the effects of implementing agricultural best management practices. For that project, a sensitivity analysis was used to narrow down the set of influential parameters that need to be optimized (Wang et al., 2006). General model verification was accomplished by comparing outputs against a number of criteria such as whether increasing tillage intensity was correlated with increased erosion, and whether the results were consistent with the NRCS Conservation Practice Physical Effects matrix (Williams et al., 2010). Intermediate levels of calibration include techniques such as calibrating and validating models using data from one location, and applying them to another location that is judged to be sufficiently similar (e.g. Daggupati et al., 2015; Gassman et al., 2006).

Recently, some authors have investigated the effect of different calibration methods on the accuracy of APEX results. Baffaut et al. (2017) compared the output obtained for 12 poorly-drained, midwest soils when APEX was (a) calibrated using edge-of-field data and (b) set up using "best professional judgment" (BPJ) without reference to runoff, sediment, and P measurements. They found that the BPJ model gave acceptable predictions for runoff, although better results were achieved with the calibrated model. However, only the calibrated model was able to satisfactorily simulate sediment and total phosphorus yields. Baffaut et al. argue that calibration of APEX with water quality data is necessary for acceptable results.

Ramirez-Avila et al. (2017) performed a similar analysis for row crop and pasture fields in the southern US, and found that uncalibrated APEX models did not give useful results. In their study, calibrated models performed well for runoff, but did not satisfactorily simulate sediment or dissolved and total P losses. It appears that the performance of APEX depends on the context in which it is being used. However, no systematic analysis relating model accuracy to specific circumstances yet exists. It therefore seems highly advisable to calibrate APEX with respect to real-world data before relying on its output.

Assuming that water quality data are available and APEX can be calibrated so as to produce sufficiently accurate results, the next question is to what extent the model can simulate conditions – soils, management, climate etc. – that differ from those used for calibration (which, obviously, is the point of modeling). In the case of management, Bhandari et al. (2016) assessed the accuracy of APEX results for runoff, sediment, and total and dissolved P (TP, DP) when the model was calibrated for one management practice and then used to simulate another. They took advantage of two field sites: one containing plots managed with and without tillage, and with various fertilizer application methods, and another that compared different fertilizer types and rates. When calibrated for specific tillage and fertilizer practices, model validation results were generally good for runoff, TP, and DP (less so for sediment because of generally low sediment yields during the experiment).

When calibrated for one management and applied to another, results for runoff continued to be satisfactory. However, predictions of TP and DP were poor, especially for larger changes in management (e.g. tillage vs no-till). Bhandari et al. recommend that either models be applied only to the management types they are calibrated for, or that data from multiple managements be included in the calibration.

As far as climate is concerned, Moriasi et al. (2016) emphasize the importance of including wet, dry, and average years in the data set used to calibrate APEX. Illustrating this, Wang et al. (2014) note that APEX gave poor results for a very dry validation period when calibrated using data from a wetter period. This is consistent with the finding of Vaze et al. (2010) that rainfall-runoff models calibrated in drier conditions are better at predicting runoff in wetter conditions than the reverse. For hydrologic and water quality models in general, Daggupati et al. (2015) advise a "differential split-sample" strategy for simulations involving climate change. To simulate a wetter climate, Daggupati et al. state that the model should be calibrated using data from a relatively dry period and validated during a wet period to verify that it could simulate the transition from dry to wet conditions (and vice versa).

In simulating Vermont farms, then, it is clear that access to relevant water quality data will give the best chance of producing reliable results. Fortunately, measurements of runoff, sediment, N, and P for this study are available from a monitoring project carried out at seven Vermont farms by a local environmental consulting firm, Stone Environmental (Braun et al., 2016). The Agricultural Practice and Monitoring (APME) study began in 2012 and gives access to 2 – 4 years' worth of agronomic records and edge-of-field monitoring data (depending on the site). The data were not collected with the intent of being used for model calibration, which will present some challenges for this study. Nonetheless, they are a valuable resource. More details about the APME project are given in Chapter 2.

A large number of studies have made use of APEX (e.g. Gassman et al., 2010), including some limited use in Vermont (Stone, 2015; Winchell et al., 2011). While these works have made valuable contributions to our understanding of the environmental effects of agricultural practices, the procedures used to set up and calibrate the models are not usually well documented (although see Baffaut et al. 2017 for a good counterexample). Calibrating APEX can be a complex and somewhat subjective process and, while some useful general guidelines are available (Daggupati et al., 2015; Wang et al., 2012), no standard, detailed procedure exists. In the light of growing concerns about transparency and reproducibility in science in general (e.g. Stokstad, 2018), this is not ideal.

This thesis aims to give a thorough and detailed account of the APEX setup and calibration process used in this work, documenting and explaining decisions taken at each step. APEX input and output files will be made available, and the ancillary

Python code and supporting documents created for this project may be requested from the author. It is hoped that this will prove useful for others who are new to APEX and may benefit from examples that can be adapted and improved upon for their own research.

1.6 Summary of this Study

In summary, this thesis will create hydrological models for two Vermont dairy farms that grow silage corn every year. These models will be run under five different climate scenarios, with and without the possibility of farm operations being delayed by weather. The models will be used to predict runoff, erosion, nutrient losses, crop yields, and scheduling issues.

Chapter 2 describes the creation of "baseline" APEX models for the two farms. These models are based on agronomic, management, and weather data collected at the site. Where no site-specific information is available, all parameters are left at their default values. In Chapter 3, the parameters of the baseline models are adjusted in order to produce outputs that better match the APME edge-of-field data. Climate scenarios are fleshed out in Chapter 4, and many runs of the calibrated APEX models are carried out for each climate, to characterize the distribution of outcomes. The results of that exercise, and lessons learned along the way, are summarized in Chapter 5.

2 Initial APEX Model Setup

This thesis uses the APEX 1501 model to simulate nutrient flows and crop growth on Vermont dairy farms. APEX was developed by the USDA Agricultural Research Service and Texas A&M University to address water quality and other environmental problems (Gassman et al., 2010). Essentially, the model takes information about weather, site, and soil, performs user-specified farm operations, and calculates how water and nutrients move and how various soil and crop properties change as a function of time and management. It is a highly flexible tool that can simulate the effects of different cropping systems, conservation and nutrient management practices, etc. Depending on the inputs and operations defined by the user, APEX calculations and outputs can include, for example:

- The amount of soil that is eroded by water and wind
- How nitrogen, phosphorus, and carbon are transformed and transported within the soil
- The amount of nutrients exiting the simulation area in different forms
- The amount of pesticide exiting the simulation area
- The crop yield that is achieved each year

- The cost of a year's operations

Simulations can be composed of several distinct, connected sites – "subareas" – in which case APEX tracks the flow of water, sediment, nutrients, and chemicals from one subarea to the next.

APEX uses several hundred parameters to describe everything from how efficiently a plant uses incident radiation to microbial decomposition rates. Some of them define the system that is being simulated: the crops that are being grown, dates of operations such as tillage and harvest, soil properties, weather, etc. Others must be set to values that provide acceptable model outputs for the system in question, and the remainder are most likely left at their default values. Assessing which of the many parameters are relevant in any given simulation, and selecting appropriate values for them, requires a number of decisions to be made. While numerical criteria and software tools are available to assist in this process, some of these decisions are ultimately based on the modeler's judgment and the experience of others reported in the literature.

In the interest of transparency and reproducibility, this chapter and the next describe in some detail the steps taken in creating APEX simulations for two small watersheds on Vermont dairy farms: the initial setup process (selecting model parameters and processes to reflect the site conditions and management at each farm) and, in Chapter §3, the calibration procedure (tuning additional parameters to improve the model's ability to represent reality). The model setup and calibration process relies heavily on the data collected in the "Agricultural Practice Monitoring and Evaluation" (APME) project introduced in Ch. 1 – both to characterize weather, site, and management, and to provide crop yield, runoff, erosion, and nutrient loss information against which the simulation output can be compared. This Chapter therefore begins by describing that study and how it is used in this work.

2.1 The Agricultural Practice Monitoring and Evaluation Project

The information in this section is adapted from Braun et al. (2016), which gives a comprehensive description of the APME project and discusses results obtained through 2015. Most of the site characterization, agronomic records, and edge-of-field measurements used in this thesis were supplied by D. Braun at Stone Environmental. Some records and results from 2016 – 2017 were made available by J. Faulkner, L. Barbieri, and C. Twombly at the University of Vermont (personal communications, 2018).

2.1.1 Sites

Seven Vermont farms took part in this study that aimed to assess the effects of implementing conservation practices on corn and hay fields across the state (Braun et al., 2016). The farm sites and some of their basic characteristics are listed in Table 2.1. The project lasted from late 2012 to early 2016 at most of these sites, except for Charlotte (2015 – 2018) and Williston (2012 – ongoing).

The study employed a paired watershed design in which two similar, small (few-hectare) watersheds were identified at each farm. Both watersheds were managed in approximately the same way during a control period lasting about one year, and runoff and water quality data obtained during that period were used to verify that the watersheds had similar hydrologic properties. At that point, alternative management practices were implemented on one watershed while the other continued as before. The advantage of a paired watershed study is that control and treatment practices can be compared while effects due to temporal variations in weather conditions are in principle canceled out.

The standard (control) management for the hay fields was broadcasting of manure, while the alternative (treatment) practice was for the soil to be aerated prior to the manure broadcasting with the aim of improving infiltration and reducing losses to ammonia volatilization. The control management of the corn fields varied from site to site (Table 2.1), but one common element is that manure was broadcast at least once a year. The treatment management at those sites included manure injection, a grassed waterway, reduced tillage, and a water and sediment control basin. Cover cropping was practiced at some of the corn sites in the control and/or treatment periods, although establishment was often poor.

Table 2.1: Summary of farms in the APME project. The farms containing the watersheds that were modeled in this thesis are indicated in **bold***.*

Site	Crop(s)	Soil Texture(s)	Control Management	Treatment Management	Tile Drains[a]
Charlotte (CHA)	Corn	Silty clay	Manure incorporation; disk tillage; cover crop[b]	Grassed waterway	No
Franklin (FRA)	Corn, hay	Silt loam	Manure broadcast, aeration, incorporation; various tillage; strip cropping; cover crop[b]	Manure injection; reduced tillage; WASCoB[c]	Yes
Pawlet (PAW)	**Corn**	**Silt loam**	**Manure incorporated; chisel & disc plowing**	**Cover crop[b]**	**No**
Williston (WIL)	**Corn**	**Silt/sandy loam**	**Manure incorporated & broadcast; disc plowing; cover crop[b]**	**Manure injection; reduced tillage**	**No**
Ferrisburgh (FER)	Hay	Silty clay loam	Manure broadcast; ash+bedding broadcast	Soil aeration	Capped
Shelburne (SHE)	Hay	(Silty) clay loam	Manure broadcast	Soil aeration	Defunct
Shoreham (SHO)	Hay	Clay	Manure broadcast	Soil aeration	No

[a] APEX has limited ability to simulate P in subsurface drains (Francesconi et al., 2016)
[b] Cover crops failed to establish at WIL; some establishment in one season at PAW; good establishment in one season at CHA; good establishment in two seasons at FRA.
[c] Water and Sediment Control Basin

Two watersheds, PAW1 and WIL2, were selected for modeling in this thesis, based partly on expectations that they would be the simplest watersheds to use for an initial, detailed study. Corn is the only cash crop grown on these watersheds, and they have no known tile drainage that could complicate the modeling effort (e.g. Francesconi et al., 2016). Also, the annual management of these watersheds is relatively uniform over the duration of the monitoring study. This in principle permits the effectiveness of APEX in different weather conditions to be examined without confounding effects from major changes in farm operations.

PAW1 is technically the treatment watershed on the PAW1 farm, implying a change of management midway through the study. However, the cover crops seeded during the treatment phase at PAW1 generally failed to establish. In fact, weeds covered more ground on PAW2 in some years than cover crops ever did on PAW1 (Braun et al., 2016), suggesting more variability on the control watershed than on the treatment one. WIL2 is the control watershed on the Williston farm.

2.1.2 Data

A tipping bucket rain gauge and air temperature sensor monitored precipitation and temperature at each farm. These data were used to supply daily weather information to APEX during the model setup and calibration phase, with some substitutions made for missing and unreliable data as discussed in §2.2.2. An extensive set of soil physical and chemical properties, including pH, texture, cation exchange capacity, and macro-/micro-nutrient and organic matter (OM) content, was measured at the start of the project. A subset of that information was used to set up the soil properties at the start of the APEX simulation, as described in §2.2.3.

The operations carried out at each site (planting, harvest, tillage, and manure and chemical applications) were recorded, including date, method, variety, rate, etc., as applicable. Data about crop yields, cover crop establishment, and weed growth were also obtained. Much of this information was used in this thesis to create an operations schedule for each watershed as described in §2.2.4. The yield records formed part of the data set used during model calibration.

The runoff from each watershed was directed into a flume using plywood wing-walls or a soil berm. A water level sensor in the flume was used to construct runoff event hydrographs, while runoff samples were collected using a pair of autosamplers designed for small-medium and medium-large events. The water level sensors and autosamplers sent data every 30 minutes to a server at the project office, and sent text messages to staff to notify them that a runoff event was occurring. Project staff traveled to the site after every event to collect the water samples and dispatch them for analysis by the state Department of Environmental Conservation.

The runoff and water quality data from the Pawlet and Williston sites consist of the total flow for the runoff event (HQ), total suspended solids (TSS), and total and dissolved nitrogen and phosphorus (TN, TDN, TP, TDP). For some events, HQ was the only quantity that was measured; water quality data are not available. In order to compare APEX model output with the edge-of-field observations, it is necessary to identify the model variables that correspond to the measured quantities. Table 2.2 gives this translation.

Table 2.2: Runoff and water quality variables measured by the Agricultural Practice Monitoring and Evaluation project, and the corresponding model output variables

Observed quantity	Units	APEX Var.	APEX Definition[a]	APEX Units
Runoff vol. (HQ)	L	Q	Surface runoff	mm
Total suspended solids (TSS)	g	Y[b]	Soil loss from water erosion	tonnes ha^{-1}
Total P (TP)	g	QP + YP	P in runoff + P loss with sediment	kg ha^{-1}
Total dissolved P (TDP)	g	QP	P in runoff	kg ha^{-1}
Total N (TN)	g	QN + YN	N in runoff + sediment transported N	kg ha^{-1}
Total dissolved N (TDN)	g	QN	N in runoff	kg ha^{-1}
Silage yield	US tons acre^{-1}	YLDF	Forage yield	tonnes ha^{-1}

[a] Some variables are defined in more than one way in the APEX User's Manual; these definitions are taken from Table 2.9 in that document.

[b] APEX includes several options for calculating water erosion including the Universal Soil Loss Equation (USLE) and Small Watershed Modified Universal Soil Loss Equation (MUSS); see §3.3.3. The code generally outputs a specific "USLE", "MUSS", etc. variable. For brevity, these will be referred to as "Y".

In the case of TSS/sediment yield, the correspondence between edge-of-field and APEX variables is not exact. The observed TSS includes only the solids that were suspended in the runoff water and measured as part of the water quality analysis, whereas the modeled erosion (Y) represents all the sediment leaving the simulation area. In some cases sediment was deposited in the flume rather than leaving the site suspended in runoff, meaning that TSS underestimates the total amount of sediment lost from the field. However, except in the smallest events, the "missing" sediment is always a very small fraction of the total amount (Braun et al., 2016, Braun 2018, personal communication), and we assume that TSS $\equiv$ Y.

Field studies, especially those involving partnerships between scientists and working farms and measurements made under challenging outdoor conditions, are – understandably – affected by many sources of error and uncertainty. In the case of the APME project, for example, the management and yield records are imprecise at times. In 2014 the Williston farmer reported applying 4 loads of manure from a 9000 gallon tanker, but it is not clear whether that was on each watershed or in total. Yields were sometimes recorded as "18-22 tons/acre".

As for weather, the tipping rain gauges are expected to be accurate to $\pm 3\%$ in general, but are known to under-record very intense events and are not designed to accurately record solid precipitation (Braun 2017, personal communication). The accuracy of the flow (runoff) measurements in the flumes depends on the size of the event in a way that has not been fully characterized, although the fractional error will be larger for small events than for large ones (Braun 2018, personal communication). Erosion and frost heaving allowed some runoff to bypass the flumes at times, and icing in the flumes compromised some winter level measurements. In some cases the data had to be rejected outright. In others, APME personnel could make reasonable corrections. Errors in runoff volume will in turn affect the measurements of sediment and nutrients.

Attempting to rigorously quantify the uncertainties on the runoff and water quality data, etc., is beyond the scope of this thesis. For some general perspective, it is noted that Harmel et al. (2006) reviewed published errors for streamflow and TSS/nutrient measurements and found a range of 8 – 110% for total N and P and 7 – 53% for TSS under "typical" sampling conditions. Overall, the data from the APME study are an invaluable tool for assessing the performance of APEX at these Vermont sites. However, the uncertainties in the data and records mean that even the most accurate model cannot be expected to exactly match all of the observed measurements.

2.2 Initial Model Setup

The first step in setting up an APEX simulation is to enter the available information about the site, weather, and soil, and create a realistic operations schedule. Several other variables related to crops, tillage, fertilizer, etc. are also set at this point. Once this "baseline" model has been created, it can be further refined through the calibration process discussed in Chapter 3.

The model setup procedure essentially consists of editing numbers in a set of interrelated text files. An overview of the files and the relationships between them is given in Figure 2.1, reproduced from the APEX User's Manual (see below). During the setup and calibration phase of this work, APEX was configured using WinAPEX 1501[1], which provides a graphical interface to the program.

[1]WinAPEX is available from https://epicapex.tamu.edu/apex/winapex. More pre-

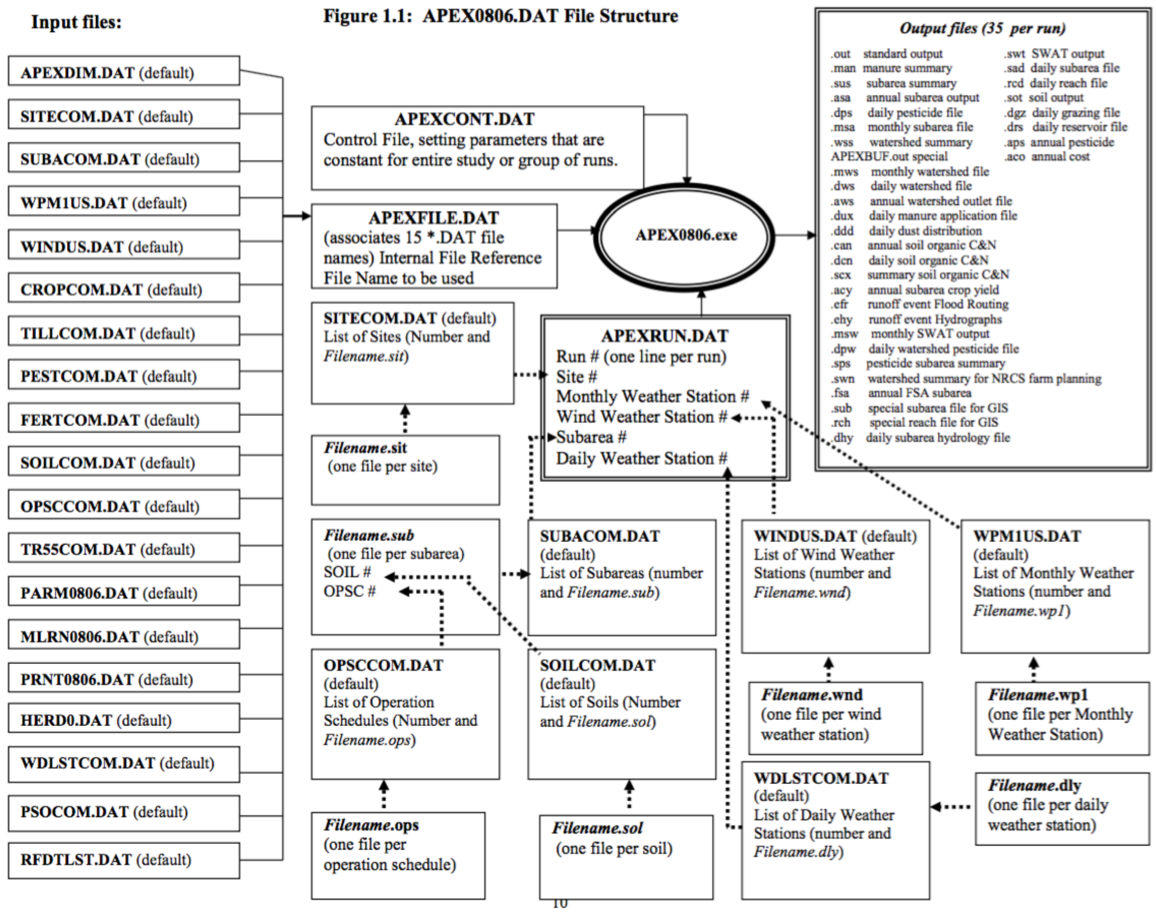

Figure 2.1: APEX input and output files, reproduced from the APEX 1501 User's Manual.

This section describes in some detail how baseline APEX models were created for the study watersheds, and the reasoning behind the decisions made along the way. It draws from several sources of information, in particular:

- The APEX Theoretical Documentation Version 0806 (Williams et al., 2012)

---

cisely, it was determined that in practice certain parameters are not translated correctly between the WinAPEX interface and the underlying APEX code (for example, see https://groups.google.com/d/msg/agriliferesearchmodeling/4F9m6EM1Pf8/vYlyOItWBQAJ). WinAPEX was therefore used to aid the initial model setup, then further parameter editing was carried out on the underlying text files and the APEX executable run directly.

  - Lays out many of the principles and equations underlying the code

- The APEX User's Manual Version 1501 (Steglich et al., 2016)
  - Defines input parameters and output files and variables

- The EPIC/APEX Modeling Forum[2]
  - APEX support staff answer questions and help troubleshoot model issues

2.2.1 Site Definition

An APEX simulation is comprised of one or more "homogeneous hydrological landuse units", otherwise known as "subareas", that can be linked together to form a whole watershed. The PAW1 and WIL2 "watersheds" in the APME project are treated in this thesis as single APEX subareas, disconnected from any other land. This means that water only enters the subarea in the form of precipitation; the simulations do not account for any contributions from runoff and subsurface flow from neighboring areas. This is an unavoidable limitation of APEX, and it may be more significant for these single-subarea simulations than for models with a smaller ratio of upland edge to simulated area.

The process of defining a subarea in APEX involves specifying basic information like the size and slope of the area, and the land condition that will be used for runoff curve number selection. These data are supplied in the Site and Subarea files. Those files also contain references to weather stations, management schedules, and soil types

[2] https://groups.google.com/forum/#!forum/agriliferesearchmodeling

that are defined in other files, which may in turn refer to sub-files of their own. (The contents of these other files are examined in the following sections.)

The subarea file contains a large number of options and parameters beyond basic site information, but many of them are not important for the purpose of this thesis. For example, channel properties do not need to be defined for a single-subarea simulation[3], and parameters to do with irrigation, automatic fertilization, reservoirs, etc. are not relevant. Table 2.3 summarizes the Site and Subarea parameters judged to be potentially relevant for this work and comments on their function and how some values were chosen. Table 2.4 gives the values of these parameters used for the PAW1 and WIL2 models.

[3]https://groups.google.com/forum/#!topic/agriliferesearchmodeling/J76A_Y_5Idc

Table 2.3: APEX/WinAPEX site/subarea parameters judged to be potentially relevant for this study

APEX Param.	Description	Comments
Site file		
	Watershed name	
	Weather station	
YLAT	Latitude	Used to estimate day length
ELEV	Elevation	Used in Penman-Monteith PET calculation
APM	Peak runoff rate-rainfall energy factor	Possible calibration parameter
Subarea file		
SNUM	Current subarea	
TITLE	Subarea name	
	Type of subarea	WinAPEX only; APEX derives from RCHL & CHL
	Downstream receiving subarea	WinAPEX only; APEX derives from RCHL & CHL
IOW	Owner	Owner should have herd size = 0
IWTH	Daily weather station	Used when reading daily weather from file
INPS	Soil number	Selects one of the soils defined in §2.2.3
IOPS	Operations schedule	Selects one of the schedules defined in §2.2.4
LUNS	Land condition	Based on site descriptions in Braun et al. (2016)
WSA	Subarea drainage area	From Table 1 of Braun et al. (2016)
CHL	Channel length	CHL=RCHL indicates this is an extreme subarea[a]
RCHL	Reach channel length	CHL=RCHL indicates this is an extreme subarea
STP	Average upland slope	From Table 1 of Braun et al. (2016)
SPLG	Average upland slope length	Estimated from contour maps in Braun et al. (2016)
UPN	Manning's N for upland	Estimated based on User's Manual Appendix F
LM	Lime application switch	0=auto lime, 1 = no lime
PEC	Erosion control practice factor	Possible calibration parameter

NVCN	Method used to adjust daily curve number	Set to same value as APEXCONT.NVCN[b]

[a] The values of the channel variables CHL and RCHL are not important per se for these single-subarea simulations, but setting them equal indicates to APEX that the subarea is not receiving material from any other subarea.

[b] There is also an NVCN variable in the Control file (§2.2.5). The APEX User's Manual does not explain how the NVCN values in the Control and Subarea files relate to one another; whether one overrides the other, for example.

Table 2.4: Physical characteristics of the sites and subareas in the PAW1 and WIL2 baseline models.

Parameter	PAW1	WIL2
Site file		
YLAT	43°.617	44°.468
ELEV	189 m	101 m
APM	1.0	1.0
Subarea file		
LUNS	Straight row, poor infiltration	Straight row, good infiltration
WSA	2.4 ha	0.81 ha
CHL	0.1 km	0.1 km
RCHL	0.1 km	0.1 km
STP	0.045	0.0006
SPLG	45 m	45 m
UPN	0.15	0.15
LM	1	1
PEC	1.0	1.0

2.2.2 Weather

The APME on-site and other local weather records were used to construct daily weather files and weather stations for APEX. On-site temperature (T) and precipitation (precip.) measurements are available for each site for the few-year duration of the monitoring project (§2.1). However, a "run-up" period is needed before the model output can be used, and long-term simulation results can help to identify trends and anomalies in the model output. Alternative weather data was therefore needed to enable longer-duration modeling.

Daily T and precip. records from the nearest Global Historical Climatology Network (GHCN) weather station were used for this purpose. For the Pawlet model, the nearest weather station is at Rutland, roughly 33 km distant (as the study farms are anonymous and their exact locations are not known to this author, only estimated distances can be given). For Williston, the nearby (∼8 km) Burlington Airport station was selected.

The on-site precipitation was measured using tipping bucket rain gauges. These instruments are not designed to record solid precipitation, meaning that the winter precip. records were likely to be inaccurate. In an attempt to roughly correct for that effect, weather station precip. values were substituted for on-site precip. on days where the on-site $T_{max} < 0°C$. Weather station data were also substituted when on-site records were obviously incorrect, specifically in cases where $T_{max} > 50°C$.

Figures 2.2 and 2.3 show weather station vs. on-site T and precip. for the two modeled locations. In both cases, the on-site and weather station temperatures are quite well-correlated. On days where the on-site $T_{max} < 0°C$, the on-site precip. is often zero while non-zero precip. is recorded at the weather station. This may reflect the inability of the on-site rain gauges to record snowfall. The precipitation at the Williston site is quite closely correlated with that at the Williston weather station, so substitutions made for that site are probably fairly accurate. Substitutions made at Pawlet are likely to be less accurate.

Once daily weather text files had been constructed for the whole simulation duration, the Weather Import program[4] was used to convert the data to daily weather files and weather stations in the format required by APEX. Note that these weather files only contain T and precip.; the remaining weather variables (humidity, solar radiation, wind speed) are generated by the model (see also §2.2.5).

[4]https://epicapex.tamu.edu/model-executables/weather-import/

2.2.3 Soils

APEX support staff kindly made available a WinAPEX database containing APEX parameters for all the soil series in Vermont, based on the Soil Survey Geographic Database (SSURGO[5]). The first step in setting up soil parameters for any site, then, was to select from the database the soil series making up the largest fraction of the watershed. For PAW1, "Bomoseen and Pittstown soil 148B" comprises 63% of the site, followed by Bomoseen and Pittstown 148C at 34% and Taconic-Macomber complex 43C at 4%. As "B" and "C" soils differ only in their slopes (2-8% vs. 8-15%), and the site slope is defined during the subarea setup (§2.2.1), PAW1 is treated in this work as effectively composed entirely of soil 148B.

The SSURGO database contains two entries for soil 148B: Bomoseen (CH-L) and Pittstown (SIL). The Bomoseen soil was selected for the baseline model. The properties of these soils are very similar, so APEX models based on each of them would not be expected to differ noticeably. At WIL2, Winooski and Limerick soils respectively cover 65% and 35% of the WIL2 site. The Winooski soil was selected for the baseline model. The Winooski and Limerick soils differ somewhat more than the Bomoseen and Pittstown soils, but tests showed that the effect of substituting the Limerick soil was small.

A soil in APEX is characterized by roughly 60 parameters that describe everything from the soil's pH to the lignin content of structural litter in the soil. The SSURGO database contains reasonable values for many of these parameters, and APEX will estimate values for those that are not available. Some of the parameters, like sand and silt fractions, pH, and cation exchange capacity (CEC), are commonly measured on

[5]https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053627

individual farm fields. These can be used to replace the SSURGO values, which may be important for watersheds that contain a mixture of soils, or where management activities may have led to site-specific changes over time.

An extensive set of soil test data was obtained at the APME watersheds. To use these data to parameterize the soils in APEX it was first necessary to identify the APEX input parameter or output variables corresponding to the soil test measurements. In some situations, the correspondence is fairly clear. For example, it is straightforward to use the soil test pH value as the APEX PH parameter. Other field measurements are less clearly related to the pools of nutrients and organic matter that APEX uses. Although the APEX SOIL file contains an "Organic Carbon Concentration" parameter, and the APME data set includes an "Organic C" measurement, for instance, the Organic C number appears to refer to a water-extractable carbon pool that is much smaller than the organic C pool in APEX (Braun and Moore, 2014).

Table 2.5 gives the translation between APEX variables and soil test data for the measurements for which a reasonable correspondence could be found. Approximate values for APEX organic C and N can be obtained from soil test organic matter (OM) by assuming that organic C $\sim$ OM/1.9 (Pribyl, 2010) and organic N $\sim$ organic C/12 (Kirkby et al., 2011; Lal, 2008). APEX soluble P can also be derived from soil test results. Winchell et al. (2011) used a database from the UVM Agricultural Testing Laboratory to derive a relation between phosphorus concentrations obtained from the Modified Morgan (MM) soil test method used in Vermont, and the Mehlich 3 (M3) P values that are required by the SWAT model. Stone (2015) used this relation, with M3 values divided in half, to initialize soluble P in APEX. At high values of MM P ($\gtrsim$10 ppm), the scatter in the relation is very large.

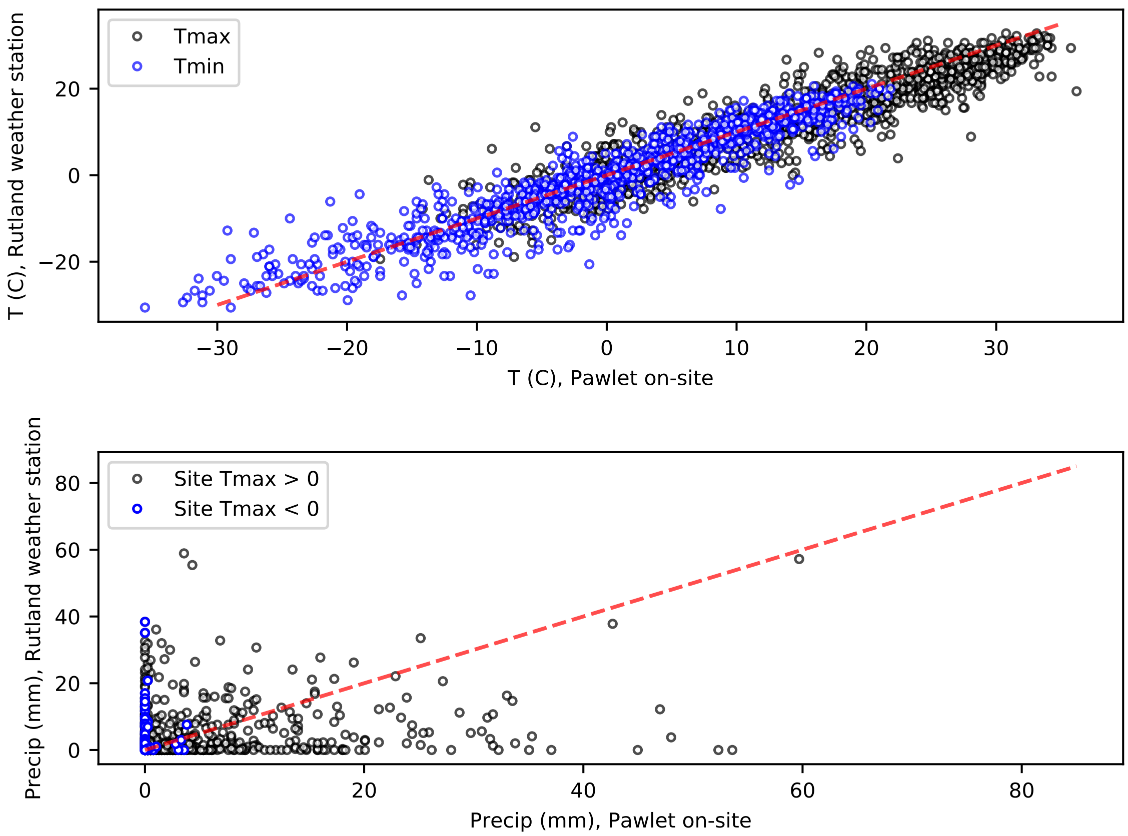

Figure 2.2: Top: Minimum and maximum daily temperature at the Rutland weather station compared with that at the Pawlet field site. Bottom: Daily precipitation at the Rutland weather station compared with that at the Pawlet field site. The blue points in this panel identify days where the on-site maximum temperature was below freezing and the tipping bucket rain gauge measurements are therefore likely to be unreliable. On those days, APEX was supplied with precipitation measurements from the Rutland weather station instead of on-site data. In both panels, the red dashed lines show the 1:1 relation.

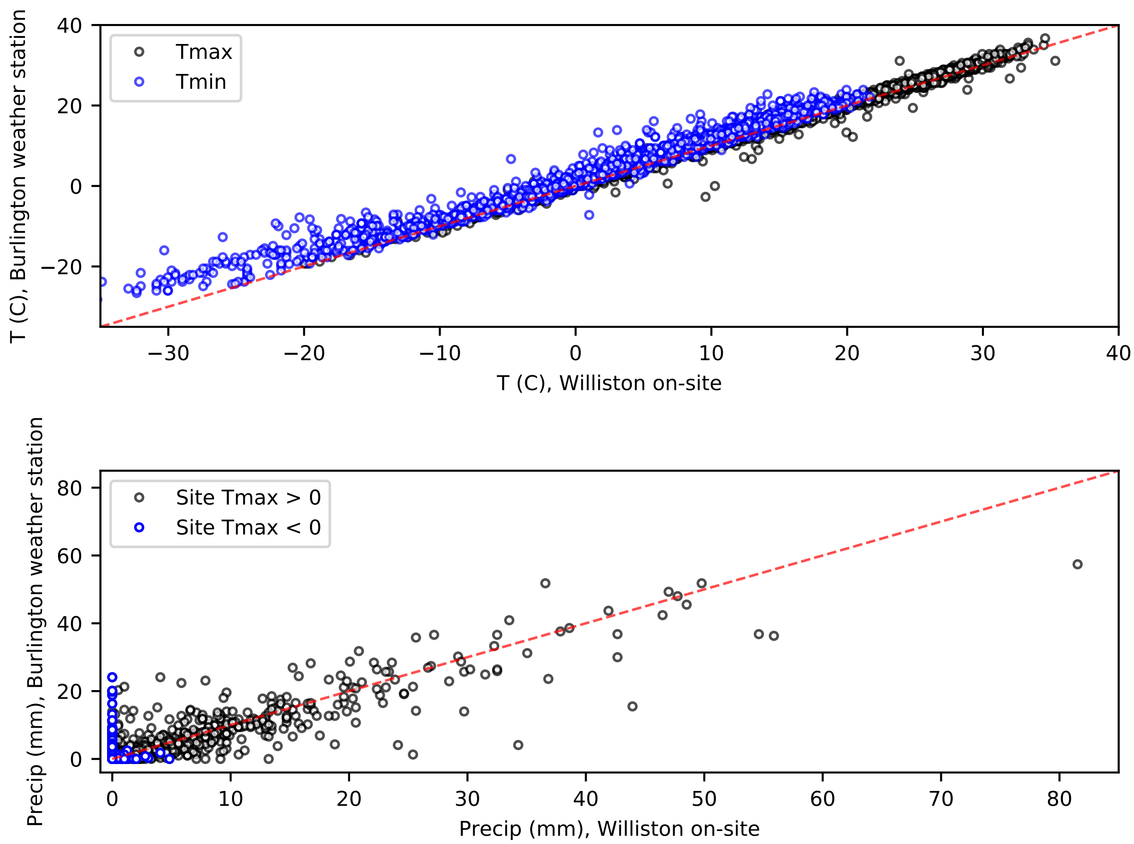

Figure 2.3: Top: Minimum and maximum daily temperature at the Burlington Airport weather station compared with that at the Williston field site. Bottom: Daily precipitation at the Burlington weather station compared with that at the Williston field site. The blue points in this panel identify days where the on-site maximum temperature was below freezing and the tipping bucket rain gauge measurements are therefore likely to be unreliable. On those days, APEX was supplied with precipitation measurements from the Burlington weather station instead of on-site data. The correlation between on-site and weather station data in this figure is much better than in Figure 2.2, probably because of the smaller distance between the field site and weather station. In both panels, the red dashed lines show the 1:1 relation.

Table 2.5: Translation between APME soil test data and APEX variables

Soil Test Measurement	APEX Param.	APEX Definition	APEX File[a]	Conversion Factor	Reference[b]
pH	PH	Soil pH	SOIL		
Ca+K+Mg (%)	SMB	Sum of bases (cmol/kg)	SOIL		
OM (%)	WOC	Organic C (%)	SOIL	WOC∼OM/1.9	P10
	WOC	Organic C (kg/ha)	ACN	WOC∼OM/1.9	P10
	WN	Organic N (ppm)	SOIL	WN∼WOC/12	K11, L08
	WON	Organic N (kg/ha)	ACN	WN∼WOC/12	K11, L08
Sand (%)	SAN	Sand content (%)	SOIL		
Silt (%)	SIL	Silt content (%)	SOIL		
Avail. P (mg/kg)	SSF	Soluble P (g/Mg)	SOIL	See refs	W11, S15
	AP15	Plow depth soluble P (g/t)	ACY	See refs	W11, S15
CEC (meq/100 g)	CEC	Cation exchange capacity (cmol/kg)	SOIL		

[a] The SOIL file contains parameters that initialize the soil at the start of the simulation, whereas the ACN and ACY files contain annual simulation output. The SOIL and ACN files contain values for each soil layer in the model.

[b] P10: Pribyl (2010); K11: Kirkby et al. (2011); L08:Lal (2008); W11: Winchell et al. (2011); S15: Stone (2015).

Table 2.6: Soil variables adjusted for the baseline models[a]

Parameter	PAW1		WIL2	
	Default	Baseline	Default	Baseline
PH	6.45	7.9	5.9	7.3
WN (ppm)	0[b]	1500	0[b]	1700
SAN (%)	32.9	35.0	60.68	31.3
SIL (%)	57.1	49.6	27.82	55.9
WOC (%)	2.35	1.95	2.06	1.0
CEC (cmol/kg)	10.06	18.8	9.63	12.9
SMB (cmol/kg)	0[b]	18.8	0[b]	12.9

[a] Soil samples were taken to a depth of 20 cm, so APEX values were only adjusted for the top model soil layer. The thickness of this layer is 23 cm at PAW1 and 20 cm at WIL2.

[b] In APEX, 0 generally indicates an unknown value that will be filled in by the model.

Some soil properties do not change significantly with time, and can therefore be used to initialize the model. These are the sand and silt fractions, pH, cation exchange capacity (CEC), and base saturation. (Actions such as fertilizer applications can certainly change the pH of a soil. However, experiments with APEX showed that the initial and final soil pH remained the same in a 40-year simulation including annual fertilizer applications, suggesting that APEX does not track changes in pH related to management activities. Similarly, although the CEC of a soil is related in part to its organic matter content, the initial and final CEC reported in the APEX log file were found to be the same.). The APME soil test results were therefore used to set SAN, SIL, PH, CEC, and SMB for the baseline models, as shown in Table 2.6.

Other soil characteristics, particularly those related to organic matter and nutrients, can change appreciably during a multi-year APEX simulation. Rather than use soil test data directly to initialize the model in these cases, initial values can be set so that the correct values are achieved later in the simulations. The initial values of

WOC and WN were therefore adjusted so that WOC and WON reported by APEX in 2012 matched the values estimated from the organic matter level measured in the field that year (Figures 2.4, 2.5).

In these simulations the soil organic C and N concentrations each year are also affected by the amount of C and N added in manure and the organic matter cycling in APEX. There are uncertainties in all of these quantities/processes, so an acceptable match between the estimated and modeled WOC and WON in 2012 could have been achieved by varying any one of a number of input parameters. The adopted initial values of WOC and WON are not unique.

APEX does not report annual soil organic P, and the initial organic P (WPO) value was left at 0 ("unknown"). For both PAW1 and WIL2, the model assigned a starting WPO value such that the initial WN/WPO$\approx$8. The initial soluble P value (SSF) was also left at its default value, partly because of the substantial uncertainty in the conversion between MM P and M3 P (especially for WIL2. where MM P = 43.5 ppm). Also, the modeled soluble P is unusually low at WIL2 in 2012 (Figure 2.4), the year the soil tests were obtained. With SSF at its default value, the plow depth soluble P concentration reported by APEX in 2012 turned out to be within a factor of two of the M3 P value estimated from the soil test data (Figures 2.4, 2.5).

In summary, soil properties in APEX were set to match field measurements as well as possible, given the significant uncertainties involved in converting between the available soil test data and APEX parameters. A further limitation is that soil parameters may not be independent; changing one parameter might imply that another property should also change. For example, a change of soil texture would probably in practice also change the bulk density. These potential shortcomings were not

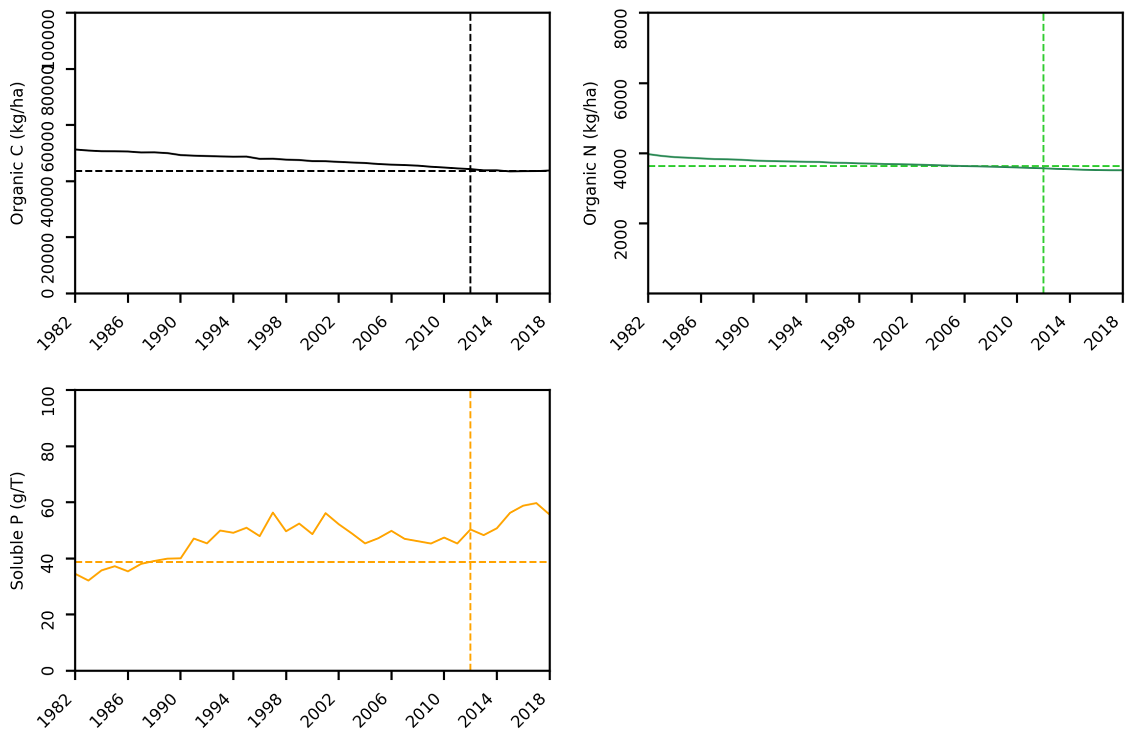

Figure 2.4: Top left: APEX annual organic C in the top ≈20 cm of the soil at PAW1. The dashed lines intersect at the organic C value estimated from soil test data obtained on the PAW1 watershed in 2012 (see text). Top right: same for organic N. Bottom left: same for soluble P in the plow layer.

addressed in this study.

2.2.4 Operations

APEX also needs to know the dates and types of the management activities that are carried out on the watershed. The actual activities that took place on the APME fields varied considerably from year to year. For example, operations sometimes had to be postponed because of poor weather, or the farmer used different tillage methods from one year to the next.

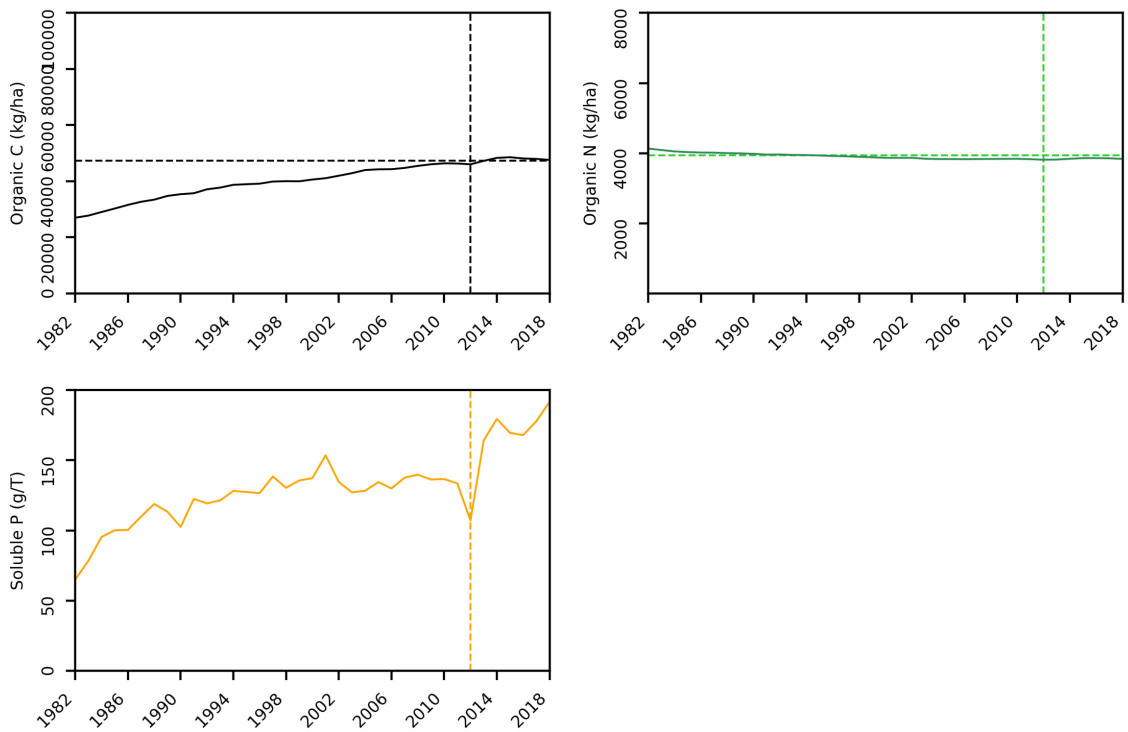

Figure 2.5: Top left: APEX annual organic C in the top ≈20 cm of the soil at WIL2. The dashed lines intersect at the organic C value estimated from soil test data obtained on the WIL2 watershed in 2012 (see text). Top right: same for organic N. Bottom left: same for soluble P in the plow layer.

Rather than construct an "average" or "typical" annual schedule, the approach was to mimic the actual operations as closely as possible in terms of dates, seeding and manure/fertilizer application rates, corn days to maturity, etc. In that way, we can ask "what would have happened if this farm had been operated in the same way but in different climates?". In model years prior to the start of the APME project, and after the end of the project, the schedule for the first APME year was used.

There were some differences between the actual and model schedules. One feature of APEX is that a planting operation will not take place until the soil temperature reaches 2° C above the crop's base temperature (T_{base} = 10° C for corn)[6]. This means that the corn planting was delayed for several days in some years.

Also, although winter cover crops were seeded on the PAW1 and WIL2 watersheds, the cover fraction observed each winter ranged from only 0 – 28% on PAW1 and ≤4% on WIL2, and no cover crop operation was included in the model schedule. In addition, the manure application rate at WIL2 in 2012 was unusually low (33000 kg ha^{-1} vs ≥65000 kg ha^{-1} in almost all other applications), so the rate for years prior to 2012 was set to a more representative value (72000 kg ha^{-1}). Finally, as the APME data set does not include pesticide masses in runoff/sediment, pesticide applications were omitted from the schedule.

[6]https://groups.google.com/d/msg/agriliferesearchmodeling/QnLo0XkZAG8/WrrBlsgzBgAJ

Table 2.7: Sample years from the APEX operations schedules.

Operation	Date	Rate	Substance	Other
PAW1, 2012				
Apply Manure	5-12	35867 kg/ha	VT Dairy Manure	
Plow, Chisel	5-12			LUN=0[a]
Plant Corn	5-29	79000 seeds/ha		GDUs/PHU=1333[b]
Apply Fertilizer Through Planter	5-29	224 kg/ha	27-9-18	
Harvest Silage	9-27			
Kill[c]	9-27			
WIL2, 2013				
Apply Manure	5-7	116569 kg/ha	VT Dairy Manure	
Plow, Disc	5-9			LUN=0
Plant Corn	5-16	84000 seeds/ha		GDUs/PHU=1333
Harvest Silage	10-9			
Apply Manure	11-10	116569 kg/ha	VT Dairy Manure	
Kill	11-9			

[a] LUN=0 implies that the plow operation does not override the land condition defined in the Subarea file (§2.2.1)
[b] Growing degree days/potential heat units to maturity; see §2.2.4
[c] APEX requires a "kill" operation for annual crops that indicates that growth has finished.

As an illustration, one year of the PAW1 and WIL2 schedules are given in Table 2.7. The management is similar in that both sites grow corn fertilized with dairy manure, but manure is applied twice at WIL2 and at higher rates than at PAW1, and no starter fertilizer is used at WIL2. To fully specify the entries in the schedule, the relevant Tillage, Crop, and Fertilizer files must also be populated with reasonable values. This is discussed in the following sections.

Tillage

Most of the operations in Table 2.7 disturb the soil. Each of the items in the "Operation" column in the table therefore has an entry in APEX's Tillage file that describes the depth of disturbance and mixing efficiency of the operation, the surface roughness created, etc. For plowing/tillage specifically, the APME management records refer to numerous methods and implements, such as "harrowed with a Sunflower disc harrow" and "disc chisel plow". These terms were simplified into just two kinds of tillage, "Plow, Disc" and "Plow, Chisel", based on advice about likely implements and soil disturbance levels from J. Faulkner (2018, personal communication).

The parameters of the tillage operations were based on pre-defined operations in WinAPEX. To create each operation used by the baseline models, a similar template operation was selected and its tillage parameters verified to (1) reflect relative soil disturbance, and (2) appear to be generally reasonable. For example, to create the Plow, Chisel operation, the existing Plow, Chisel, 20 ft operation was selected and renamed (the "20 ft" has only economic, not physical implications), and verified to cause less disturbance (have a lower mixing efficiency) than the pre-defined Plow, Disk, 22 ft operation. Table 2.8 lists the final tillage parameters that were used.

Table 2.8: Tillage-related parameters[a] for the APEX operations.

APEX Variable	Description	Apply Manure	Plow, Disk	Plow, Chisel	Fertilize, Planter[b]	Fertilize, Surface[c]	Plant Corn	Harvest Silage
EMX	Fraction mixing efficiency	0	0.85	0.3	0	0	0.1	–
RR	Random surface roughness (mm)	0	50	20	0	0	10	–
TLD	Application/tillage depth (mm)	0	100	150	40	0	40	-50[d]
RHT	Ridge height (mm)	0	0	50	0	0	75	–
RIN	Ridge interval (m)	0	0	0.3	0	0	1	–
HE	Harvest efficiency[e]	–	–	–	–	–	–	0.95
ORHI	Override harvest index[f]	–	–	–	–	–	–	0.9
FPOP	Frac. plant population reduced	0	0	0	0	–	–	–

[a] The APEX tillage file also contains variables that influence only the economics of the simulation; these are omitted here

[b] Actual name, "Apply Fertilizer Through Planter": Starter fertilizer applied through the planter at same depth as corn seed and with no extra soil disturbance

[c] Actual name, "Apply Fertilizer On Surface".

[d] Negative tillage depth indicates height at which crop is harvested

[e] Ratio of yield leaving the field to total crop yield

[f] Ratio of harvestable yield to total crop biomass. For forage crops, overrides the value of HI set in the crop file, allowing different amounts of the crop to be harvested with different types of equipment.

Crops

APEX's Crop files include numerous parameters describing plant nutrient uptake, heat units required for germination, efficiency of converting radiant energy into biomass, etc. Files with pre-defined values are available for a wide range of crops, including silage corn. The User's Manual states that "The crop-parameters should not be changed without consulting the model designers or without solid knowledge of plant growth and development" (p104).

For the baseline model, the only crop-related parameters that were changed from the defaults were the plant population (OPV5), heat units required for germination (GMHU), and the heat units at maturity (PHU)[7]. OPV5 was based on agronomic records, and GMHU was changed from its default of 100 (Fahrenheit scale; Darby and Bosworth, 2004) to 55 (Celsius scale).

The PHU parameter governs the number of heat units (growing degree days; GDUs) that are required to bring a plant to maturity, so for correct crop yields it must be set appropriately for the crop variety and climate at the site. Both farms grew corn with hybrids requiring ∼95-105 days to maturity, which translates to approximately 2200 - 2500 GDUs in Fahrenheit, or 1220 - 1390 GDUs in Celsius (Carter, 1992; Nielsen, 2012).

Fertilizers

Two kinds of fertilizer were used on the study farms: cow manure from each farm's storage pit, and synthetic NPK fertilizers (e.g. Table 2.7). To simulate the effects

[7]The OPV5 and PHU parameters are actually set in the Operations file (Table 2.7), but are discussed here along with the other crop-related parameters.

of applying a fertilizer, the amounts of N, P, and K in their various forms must be specified in the APEX Fertilizer file (Table 2.9). For a fertilizer such as 27-9-18 this simply requires translating the fertilizer's P and K fractions from P_2O_5 and K_2O into elemental P and K (so FN=0.27, FP=0.04, FK=0.15). It was also assumed that the N in the synthetic fertilizers was in urea form (J. Görres, personal communication, 2018), so the FNH3 (Ammonia N fraction) parameter was set to 1.0.

Manure analyses from the farms in the monitoring study are not available for the Pawlet and Williston farms, but typical values of mineral and organic N, mineral P, and mineral K in liquid manure from Vermont dairy farms can be found in Jokela et al. (2004). These were converted from lbs/1000 gallons to lbs/lb (=kg/kg) assuming a density of 8 lbs/gallon (Beegle and Peters, 2011), and from P_2O_5/K_2O to elemental P/K as appropriate.

The APEX fertilizer file also specifies the organic P, NH_4-N, and organic C in the fertilizer. Roughly 10-40% of the P in manure is in the organic form (Sharpley and Moyer, 2000), and a value of 25% was assumed for this work (i.e., FPO = FP/3). All of the N was assumed to be in the form of NH_4-N (D. Ross & J. Tilley, personal communication, 2017).

Less information is available about the organic carbon fraction of the manure in Vermont's farm pits. Pettygrove et al. (2009) found that, in lagoon water from California dairies, suspended C plus dissolved organic C totaled aproximately 2230 mg/L. Assuming the same density as above, this corresponds to FOC= 0.23%. In another study, Moral et al. (2005) found a mean organic C fraction of 23% dry matter in samples of Spanish cow manure on a dry matter basis. This converts to FOC= 1.6%, assuming a dry matter fraction of 7% (Jokela et al., 2004). The manure pits

Table 2.9: Properties of the dairy manure used for the baseline models

Param.	Description	Value	Derived From
FN	Mineral N[a]	1.5×10^{-3}	Jokela et al. (2004)
FP	Mineral P[a]	4.0×10^{-4}	Jokela et al. (2004)
FK	Mineral K[a]	2.1×10^{-3}	Jokela et al. (2004)
FNO	Organic N[a]	1.6×10^{-3}	Jokela et al. (2004)
FPO	Organic P[a]	1.3×10^{-4}	Sharpley and Moyer (2000)
FNH3	NH_4-N[b]	1.0	D. Ross & J. Tilley 2017, pers. comm.
FOC	Organic C[a]	5×10^{-3}	Moral et al. (2005); Pettygrove et al. (2009)
FSLT	Salt[a]	5.0×10^{-2}	University of Arizona (2000)

[a] Judging by the worked example in section 2.18 of the APEX User's Manual, this is the fraction of the total wet weight of manure that is composed of this substance.
[b] Fraction of the mineral N that is in the NH_3 form

in the APME study are often described as "well-agitated", and may contain more suspended C than the Pettygrove lagoon water, but probably less bedding material than the Moral samples. An intermediate value of 0.5% (FOC=0.005) was chosen for this work.

2.2.5 Control File

The Control file contains a number of parameters that govern global simulation properties. This includes specifying equations to use for calculating processes like runoff and erosion, and quantities such as daily soil water and curve number. The baseline values of the control parameters that could be relevant to these simulations are shown in Table 2.10. Several of these options were later varied as part of the model calibration process described in Chapter 3.

The Control file also contains information about the simulation starting date and duration. APEX requires a "run-up" period to allow the model to stabilize before reliable results can be obtained. Studies in the literature have used run-up periods as short as two years (e.g. Stone, 2015), while periods as long as 15 years have also been recommended[8]. To allow an adequate run-up period, as well as permitting visibility of long-term trends, the baseline models begin in 1982 and run for 37 years (through 2018).

[8]https://groups.google.com/d/msg/agriliferesearchmodeling/_OaXZg-F7_4/6B4x2WR_AwAJ

Table 2.10: Baseline values for potentially relevant variables in the control file, used for both PAW1 and WIL2.

Parameter	Description	Default	Baseline	Comments
NBYR	Simulation duration (yrs)	40	37	Run through 2018, long-term trends visible
IYR0	Beginning year	1960	1982	Start 30 years prior to APME[a]; 15-year run-up
IPD	Print code	2	6	Print daily output
NGN	Weather input code	2345	2	Read T & P, generate all other weather variables
IGN	Num. of random number cycles	0	0	Same seed used for all simulations
IET	Potential evapotranspiration eq[n]	4	4	Hargreaves[b]
ISCN	Stochastic CN[c] estimation	1	1	Deterministic (not stochastic) CN calculation
ITYP	Peak rainfall rate code	1	3	SCS type II rainfall distribution, OK for US NE
ISTA	Soil profile code	0	0	Normal soil erosion (not static profile)
IHUS	Auto heat unit scheduling	0	0	Operations as scheduled (not by heat units)
NVCN0	Daily CN calculation method	4	4	Based on soil moisture index
INFL0	Runoff estimation methodology	0	0	CN (not Green & Ampt)
IERT	Enrichment ratio method code	0	0	EPIC (not GLEAMS)
LBP	Soluble P runoff method	0	0	GLEAMS (not “modified nonlinear”)
NUPC	N and P plant uptake code	1	1	s-curve (not Smith curve)
ISLF	Slope length/steepness factor	0	0	Use RUSLE slope factor in erosion equations
ICO2	Atmospheric CO_2	0	0	Constant CO_2
ISW	Soil water calculation code	0	0	Rawls method, dynamic
RCN0	N in rainfall (ppm)	0.8	0.8	Default
CO20	Atmospheric CO_2 (ppm)	330	400	www.esrl.noaa.gov/gmd/ccgg/trends/index.html
RTNO	Yrs of cultivation at sim. start	0	0	Affects C and N mineralization
DRV	Equation for water erosion	3	3	MUSS (small watershed MUSLE)

[a] Starting in 1982 also avoids starting the simulation on a leap year. A leap year start means that operations are delayed by one day in that year and all subsequent years (https://groups.google.com/d/msg/agriliferesearchmodeling/QnLo0XkZAG8/WrrBlsgzBgAJ).

[b] The Hargreaves equation requires only T and precip. as input, and is recommended by APEX support staff when only those weather data are available (https://groups.google.com/d/msg/agriliferesearchmodeling/maz5IDXcuRk/jpoCi41_IgAJ).

[c] Curve number, see §3.3.1.

2.2.6 Parameter File

The Parameter file consists of two parts: 30 s-curve parameters and >100 other miscellaneous parameters. An s-shaped curve is used to represent many processes in APEX (such as soil water evaporation as a function of depth) and two parameters per process specify the shape of the curve. The set of miscellaneous parameters affects everything from the effect of soil bulk density on root growth to the day length required for dormancy in winter crops. To distinguish this second set of parameters from the s-curve variables, and for consistency with the notation in the APEX documentation and tools, I will refer to them as the "PARMs".

The User's Manual states that "The PARMCOM.DAT file is a very sensitive part in APEX, because many coefficients of equations are maintained in this file. The equation coefficients should not be changed without consulting the model designer first." Nonetheless, Wang et al. (2012) identify several PARMs that can be useful for model calibration, the APEX-CUTE auto-calibration tool (§3.1; Wang et al., 2014) includes PARMs 1 – 100, and published studies routinely adjust PARMs to improve the model's performance (e.g. Baffaut et al., 2017). For the baseline models in this thesis, all the s-curve parameters and PARMs were kept at the default values in the WinAPEX Parm Table. This includes PARM78, whose default value (10) means that operations will not be delayed because of excessive soil moisture (§4.2).

The meaning of "default" may be ambiguous for two of the PARMs. In the interest of clarity, these are:

- PARM 101 (Century Passive Humus Transformation Rate): The WinAPEX Parm Table lists the default value of this parameter as 0.000012, and the User's

Guide states that this is the "original value" for PARM 101. However, the actual value in the Parm Table is 0.2918. The selected value was set back to the default for the baseline models.

- PARM 102 (Lower Nitrogen-Carbon Ratio of Biomass): The WinAPEX Parm Table lists the default value of this parameter as 0.0667, but the actual value in the Parm Table is 7.35. The selected value was set back to the default for the baseline models.

At this point the relevant APEX parameters for the baseline have been set to reasonable values and the first baseline simulation can be executed. The following sections explore in some detail the output of the baseline models, and compare the output to the APME edge-of-field and crop yield data.

2.3 Baseline Models: Overview of Output

A useful first step in evaluating the output from an APEX run is to simply visualize the overall properties of the simulation. Are soil nutrient levels stable over time or are they building up or being depleted? What are the dominant pathways of nutrient loss? What are the main factors affecting the crop yield? Answering questions like these can help identify problems with the setup, show whether the period for which comparison data are available differs in any important ways from the simulation period as a whole, and provide insight into areas that may be candidates for further calibration.

This kind of simulation overview relies on the user selecting variables of interest from among APEX's numerous outputs, so it is necessarily a narrow and subjectively-chosen view of the model. However, it provides the background for a more detailed and quantitative evaluation of the model's performance, in which its output is compared with the available real-world data in a systematic way. The following sections, then, present a set of plots illustrating the baseline models, describing the characteristics and trends observed. The models are then compared with the data obtained for the APME project.

In the hope of aiding the reader in navigating the many APEX output plots, standard colors will be used in Chapters 2 and 3 to denote quantities related to water, temperature, carbon, etc.:

- Water (precipitation, percolation, water stress, etc.): blues
- Carbon: grays
- Nitrogen: greens
- Phosphorus: red/orange/yellow
- Erosion/sediment: browns
- Crop yield: purples
- Temperature: pinks

2.3.1 PAW1

Figures 2.6 and 2.7 give an overview of the PAW1 baseline model in terms of the total inputs and outputs of water, carbon, phosphorus, and nitrogen that are reported in

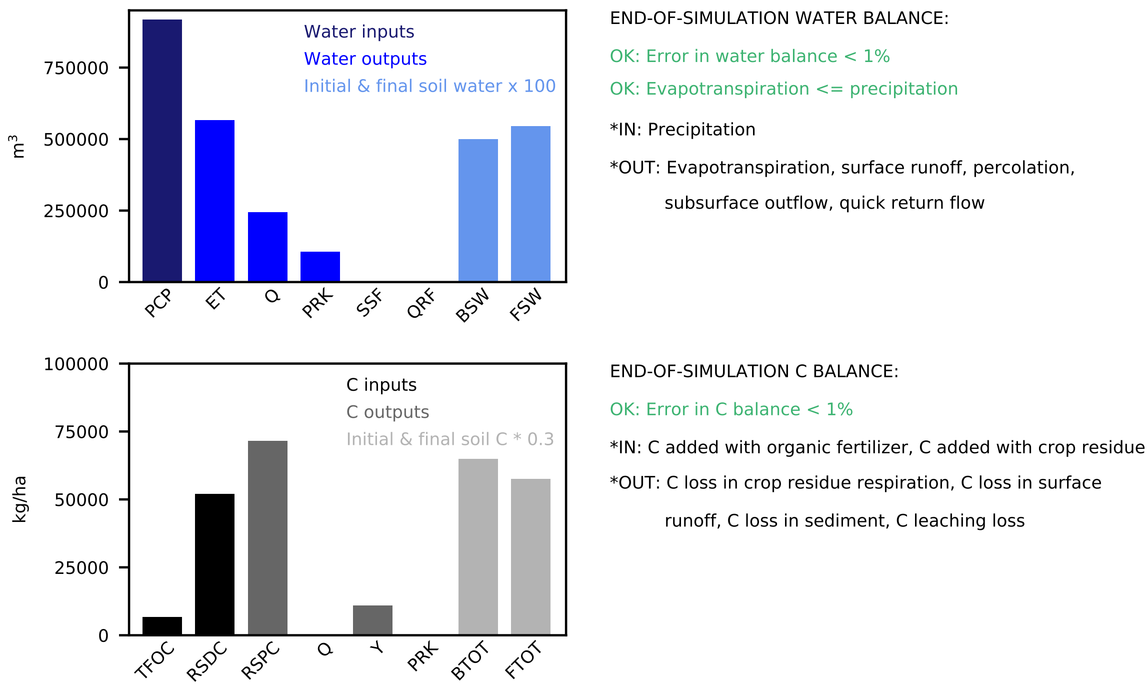

Figure 2.6: Summary of water and carbon inputs and outputs over the course of the PAW1 baseline simulation.

the model's main log (".OUT") file. The figures illustrate the main pathways of addition and loss of these materials, as well as comparing their initial and final amounts in the soil. In the case of water, for example, evapotranspiration is the main source of loss from the system, with runoff and percolation accounting for lesser amounts. Quick return flow and lateral subsurface outflow contribute negligible quantities.

Carbon is added to the soil with organic fertilizer (manure) and when crop residue decays. Crop residue also accounts for the predominant losses of C (via respiration), with a much smaller fraction lost as soil erodes. The overall C balance in the simulation is negative, with the final soil C at 88% of the initial value. P, in contrast, accumulates slightly over the 37 year simulation period. Inputs in the form of min-

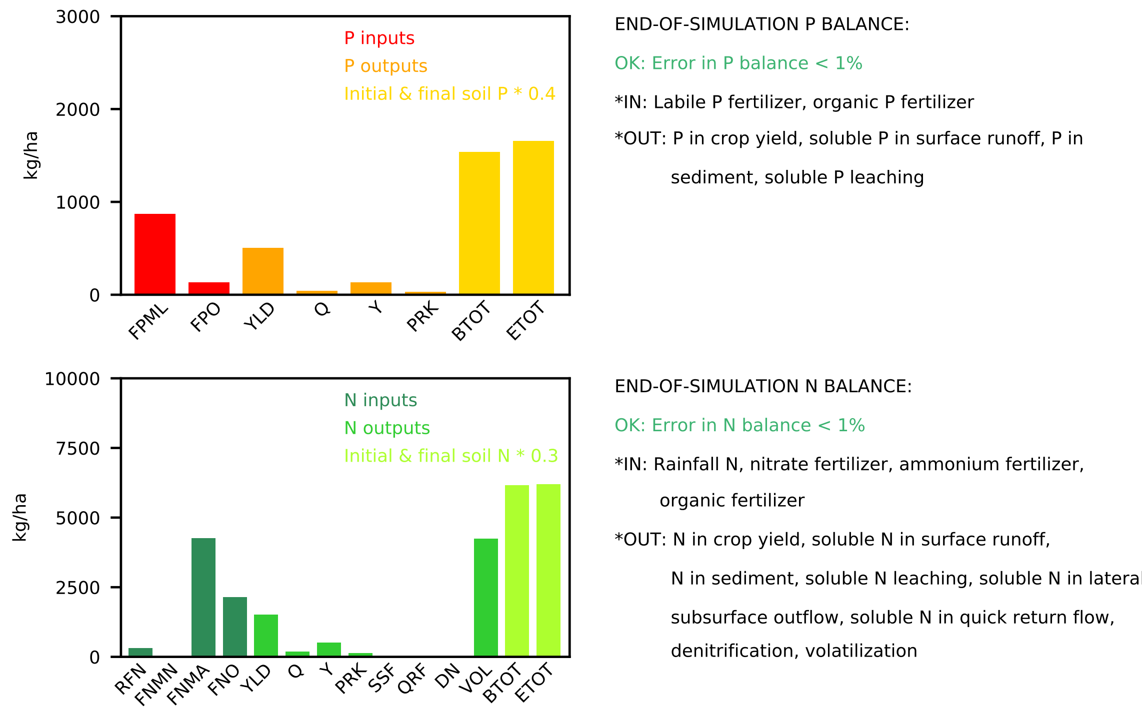

Figure 2.7: Summary of P and N inputs and outputs over the course of the PAW1 baseline simulation.

eral/labile, and to a lesser extent organic, P fertilizer outweigh P losses to harvest, erosion, runoff, and leaching.

Most of the N in the simulation is added in fertilizer, with ammonia contributing about twice as much as organic N. The most important losses are due to ammonia volatilization, followed by harvest, erosion, runoff, leaching, denitrification, quick return flow, and lateral subsurface outflow in that order. Soil nitrogen is about the same at the end of the simulation as at the beginning.

Visualizing some of the year-by-year output gives further insight into the processes occurring in the model; scatter and temporal trends in the results; and the relative influence of various factors in determining important outcomes such as crop yield.

Figure 2.8 shows a small subset of the PAW1 model output: precipitation, runoff, percolation, and erosion for each year of the simulation, and the amounts of N and P lost via those pathways.

Runoff loosely tracks rainfall over the course of the simulation, with an overall rise in both quantities from 1982 until roughly 2011 followed by an abrupt drop. Percolation, on the other hand, is relatively stable, suggesting that a significant fraction of the "extra" water entering the system runs off rather than infiltrating. Erosion gradually increases in a similar way to runoff, albeit with large interannual variations.

N and P can exit the system dissolved in the water that runs off or percolates, and attached to sediment particles. The quantity of P lost via these pathways is generally smaller than that of N, and most P tends to be lost with sediment. P losses with runoff and percolate are a few times smaller than the sediment losses, and generally quite similar to each other. Sediment is also the major (non-gaseous) loss pathway for N, with runoff and leaching being both generally smaller and more variable. Of the two dissolved N loss pathways, runoff usually dominates in any given year.

To understand the silage yield, it is helpful to know the basics of the crop growth component in APEX. Plants grow by accumulating heat units. The heat unit index (HUI) shown in Figure 2.10 is defined in the following way:

$$HU = 0.5 \times (TMX + TMN) - TBSC; HU > 0.0 \tag{2.1}$$

$$HUI = HU/PHU \tag{2.2}$$

where TMX and TMN are the daily maximum and minimum temperatures, TBSC is the base temperature for crop growth, and PHU is the number of heat units required

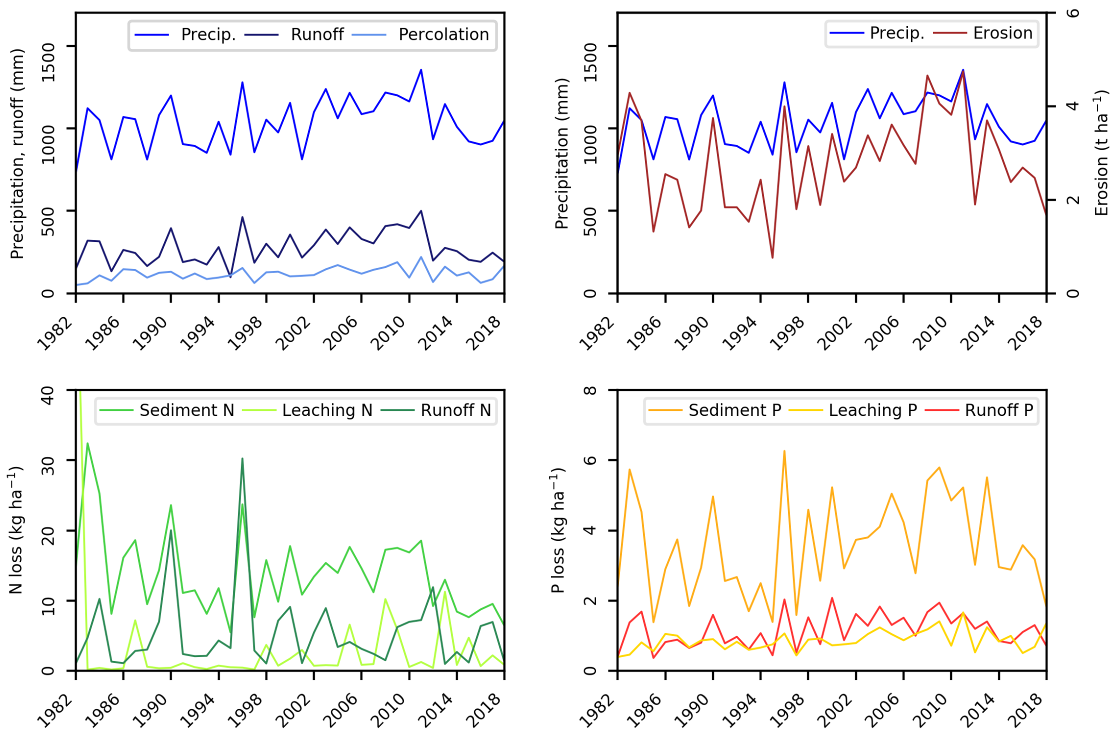

Figure 2.8: Annual precipitation, runoff, percolation, erosion, and N and P losses in the PAW1 baseline simulation.

for the crop to reach physiological maturity (Table 2.7; Williams et al., 2012).

Heat unit uptake translates into leaf area increase and dry matter production at a rate governed by various crop parameters (§2.2.4). This growth is then modulated by various sources of stress. Although APEX performs a number of complex daily calculations relating to soil concentrations and plant uptake of water and nutrients, these variables are translated into a set of relatively simple stress factors. For example, N and P stresses are a function of the actual N and P concentrations in plant tissue compared to their optimal concentrations. The stress factor with the lowest value becomes the "regulating stress" for that day, and the potential daily biomass increase is multiplied by that factor to give the actual daily increase.

The year-to-year corn silage yield is quite stable in the PAW1 baseline model (Figure 2.9). There is no strong relationship between yield and precipitation, and Figure 2.9 shows that in most simulation years there are relatively few days when water (drought) stress is the yield-limiting factor. There are no days of aeration (excess water) stress in the model, either[9].

Temperature stress is present every year to some degree, but the most common source of crop stress is insufficient N uptake. Most years have almost 50 days on which N stress is the limiting factor, which is about half the growing season for the 95 – 105 GDD corn varieties grown by the farmer at PAW1. It may be useful to understand the cause of the N stress because a farmer may then be able to take actions that limit it – or the model can be adjusted to reduce the stress, if it is judged to be unrealistic. In 1996, a year of especially high rainfall, high N stress and lower-than-usual crop yield appear to be caused by high levels of N loss in runoff and sediment. More generally, though, the N stress may be related to the large amounts of NH_3 volatilization predicted by the baseline model.

In the PAW1 baseline model, N is added when manure is spread in the spring and in starter fertilizer applied when the corn is planted. N application rates at these times are roughly 120 kg ha^{-1} (half of which is NH_4-N) and 60 kg ha^{-1} (urea form). Manure at PAW1 was incorporated on the same day as being broadcast, or the day after. Immediate chisel plowing can reduce manure NH_3 losses to a few per cent of NH_4N, while unincorporated manure may lose half of its NH_4N (Thompson and Meisinger, 2002). NH_4N losses of $\sim$20% from surface-applied urea have been reported (Jones et al., 2007), although losses at PAW1 are probably lower as in 3/4 years the

[9] As well as the stresses shown in Figure 2.9 APEX also calculates P, K, aeration (waterlogging) and salt stress. Days of stress for those factors are zero throughout the simulation period.

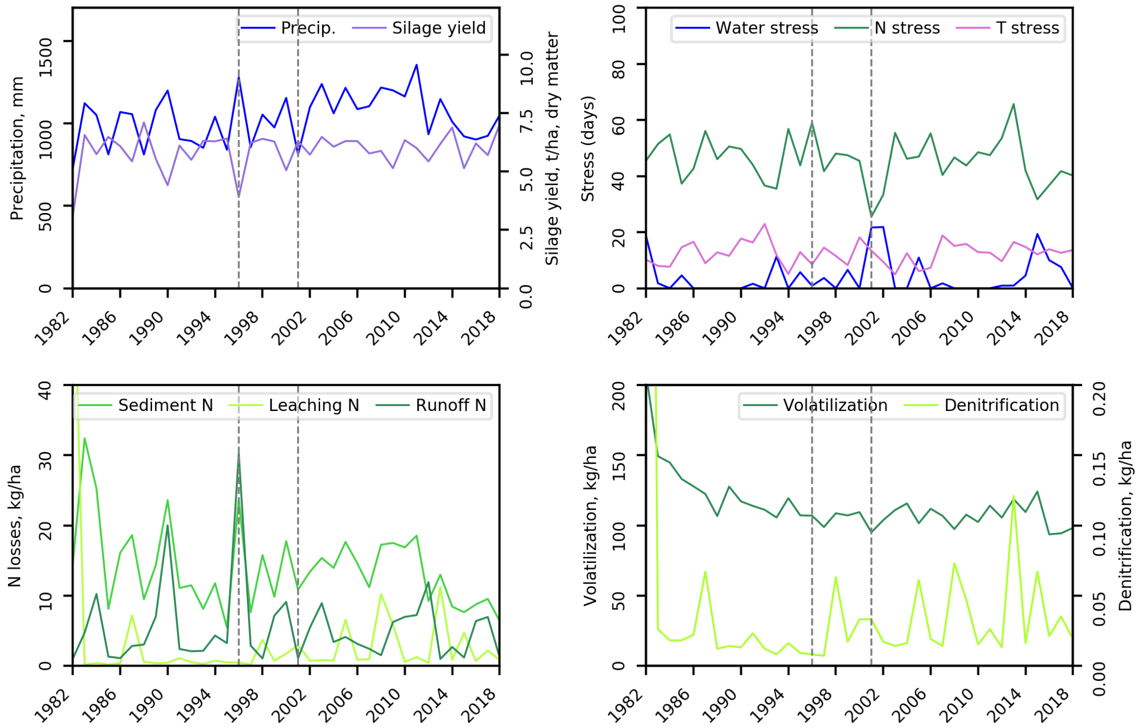

Figure 2.9: Annual precipitation, corn silage yield, crop stresses, and N loss mechanisms in the PAW1 baseline simulation. Stress days indicate the number of days each year in which a particular source of stress is the stress factor that limits crop growth (see text). Vertical dashed lines denote the years (1996 and 2001) that are examined in more detail in Figure 2.10.

starter fertilizer was applied through the corn planter.

If 50% of the manure NH4-N were lost as NH_3, and 20% of the starter fertilizer, that would imply total annual volatilization of ∼42 kg ha^{-1}. Even this number, which assumes management that encourages volatilization, is less than half the value predicted by APEX. This suggests that it will be necessary to adjust model parameters governing NH_3 volatilization during the calibration process.

The relationship between stress and crop growth (biomass production) is further illustrated in Figure 2.10. In 1996, a year of relatively low yield in Figure 2.10, severe

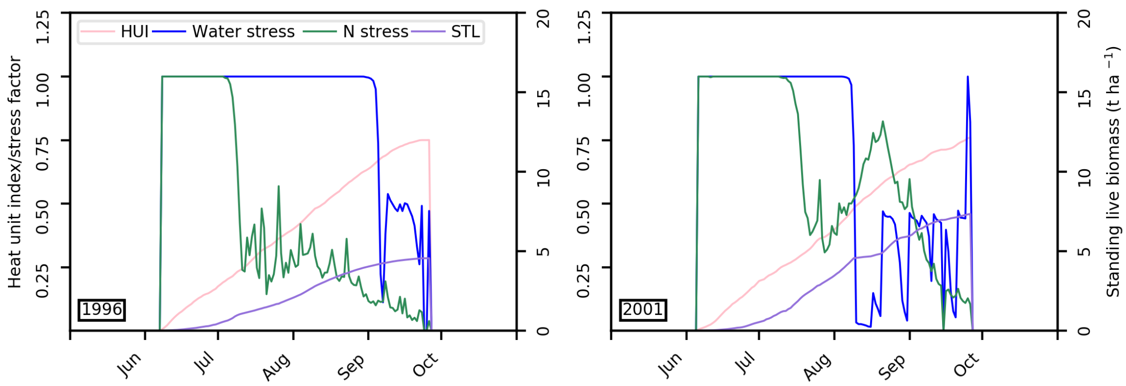

Figure 2.10: Daily heat unit index (HUI), water (drought) stress factor, N stress factor, and standing live crop biomass (STL) for the growing seasons in the PAW1 simulation years indicated in Figure 2.9. Low stress factors imply high crop stress. 1996: N stress begins early in the growing season, leading to depressed crop growth. 2001: N stress is less severe this year but this is partially balanced by drought stress in late summer. The heat unit index at harvest both years is lower than the expected value of 1.0 – 1.2.

N stress sets in in July during the period when the crop should be rapidly growing and taking up N (e.g. Alley et al., 2009). In 2001 when the yield is a little higher, N stress begins later and is less significant than in 1996. However, drought stress in late summer affects crop growth in this model year. Temperature stress is the only other stress that is ever significant in this model, but it generally only becomes the regulating factor for short intervals towards the start of the season. (For clarity, T stress is not shown in Figure 2.10).

The accumulation of heat units that drives the crop growth process depends on daily temperatures: the higher are TMAX and TMIN, the more HUs will build up (although biomass production may be reduced by T stress if temperatures climb too high). If the crop is harvested at maturity and the PHU parameter is set appropriately for the climate of the simulation, HUI should approach 1.0 in a typical year (Eq. 2.2). In fact, the User's Manual states that the heat unit index at harvest time "should normally range from 1. to 1.2”. However, the HUI values shown in Figure 2.10 do not reach 1.0. PHU in the PAW1 baseline model has been set to match the corn varieties reported for the PAW1 watershed. Nonetheless, the low HUI values calculated by APEX may suggest that PHU is another parameter that could be adjusted during calibration.

2.3.2 WIL2

WIL2 is the “control” watershed at the Williston study site. In many ways it resembles the PAW1 watershed. For instance, both farms grow silage corn fertilized with dairy manure, and the soil properties at both sites are quite similar (e.g. silt loam texture, 3–4% organic matter content, mildly alkaline pH; Braun et al., 2016). However, there are some potentially significant differences. The WIL site is rather flat, with a slope of just 0.6% compared to 4.5% at PAW1, decreasing the likelihood of erosion at this location. The two watersheds also differ somewhat in the methods used to grow the corn during the study.

On the WIL2 watershed, manure was applied in the spring, but usually at a higher application rate than on PAW1 (§2.2.4). The manure was incorporated on the same day as it was broadcast or within the next few days, and a second round of tillage was usually performed shortly before the corn was planted around two weeks later. Unlike Pawlet, no starter fertilizer was used on the Williston watershed.

Following corn harvest in late September, a cover crop was seeded on both watersheds. As with Pawlet, the cover crop generally failed to establish during the APME project. Unlike Pawlet, though, manure was surface applied at Williston at some point between late October and early December, this time without being in-

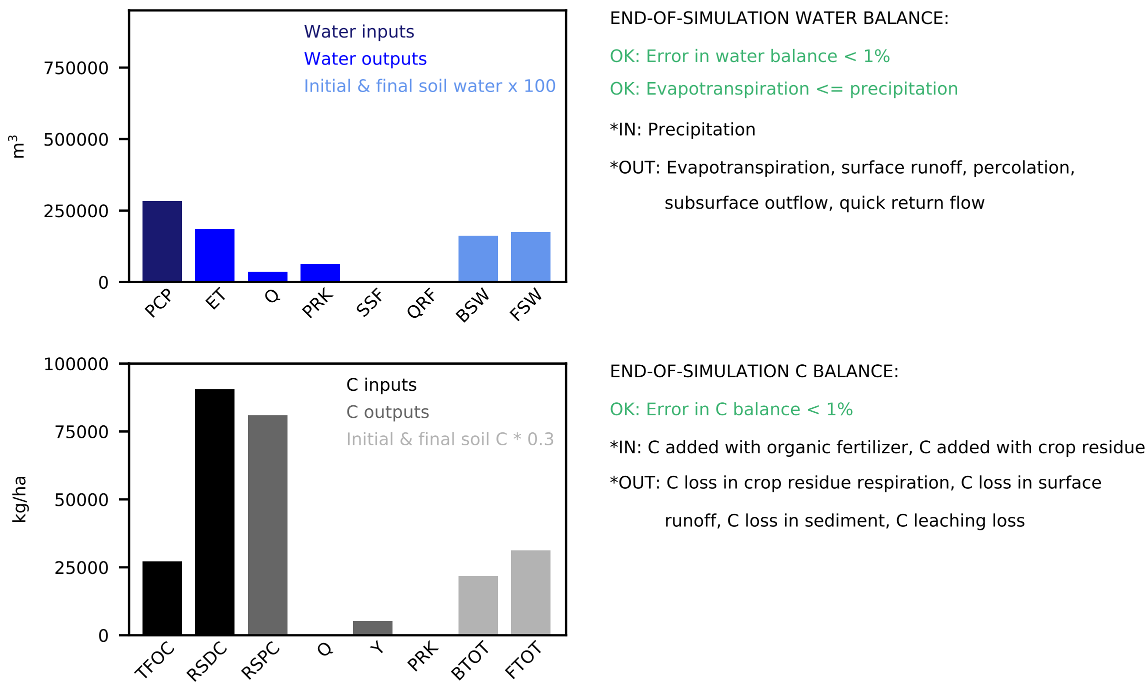

Figure 2.11: Summary of water and carbon inputs and outputs over the course of the WIL2 baseline simulation. For ease of comparison, these plots are on the same scale as Figure 2.6.

corporated. Again, application rates were relatively high. The differences in manure application rates and incorporation, in conjunction with the different weather experienced at both sites, may be reflected in the nutrient losses predicted by the baseline models.

Results from the WIL2 baseline model, aggregated over the whole 37-year simulation period, are shown in Figure 2.11. They differ from the PAW1 baseline model in several respects. First, WIL2 received much less precipitation than PAW1 over the duration of the simulation. Also, although evapotranspiration is still the main loss pathway, percolation is more important than runoff at WIL2. This is true both in aggregate (Figure 2.11) and on an annual basis (Figure 2.13); there are only a handful

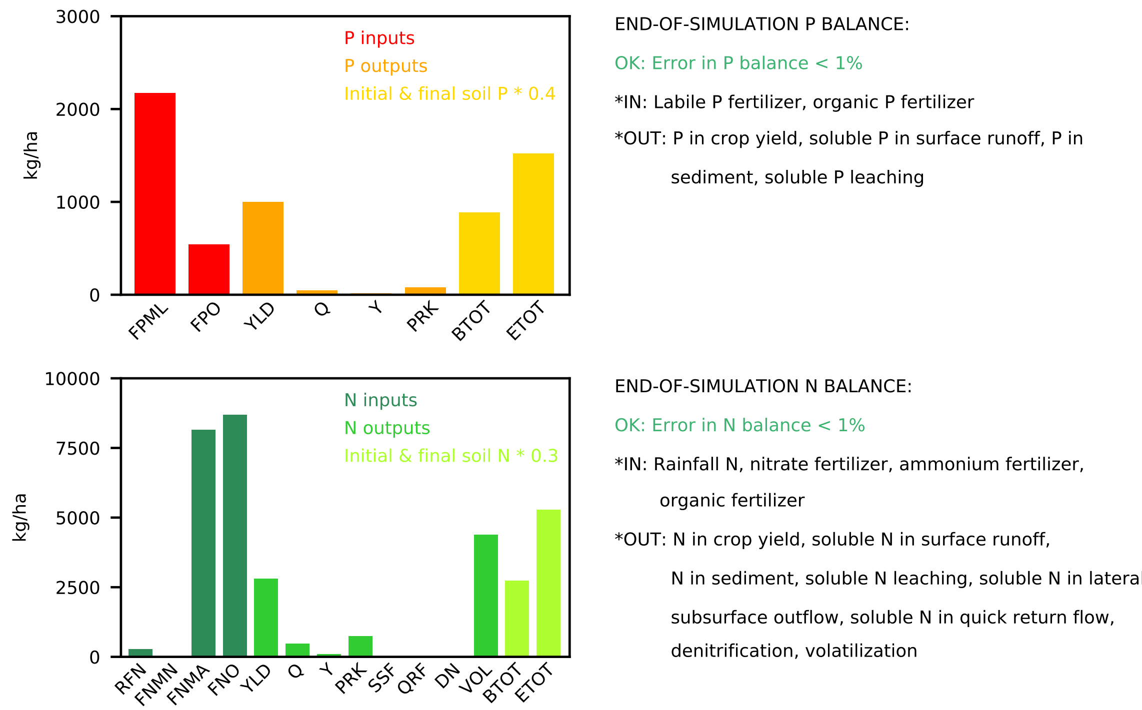

Figure 2.12: Summary of P and N inputs and outputs over the course of the WIL2 baseline simulation. For ease of comparison, these plots are on the same scale as Figure 2.7.

of years in which runoff exceeds percolation at this site, notably the high-precipitation years of 2011 and 2013.

Organic fertilizer accounts for a much larger fraction of C inputs at WIL2 than at PAW1, and carbon stocks build up over the course of the simulation instead of being depleted. P and N also build up to a larger extent than at PAW1 (Figure 2.12). All of these phenomena are presumably due to the larger amount of manure used on the WIL2 watershed compared to PAW1.

The distribution of N and P between loss pathways also differs between the two corn simulations (Figure 2.12). While P losses in sediment dominated at PAW1, the relative importance of P dissolved in percolate and runoff increases at WIL2. Total P

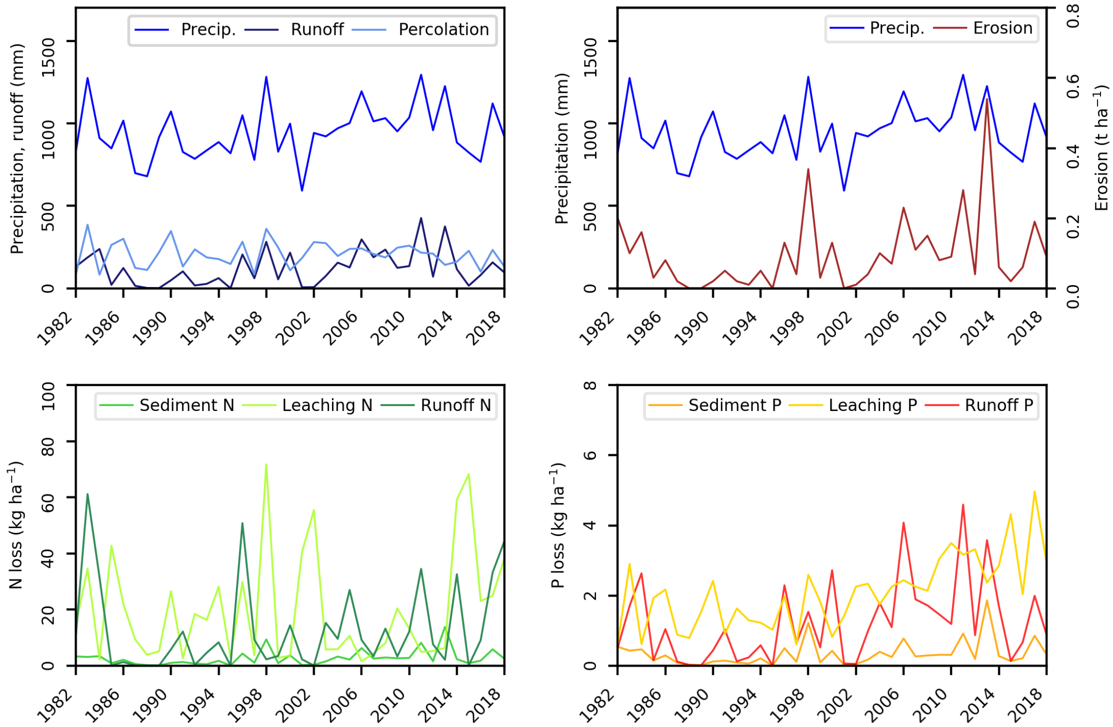

Figure 2.13: Annual precipitation, runoff, percolation, erosion, and N and P losses in the WIL2 baseline simulation.

removal in crop yield is also higher at WIL2, likely reflecting higher yields (see below). Similar trends are seen for N losses, with dissolved N becoming more important at WIL2 (although volatilization losses at both sites are high). The small amounts of sediment-bound N and P loss (and the flat site) suggest that erosion is low at this farm, and Fig. 2.13 confirms that this is the case.

Looking at the data on an annual basis shows that the relative importance of N loss pathways at WIL2 is highly variable. In 2011, for example, N losses on the order of 40 kg ha^{-1} are dominated by runoff N. In 2015, on the other hand, losses of over 60 kg ha^{-1} are almost entirely due to leaching. The dominant P losses at WIL2 also fluctuate between runoff and leaching, and overall P losses increase in the second half

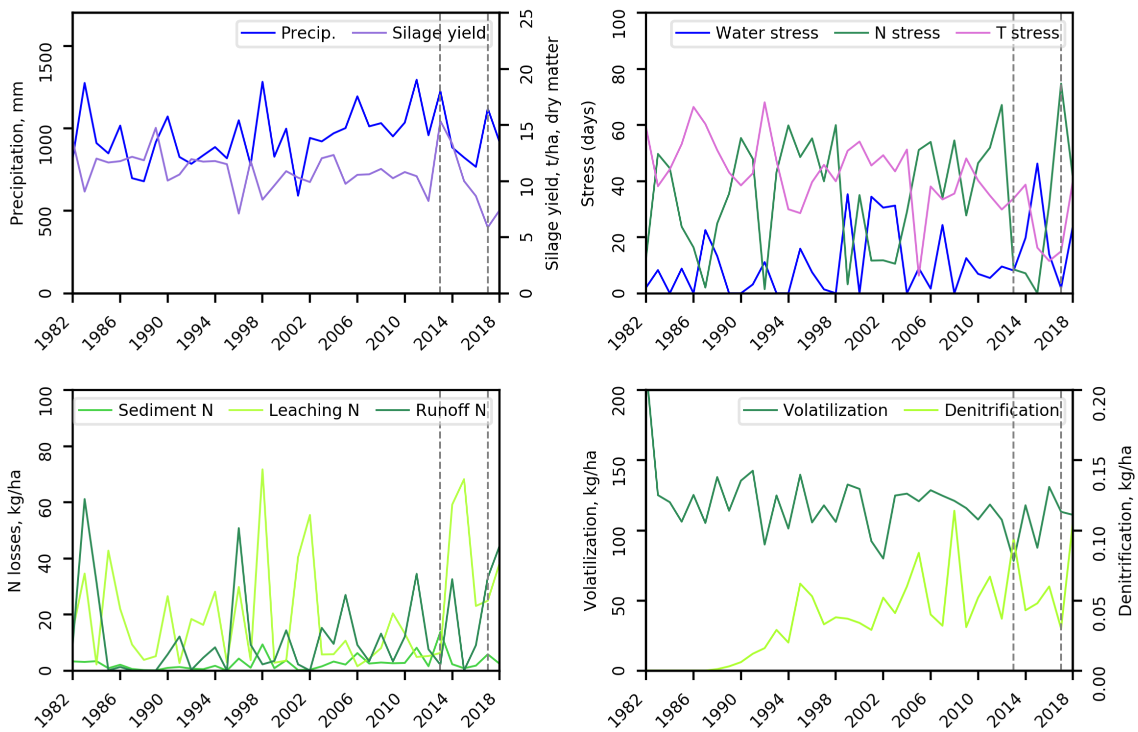

Figure 2.14: Annual precipitation, corn silage yield, crop stresses, and N loss mechanisms in the WIL2 baseline simulation. Stress days indicate the number of days each year in which a particular source of stress is the stress factor that limits crop growth. Vertical dashed lines denote the years (2013 and 2017) that are examined in more detail in Figure 2.10.

of the simulation. This may be partly due to the buildup of soluble P in the soil as the simulation progresses (Figure 2.5).

The predominant crop stress at PAW1 was N stress, and N stress is also significant at WIL2. This may well reflect the fact that NH_3 volatilization losses are of a similar magnitude on both watersheds (Figure 2.14). The number of days where drought stress regulates growth is generally fairly low, although it can occasionally be comparable to N stress.

In contrast to PAW1, temperature stress is often a major factor at WIL2. T

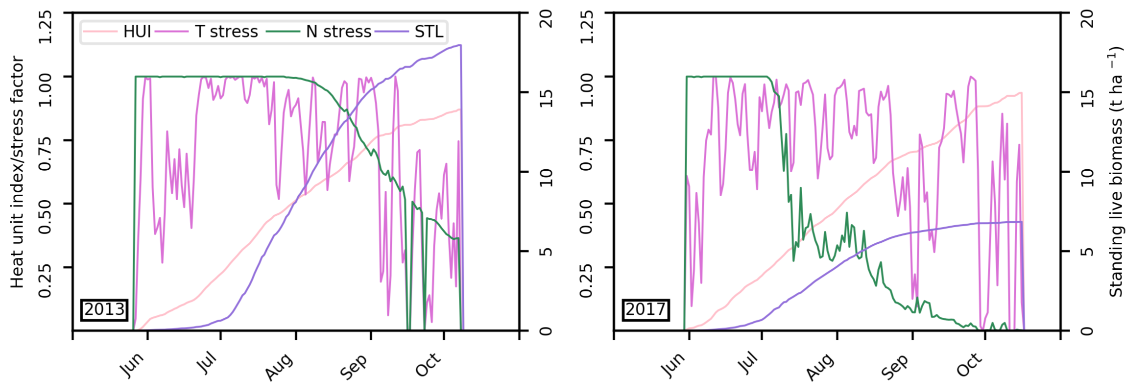

Figure 2.15: Daily heat unit index (HUI), temperature stress factor, N stress factor, and standing live crop biomass (STL) for the growing seasons in the WIL2 simulation years indicated in Figure 2.14. For clarity, drought/water stress is omitted. Low stress factors imply high crop stress. 2013: The crop experiences moderate temperature stress early and late in the growing season, but N stress is low and yields are high. 2017: Severe N stress begins in July and persists until harvest, leading to depressed yields. The heat unit index at harvest both years is a little lower than the expected value of 1.0 – 1.2, but higher than at Pawlet (Figure 2.10).

stress occurs when the mean daily temperature is above or below the crop's optimal growing temperature, and in these Vermont simulations it tends to be highest early and late in the season when temperatures are relatively low. T stress decreases with time at WIL2, which could be a reflection of the warming climate. However, it could alternatively indicate that water and N stress became more severe towards the end of the simulation, leading to fewer days when T stress was the regulating factor.

Corn yields at WIL2 are substantially higher than at PAW1, although also more variable. The higher yields are perhaps surprising given the large number of days on which stress is experienced at WIL2. However, it is possible that if the timing of the stress is favorable and the severity is low, this could offset the larger number of days when stresses occur. The corn crop at WIL2 also tends to take up slightly more heat units than the corn at PAW1 (Figure 2.15), which will also increase growth.

2.4 Baseline Models: Comparison with Observations

§2.3 gave an incomplete but perhaps illuminating description of the baseline simulations, illustrating the relative importance of several processes and their influence on some important model outcomes. The next step is to compare the models with reality as represented by the APME measurements. This section briefly summarizes the overall characteristics of the runoff and water quality measured at PAW1 and WIL2, compares the model output to these data using a set of graphical model performance indicators, and quantifies the model performance in terms of the total observed and modeled runoff etc. at the end of the APME monitoring period. (More detailed and quantitative model performance statistics will be introduced in Chapter 3.)

Comparing the models and data requires reconciling the different timescales upon which each one is based. The APME data represent discrete runoff events that could last <1 day or for multiple days, but APEX reports results on a daily basis. To compare model and data, APEX output was summed over the number of days included in each runoff event. No attempt was made to account for partial days (e.g, if an event began on a Wednesday night and ran through Thursday, all of Wednesday's APEX output was included in the sum). However, to avoid double-counting, when an event ended on the same day as another event began (e.g. an event ended on a Wednesday morning and another event started on the Wednesday afternoon), both the APME data and APEX output for the two events were merged into a single event. Reconciling these different time scales is another limitation of attempting to model

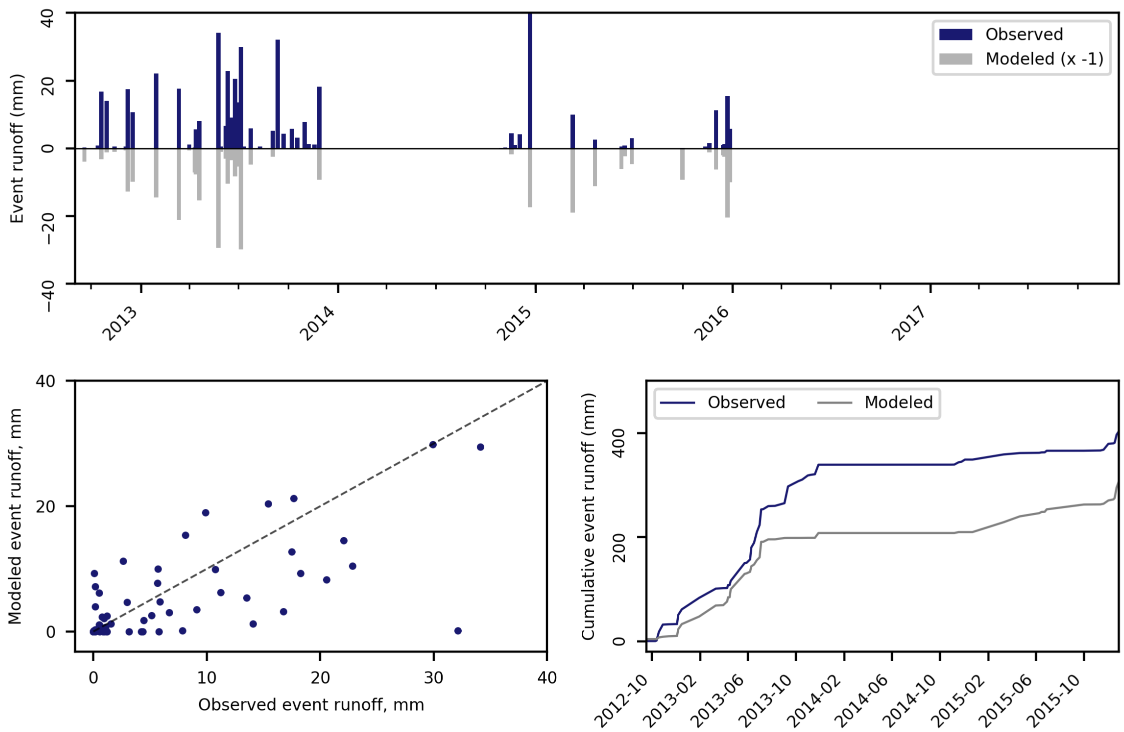

Figure 2.16: Upper panel: time series plot of observed and modeled runoff events on the PAW1 watershed. Lower left: scatter plot of modeled vs observed event runoff. The dashed line denotes the 1:1 relation. Lower right: cumulative plot of observed and modeled runoff. Note that the bars in the time series plots are of a constant width that does not reflect the duration of each event.

field data that were not collected for the specific aim of calibrating simulations.

2.4.1 PAW1

Runoff (Q)

Runoff and water quality monitoring at PAW1 began in late 2012 and ran through 2015, but was suspended during the spring and summer of 2014 due to a deviation from the planned management. Therefore two distinct periods of monitoring data

– ∼2013 and ∼2015 – are available at this site The frequency and volume of runoff events was quite different between these two periods (Figure 2.16).

Frequent, high-volume runoff events took place at PAW1 in 2013. Braun et al. (2016) note that across Vermont, the period between late May - early July 2013 contained "record-breaking rainfall totals, saturated soil conditions, and large runoff events". At PAW1, above-average rainfall was experienced in May, June, July, and September. Apart from the large event in December 2014 (see below), the highest-volume runoff events of the monitoring period occurred during those months.

The largest runoff event at PAW1 occurred between December 22nd – 29th, 2014. Braun et al. (2016) describe this as "the major Christmas rain-on-snow event". Despite a number of experiments with APEX, no model was ever found that could come close to predicting the large magnitude of this event while also making reasonable predictions for the remaining events. In addition, it is not clear that APEX's runoff algorithms can correctly handle rain falling on snow[10]. *For completeness, the "Christmas event" is shown in the time series plots in this Chapter. However, it is excluded from the scatter and cumulative plots in the figures in this section, and will also be excluded from all of the calibration plots and statistics presented in Chapter 3.*

In 2015, rainfall in April and May was below average. June, however, was wetter than usual; almost 60% above the long-term average precipitation. However, the runoff events that occurred at PAW1 in June 2015 were small in comparison with the events of mid-2013, perhaps because the preceding dry months allowed greater

[10]In the baseline models, APEX uses the NRCS curve number (CN) method to predict daily runoff. The curve number each day is calculated based on a "soil moisture index" that takes into account precipitation and potential evapotranspiration (§3.3.1). When rain falls on snow, that day's effective CN should be high. However, although APEX does increase runoff for frozen soil, its daily CN calculation methods do not appear to take into account the presence of snow on the ground, suggesting that the code cannot be expected to correctly predict runoff arising from rain on snow.

infiltration into the soil. In contrast with 2013, the biggest runoff events at PAW1 in 2015 took place in November and December.

The time series plot in Figure 2.16 can reveal how well the model reproduces these general features of the data, and can help identify trends in performance and data points that are particularly problematic for the model. The figure shows that there are several points in both the 2012 – 2013 and 2014 – 2015 periods where the PAW1 baseline model reproduces the observed runoff with reasonable accuracy. The general shape of the model follows the rises and falls in the data relatively well in December 2012 – June 2013, and again in November – December 2015.

However, there are also places where the baseline model over- or under-estimates the runoff. The model strongly underestimates several events in roughly September – November of both 2012 and 2013, but it overestimates most of the small, infrequent events in 2015 (until November). It is possible that the model calculates appropriate runoff curve numbers when the soil is consistently fairly moist, but picks too high a curve number (predicts too little infiltration) when the soil dries out.

The scatter plot in Figure 2.16 shows that the observed and modeled runoff are correlated, but it also shows that the model tends to underestimate more frequently than it overestimates. The general tendency for the model to predict too little runoff is emphasized by the cumulative graph, which makes it obvious that the model predicts too little runoff in total.

Erosion (Y)

Runoff is one of the factors that determines water erosion, and the plot of erosion vs time (Figure 2.17) demonstrates that, to zero order, the observed sediment loss

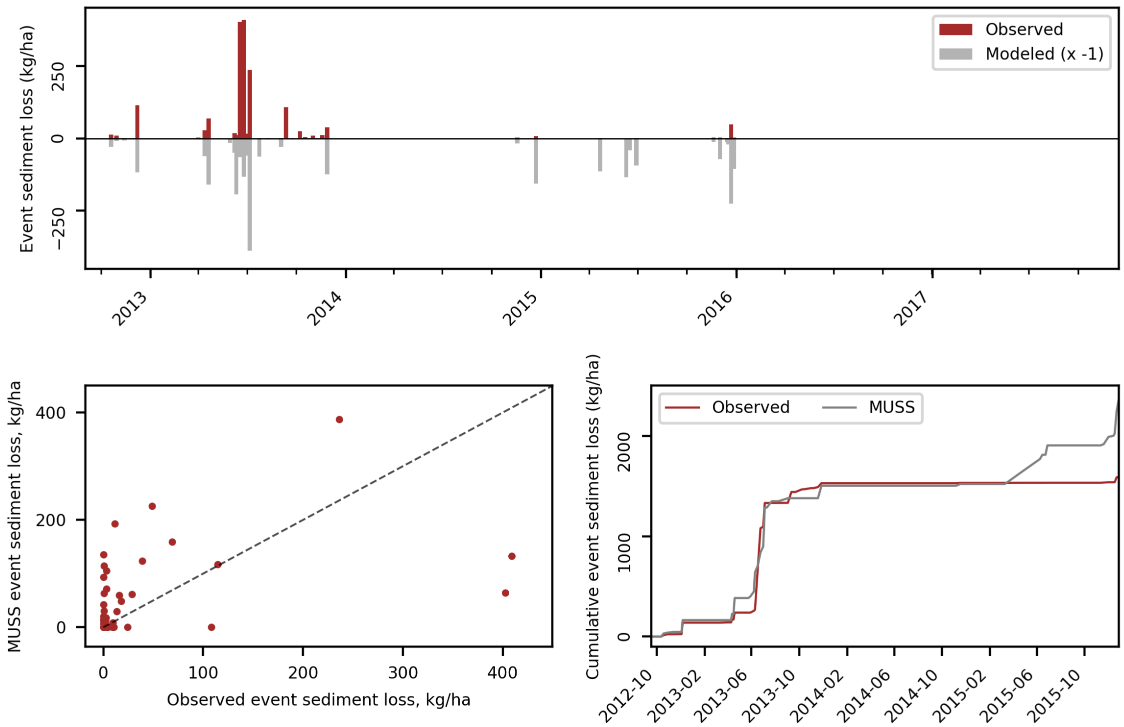

Figure 2.17: As for figure 2.16 but for erosion (PAW1 watershed). The baseline model uses the MUSS equation to calculate sediment loss.

at PAW1 changes between 2013 and 2015 in a manner that is broadly similar to runoff: the largest and most frequent erosion events occurred in 2013, while very little erosion occurred in 2015[11]. The relative magnitude of the erosion events bears some resemblance to that of the runoff events, but the correspondence is far from exact. For example, while one of the largest runoff events of 2013 occurred in September of that year, relatively little sediment was lost at that time.

[11]There are fewer data points in Figure 2.17 than in Figure 2.16 because not all runoff samples were sent for laboratory water quality analysis. Also, as explained at the beginning of §2.4, data from some adjacent runoff events was combined into a single event to avoid "double counting" problems. In a few cases, water quality data were only obtained for one of the combined events. These points may appear to have large runoff volumes accompanied by small sediment and nutrient masses. This does not affect the comparison of model vs data for a single variable, but it should be borne in mind when using these figures to compare runoff with sediment and other water quality variables.

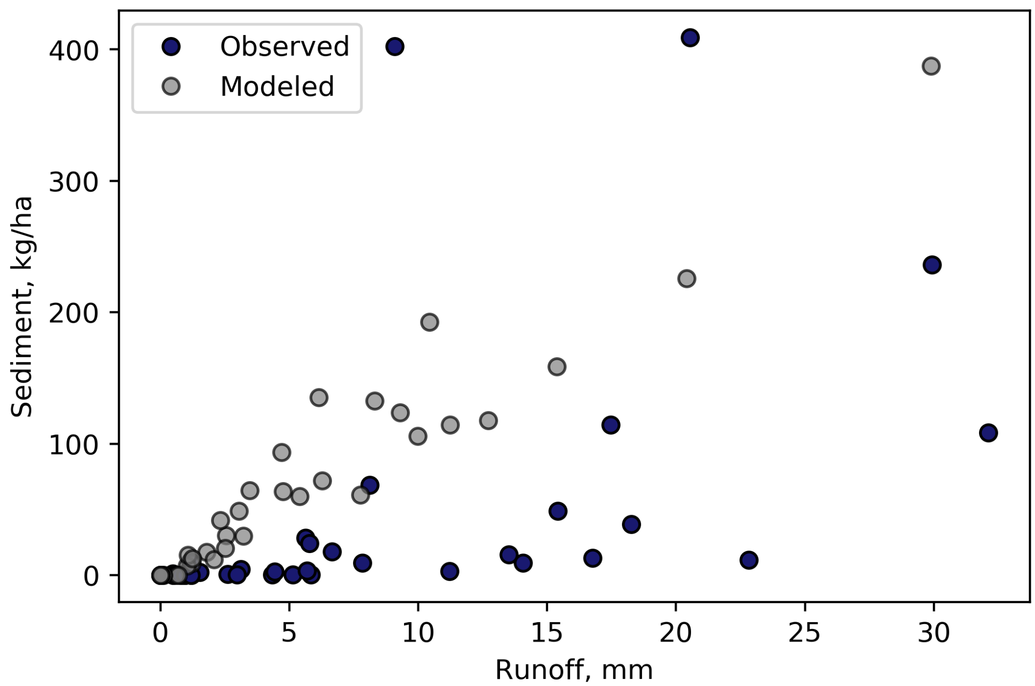

Figure 2.18: Relation between sediment and runoff at PAW1, in both the field measurements and the baseline APEX model.

Figure 2.18 gives a closer look at the connection between runoff and erosion. In the field, most runoff events generated little sediment, and the largest quantities of sediment did not necessarily result from the largest runoff events. In fact, there was no clear relation between runoff and sediment at the PAW1 site. In the APEX model, however, runoff and erosion are quite strongly related and the model generally expects a higher ratio of sediment to runoff than was actually observed.

As with runoff, the baseline model makes better predictions for erosion in 2012 – 2013 than in 2014 – 2015. The model both over- and under-estimates in 2013, but the cumulative plot shows that the total sediment loss at the end of 2013 matches the observed value very well. In 2014 – 2015, where relatively weak runoff events

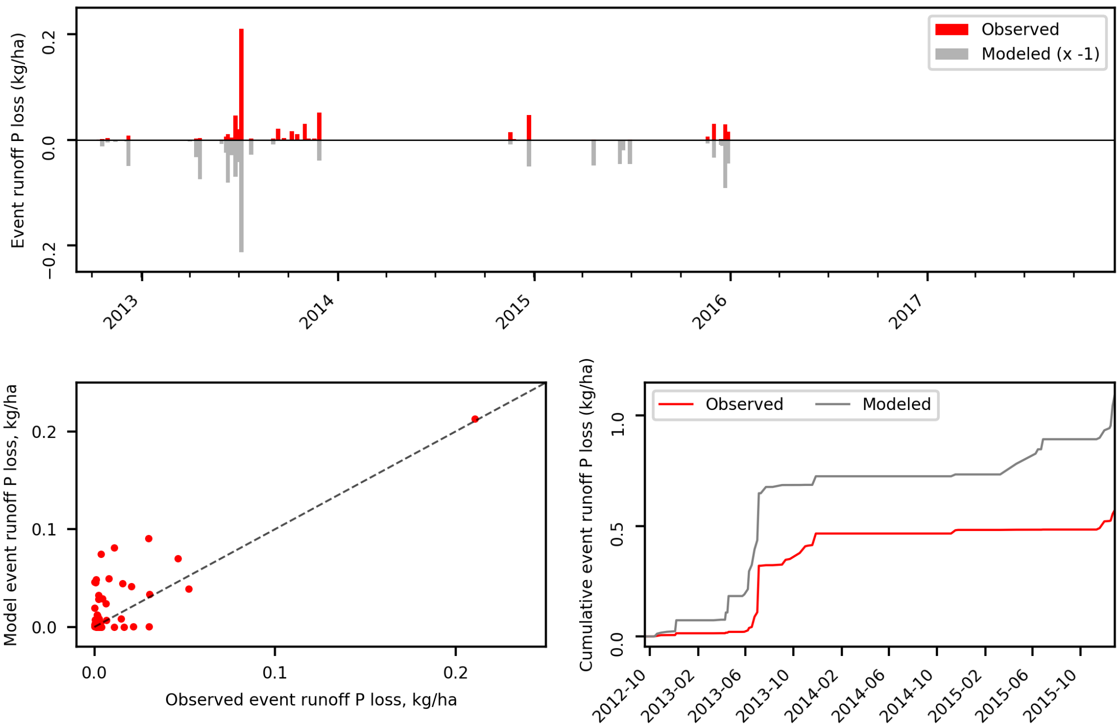

Figure 2.19: As for figure 2.16 but for P dissolved in runoff (PAW1 watershed).

generated little sediment loss, the baseline model systematically overestimates. At the end of the monitoring period, APEX has predicted 2350 kg/ha of sediment loss whereas only 1600 kg/ha actually occurred, an excess of almost 50%.

Phosphorus losses (QP, YP)

The pattern of runoff P loss (QP) observed in the field (Figure 2.19) is rather different from the pattern of runoff itself. Even in 2013, most QP events are relatively small. The exception is a single event in early July, in which several times as much QP was lost as in even the next largest event. This event did not coincide with any reported farm activities – it occurred roughly two months after manure was spread

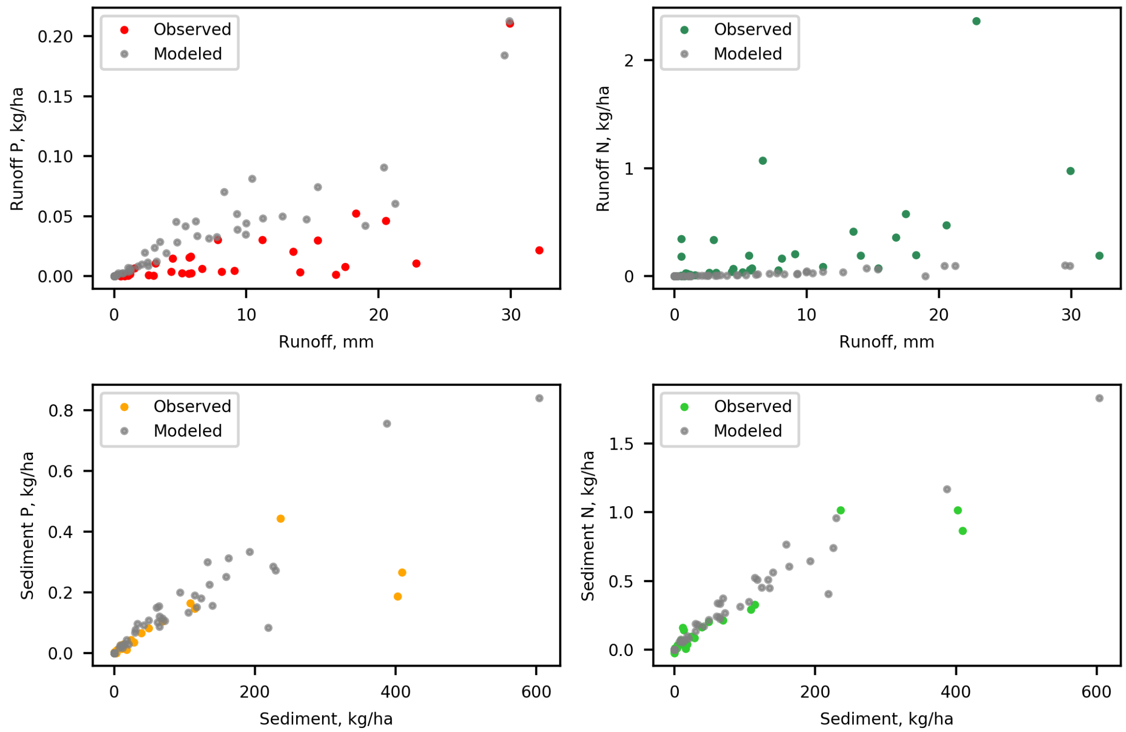

Figure 2.20: Model and observed relationships between nutrient losses and runoff (top) and sediment (bottom), on the PAW1 watershed.

and incorporated (May 2-3) and corn planted and fertilized (May 8), and during the prolonged period of above-average rainfall at Pawlet that year.

Figure 2.20 confirms the rather weak relation between QP and Q in the field data. The model relation between QP and Q is perhaps somewhat stronger, but the model expects a higher ratio of runoff P to runoff than was observed. This, together with the modest overestimate of Q, leads to the total QP loss over the monitoring period being overestimated by about 90%. Perhaps encouragingly, the model is able to reproduce the major July P loss event. However, calibrating the model to better match the total QP loss may change that.

In contrast to QP and Q, the measured P losses with sediment (YP) closely

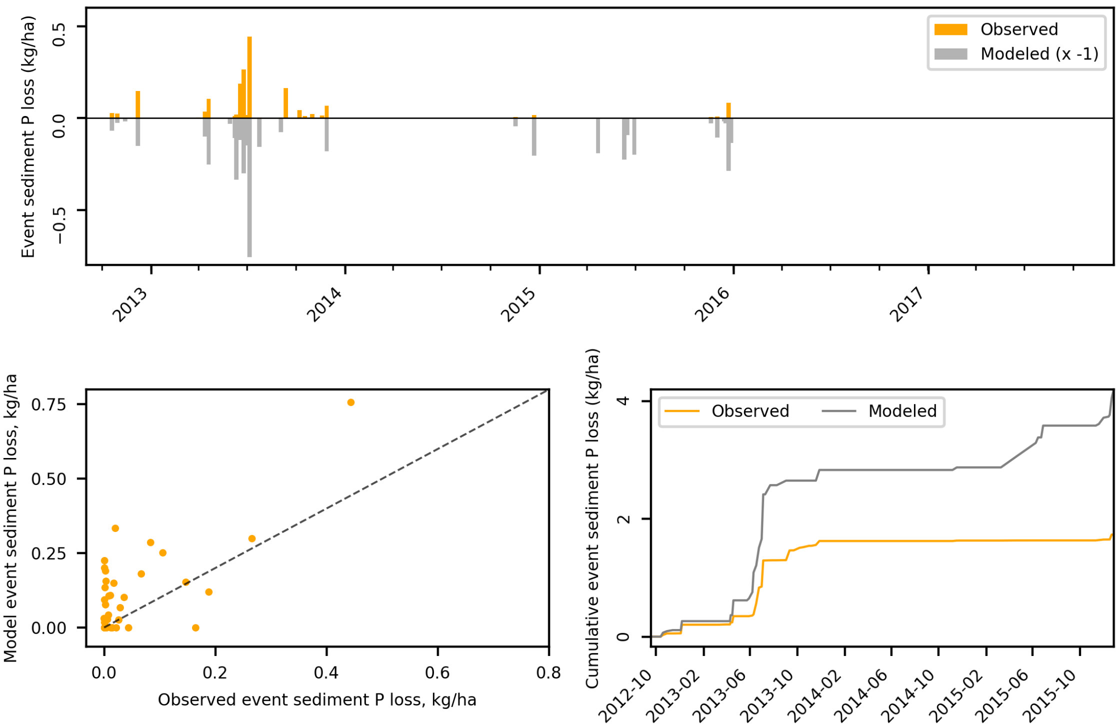

Figure 2.21: As for figure 2.16 but for sediment-bound P (PAW1 watershed).

follow the pattern of sediment loss (Y) itself (Figure 2.20). The baseline model captures this relation very well. In total, though, the model predicts about 150% more YP loss than was observed during the APME project (Figure 2.21). While this is undoubtedly related to the model's overprediction of sediment losses, the 150% excess of YP exceeds the ≈50% sediment excess.

In summary, the baseline model for PAW1 overestimates both QP and YP by a large amount. However, it does correctly predict that more P is lost in sediment than in runoff at this watershed.

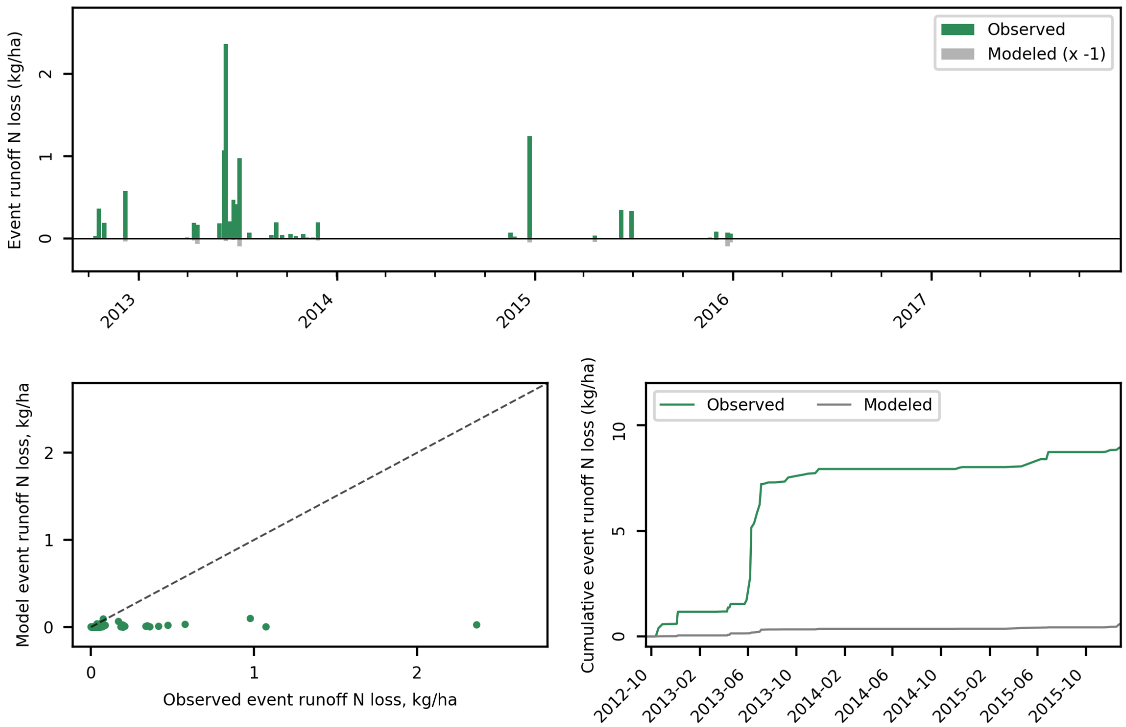

Figure 2.22: As for figure 2.16 but for N dissolved in runoff (PAW1 watershed).

Nitrogen losses (YN, QN)

Runoff N (QN) losses at PAW1 were about 15 times higher than QP losses, and the ratio of QN to QP differed greatly between events. For example, runoff events in late June and early July 2013 were accompanied by high and low rates of QN loss, respectively, but loss of QP was low in the June event and very high in the July event. Also, as Figure 2.20 shows, there is considerable scatter in the relationship between QN and Q.

The baseline model fails to predict QN losses of the observed magnitude for all but the smallest events. The total N loss over the monitoring period was almost 9 kg ha^{-1}, whereas the model calculates a loss of only 0.6 kg ha^{-1}. The reason for this is

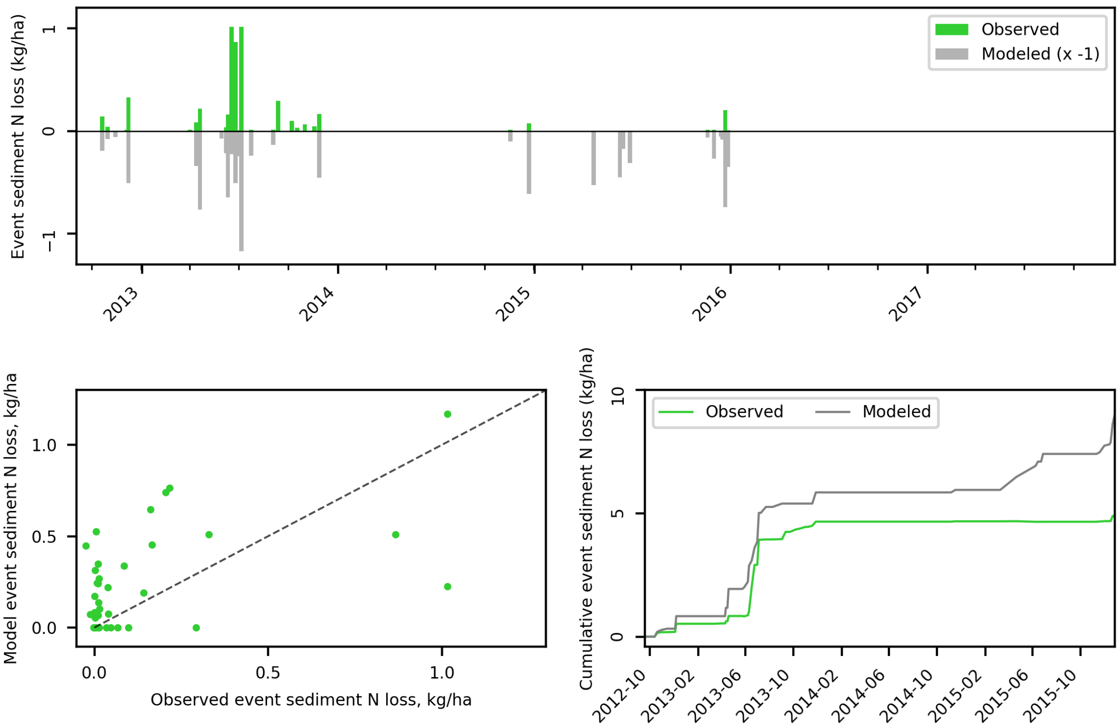

Figure 2.23: As for figure 2.16 but for sediment-bound N (PAW1 watershed).

not clear to this author at this point. However, it may be that the model is directing too much N into volatilization and other loss pathways, leaving little available to be lost via runoff.

Like YP, sediment-bound N (YN) losses at PAW1 closely followed the pattern of sediment loss (Figure 2.20). And as with runoff N and P, sediment N losses exceeded P losses. The APEX predictions for YN resemble those for YP, just generally scaled up in magnitude. That is, the model predictions for N loss in sediment are clearly related to the predictions for sediment loss itself. As with YP, APEX overestimates the total YN loss during the monitoring period, although by only 80% compared with the 100% overestimate for YP.

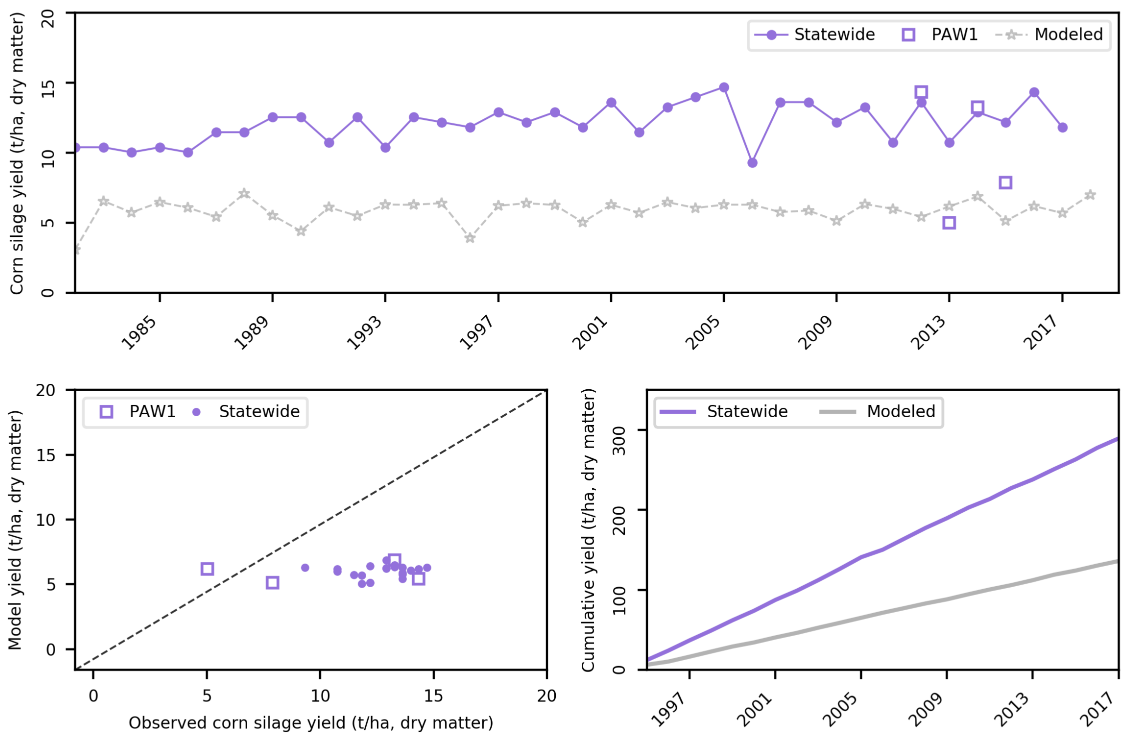

Figure 2.24: Upper panel: time series plot of model corn silage yield, statewide historical average yield, and the yields reported by the farmer for the PAW1 watershed during the APME project. Lower left: Model yield vs statewide historical and farmer-reported yield. Lower right: Cumulative model and statewide historical yields. The scatter and cumulative plots contain only data from 1997, conservatively allowing for a 15-year model run-up period.

Yield

Figure 2.24 shows the statewide average yield records obtained from the National Agricultural Statistics Service[12], and the yields estimated by the farmer at PAW1 during the APME project. The farmer-reported yields varied from 7 – 22 tonnes/ha (dry matter), and experimentation with APEX did not result in any models that produced year-to-year variation of this magnitude. This may be due to factors not taken into account in the baseline APEX model, such as pest damage (although no insect, disease, etc. problems are mentioned in the APME report). For the purpose of evaluating and calibrating the models, then, the reference yield value will be the long-term Vermont historical mean yield, which is intermediate between the low and high values reported at the study sites. Throughout the baseline simulation, the model silage yield is roughly half of the historical mean yield.

2.4.2 WIL2

Runoff (Q)

The WIL2 watershed experienced fewer runoff events than PAW1, even though the monitoring period at the Williston farm was longer. Apart from an event in May 2013 that produced ∼35 mm at both sites, runoff volumes at WIL2 also tended to be smaller than at PAW1. In common with PAW1, runoff occurred quite frequently during the unusually wet conditions in May and June of 2013. Runoff was detected with quite high frequency (but fairly low magnitude) over the winter of 2015 – 2016 as well. Other than this, events at WIL2 were sporadic and/or small.

Unlike at PAW1, where the baseline model fitted at least the 2013 data reasonably well, the WIL2 baseline model is only able to correctly reproduce a few of the very small runoff events. There is perhaps a tendency for the model to over-estimate events in the first ∼half of the year while under-estimating events later in the year, but on the whole it is about as likely to predict too much runoff as too little. The total quantity of model runoff exceeds the observed value by a factor of three.

[12] www.nass.usda.gov/. County-level corn silage production records are not available for Vermont.

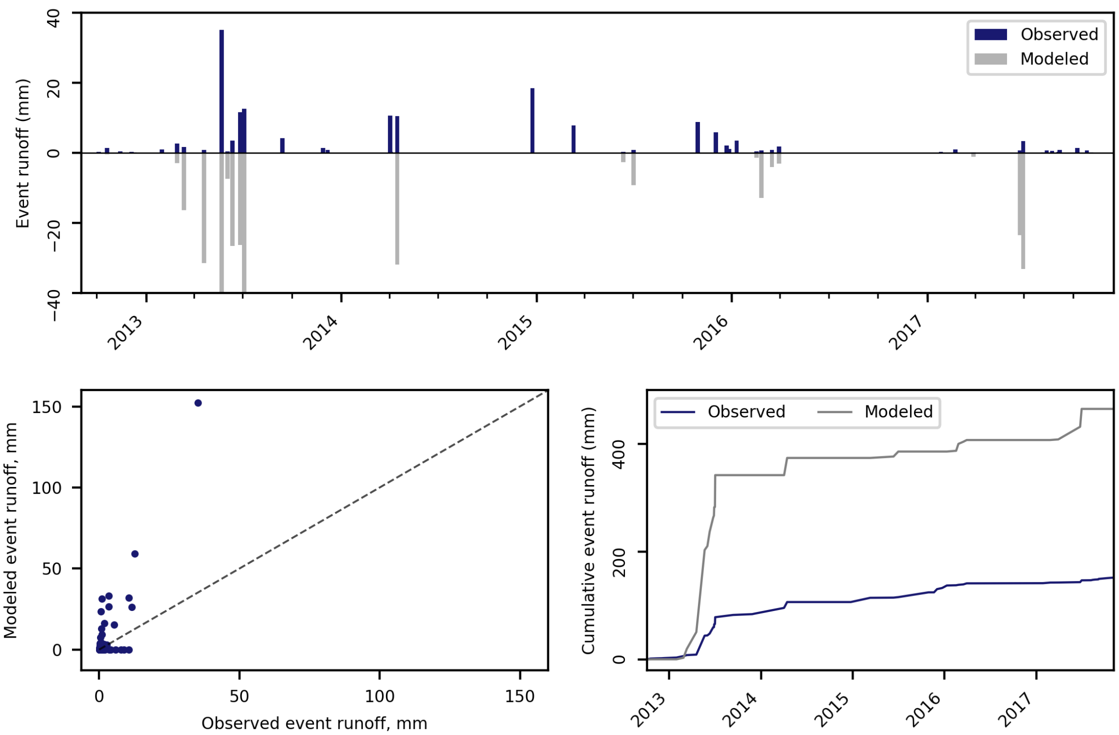

Figure 2.25: Upper panel: time series plot of observed and modeled runoff events on the WIL2 watershed. Lower left: scatter plot of modeled vs observed event runoff. The dashed line indicates the 1:1 relation. Lower right: cumulative plot of observed and modeled runoff.

Erosion (Y)

The measured sediment losses at WIL2 followed the pattern of runoff quite closely (Figure 2.26). The relation between runoff and sediment in the APEX model is also fairly clear. In contrast to PAW1, the baseline WIL2 model systematically *underestimates* sediment losses relative to runoff. However, because the model overestimates runoff by such a large factor, the total model sediment loss is also too high (Figure 2.27).

Compared to its predictions for runoff, the WIL2 baseline model results for sedi-

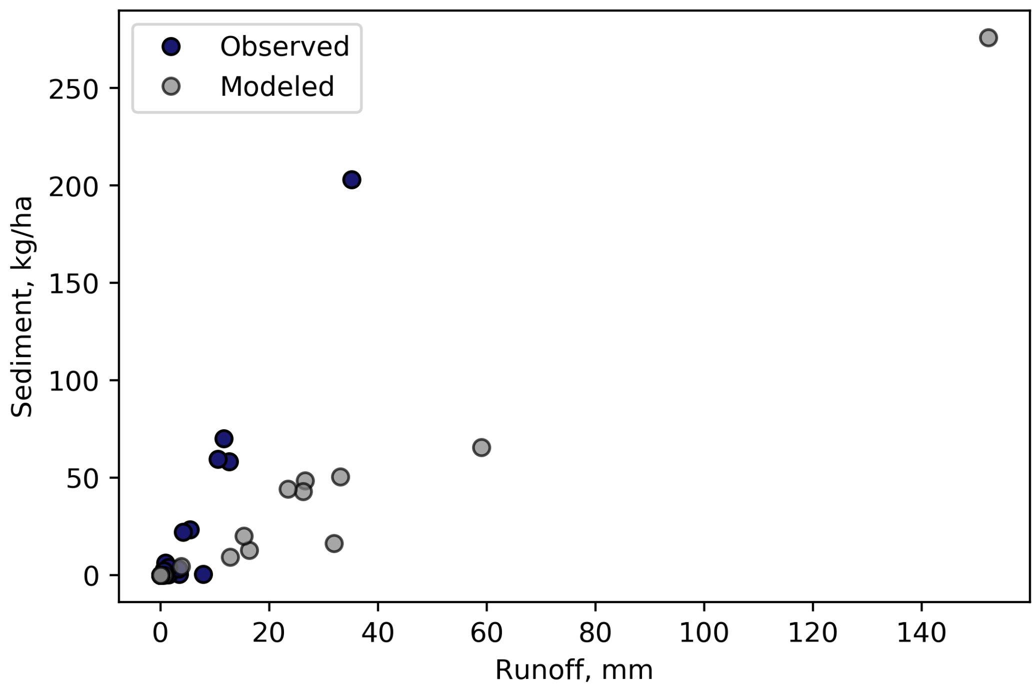

Figure 2.26: Relation between sediment and runoff at WIL2, in both the field measurements and the baseline APEX model.

ment arguably bear more resemblance to the data, particularly in 2013. TSS measurements after 2013 are sparse, however, so the performance of the WIL2 erosion model is difficult to evaluate during that period. In total, the baseline model overestimates sediment losses by 26%.

Phosphorus losses (QP, YP)

Contrary to PAW1, dissolved P losses measured at WIL2 roughly follow the pattern of the runoff itself (Figure 2.28). The model runoff P losses also track the model runoff, which was poorly modeled, meaning that the model does a poor job of predicting QP losses (Figure 2.29). At the end of the monitoring period APEX estimates total QP

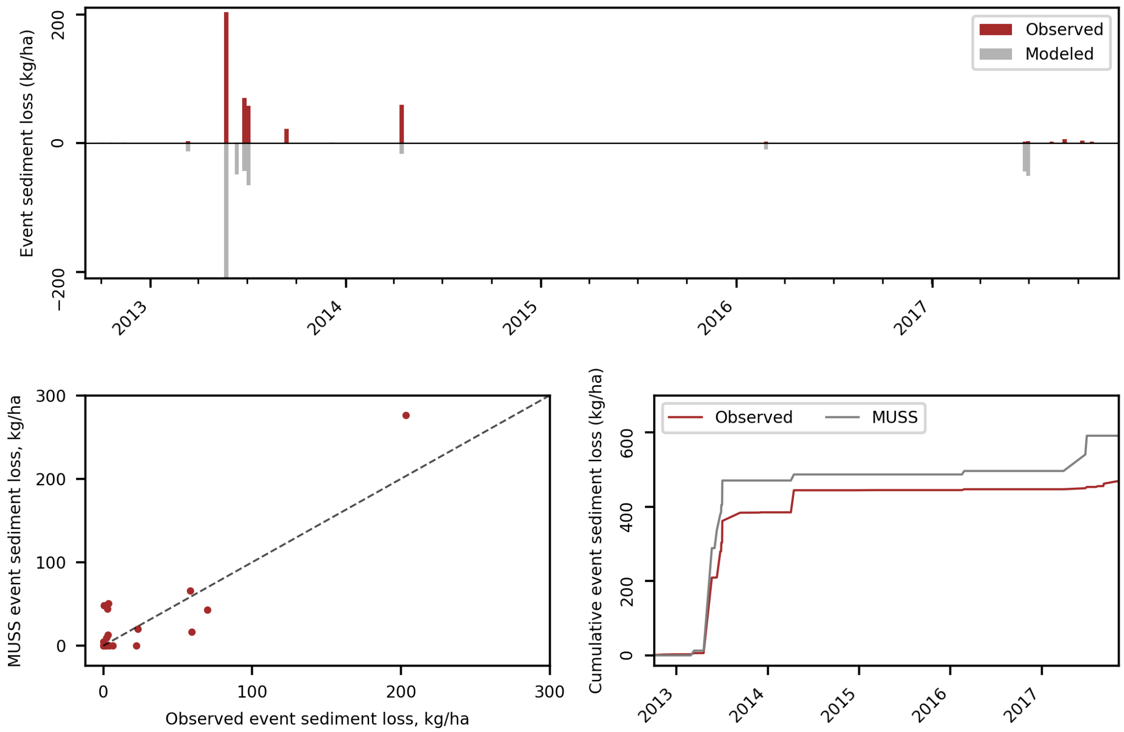

Figure 2.27: As for figure 2.25 but for erosion (WIL2 watershed). The baseline model uses the MUSS equation to calculate sediment loss.

losses of 4.0 kg ha^{-1}, compared to the observed value of 0.73 kg ha^{-1}.

The observed sediment P losses are quite closely related to the observed TSS measurements, and the same is true for the modeled YP (Figure 2.28). However, the model relation appears steeper than the observed one. The modeled sediment losses are too high, and the modeled YP losses even more so (Figure 2.30). At the end of the APME project, APEX has overestimated the total YP loss by 230%.

Nitrogen losses (QN, YN)

As was the case at PAW1, the scatter in the observed relation between QN and Q is large, while the model relation is much cleaner and with a very different slope (Figure

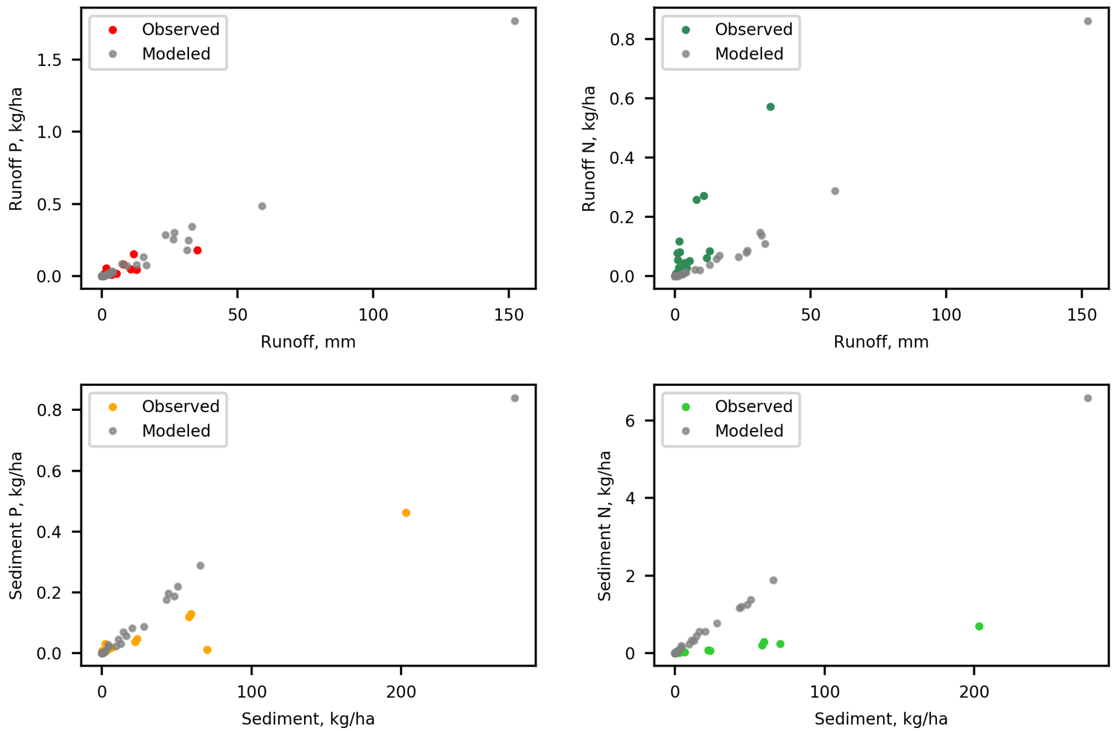

Figure 2.28: Model and observed relationships between nutrient losses and runoff (top) and sediment (bottom), on the WIL2 watershed.

2.28). Despite this, the model does a remarkably good job of estimating the total QN loss during the project, coming within a few per cent of the actual value (Figure 2.31). This is in stark contrast to the gross underestimate of QN in the PAW1 model.

For sediment-bound N, the observed losses at WIL2 quite closely mimic the observed pattern of sediment loss. The WIL2 baseline model overestimates the magnitude of almost all of these events, resulting in a total model YN loss of 15.25 kg ha^{-1}, compared to the observed value of 1.82 kg ha^{-1}. The PAW1 baseline model also overestimated YN, but by a much smaller amount.

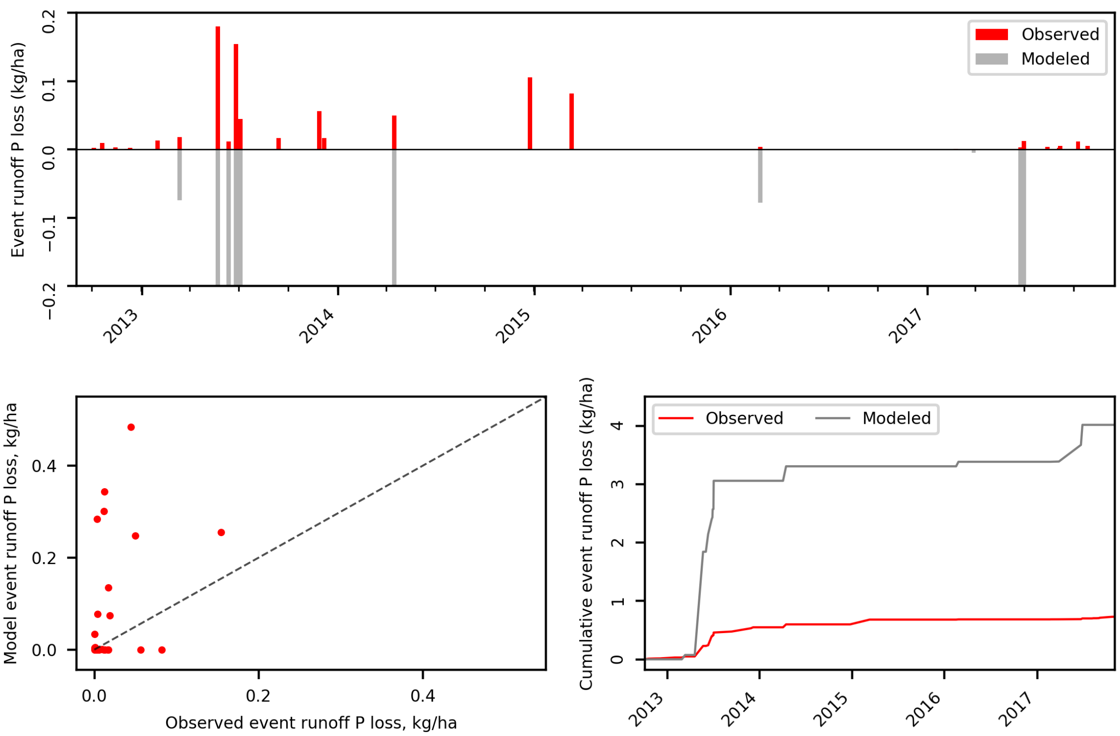

Figure 2.29: As for figure 2.27 but for P dissolved in runoff (WIL2 watershed).

Yield

The model silage yield at WIL2 compares more favorably with the state average than the PAW1 model did. It is still slightly underestimated, but this time the mean annual yield falls short by just under 20%. The model is close to the farmer-reported yields in 2013 – 2015, but it grossly overestimates the low yield in 2012. As for PAW1, the farmer-reported yields are more variable than the model yields.

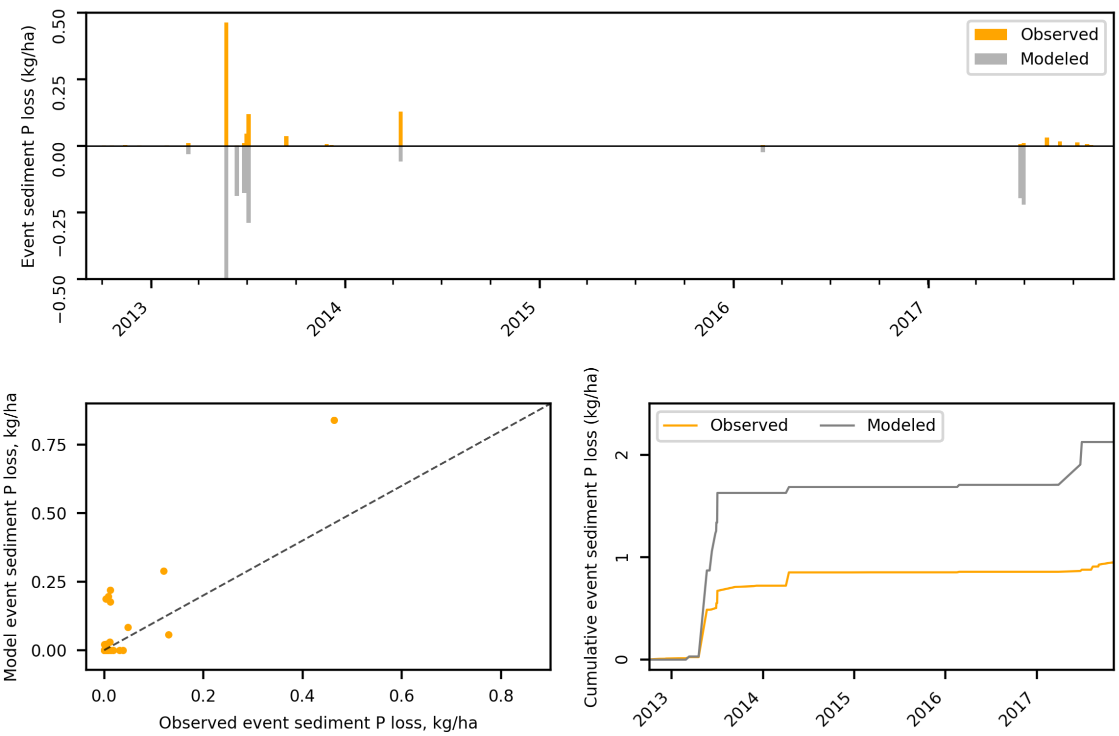

Figure 2.30: As for figure 2.27 but for sediment-bound P (WIL2 watershed).

2.5 Discussion

This Chapter has examined in some detail the process of setting up APEX models for two small watersheds using data collected for a project that aimed to monitor the effectiveness of field-based best management practices on Vermont dairy farms. The APME project provides a wealth of farm operations records and weather and site information that are not commonly available for working farms, along with an extensive set of measurements of runoff, erosion, and N and P loss at the field edge. This data set has enabled the creation of APEX models that quite closely mimic the conditions and operations experienced on the study watersheds, and adds an extra

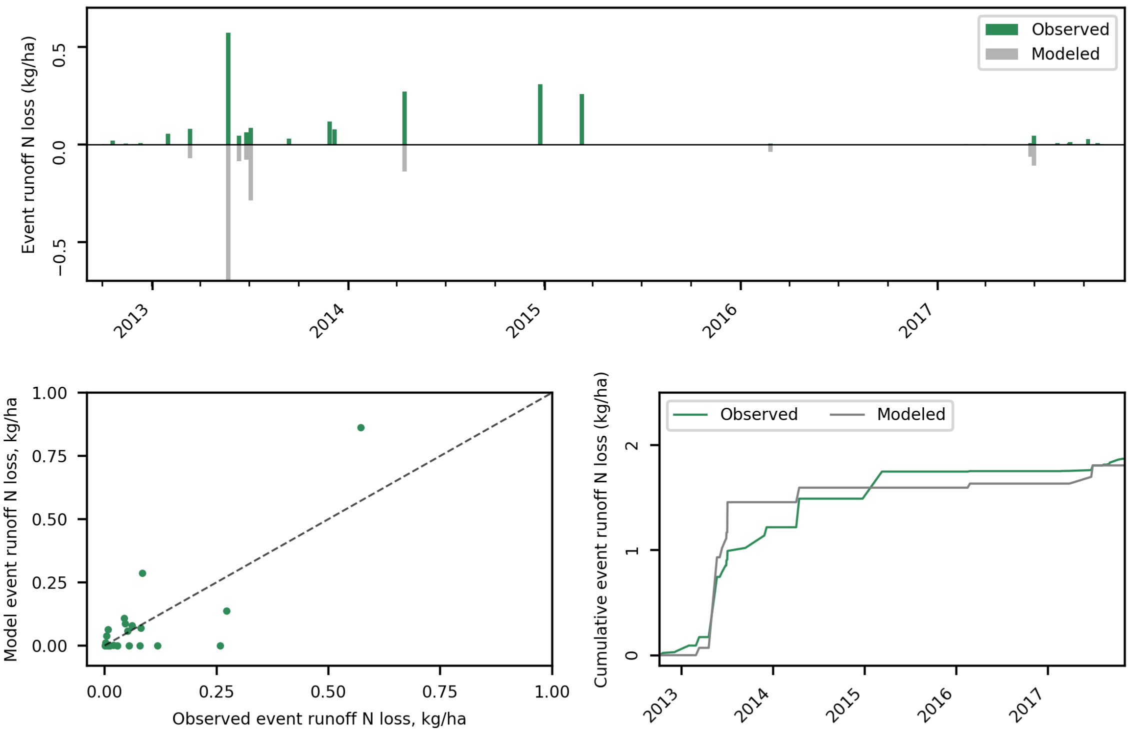

Figure 2.31: As for figure 2.27 but for N dissolved in runoff (WIL2 watershed).

layer of value to the APME project itself.

At the same time, the APME data were not collected for the specific purpose of modeling. This, combined with challenging site conditions and the errors and uncertainties that are inevitably present in field measurements, posed some challenges for the modeling effort. These include:

- Identifying and replacing unreliable meteorological data (such as solid precipitation measurements)
- Identifying APEX input and output variables corresponding to agronomic, soil test, and edge-of-field data, or making reasonable conversions

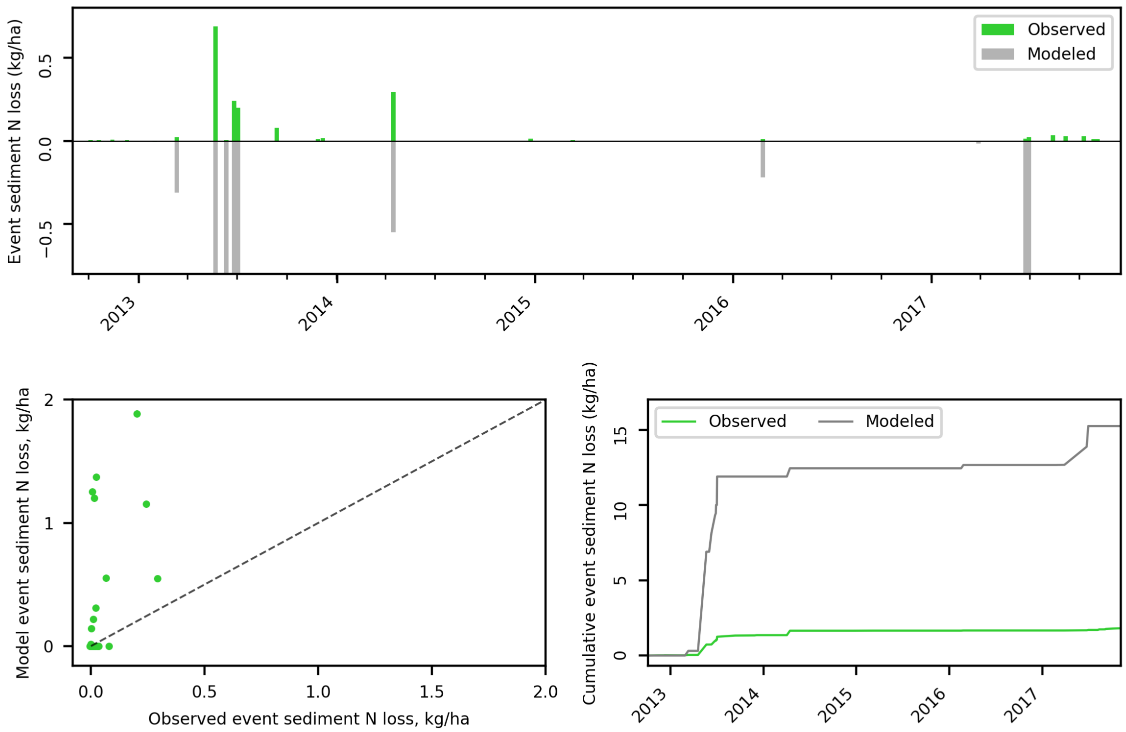

Figure 2.32: As for figure 2.27 but for sediment-bound N (WIL2 watershed).

- Reconciling different timescales: daily model output vs arbitrary runoff event durations
- Evaluating model performance based on field data with poorly-understood errors

Others may wish to bear these things in mind when planning to use similar studies for modeling work, or while designing field experiments that will be used to parameterize and calibrate models.

The two watersheds that were modeled are similar in a number of ways, including their soil properties and the fact that they grow silage corn fertilized with dairy manure. However, they also differ in some respects, perhaps most importantly in the

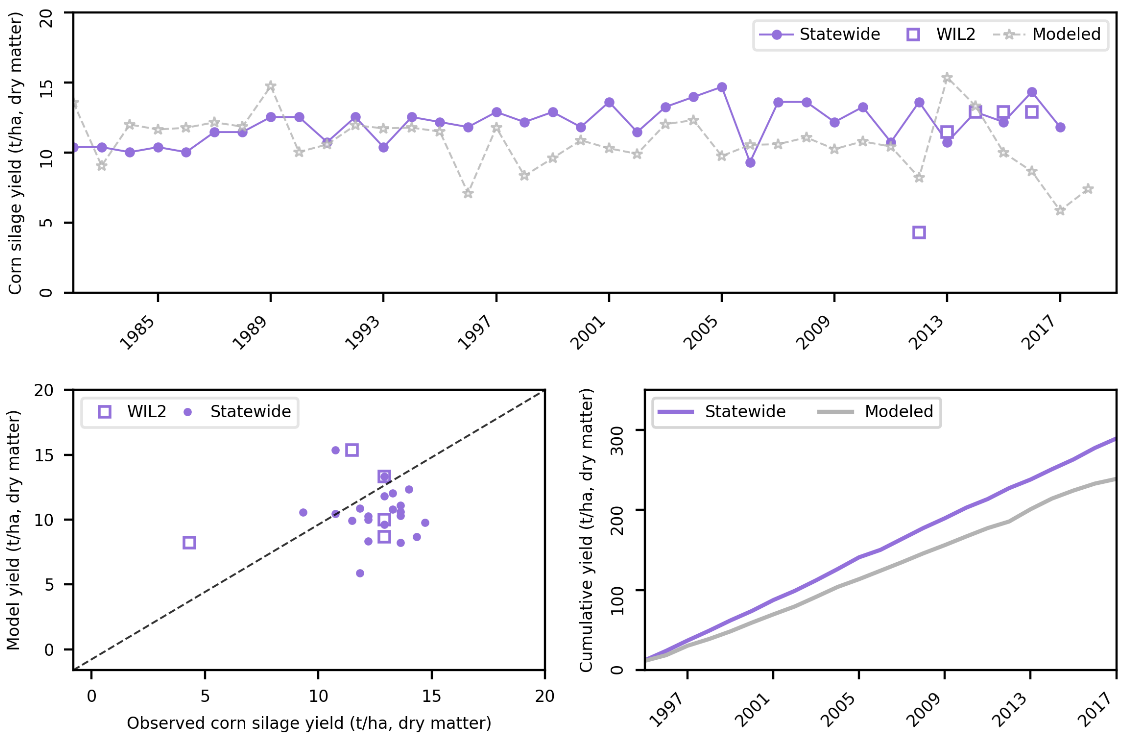

Figure 2.33: Upper panel: time series plot of model corn silage yield, statewide historical average yield, and the yields reported by the farmer for the WIL2 watershed during the APME project. Lower left: Model yield vs statewide historical and farmer-reported yield. Lower right: Cumulative model and statewide historical yields.

dates, rates, and methods of applying manure and starter fertilizer, and the fact that one site is considerably flatter than the other. In terms of the runoff, sediment loss, and N and P losses measured in the field, the sites also showed some similarities and differences.

The largest and most frequent runoff events at both sites occurred in 2013, which experienced several months of above-average rainfall. Overall, though, runoff events were much more numerous at the PAW1 site. The pattern of nutrient loss in runoff was only weakly related to that of runoff itself, except for dissolved P at Pawlet. In the case of dissolved N, this could perhaps be due to the complex set of transformations undergone by N in the soil.

Sediment losses at both sites did not bear a strong resemblance to runoff. Fairly large runoff events often generated little sediment, whereas large erosion events could occur when runoff was moderate. However, the pattern of sediment-bound N and P losses quite closely resembled that of the erosion events themselves.

The performance of the baseline models was similarly mixed. Figure 2.34 summarizes the observed runoff, sediment, etc. summed over the duration of the monitoring project. At both sites, the models overestimate erosion and underestimate crop yield. However, they predict too little runoff at PAW1 and too much at WIL2.

Both models overestimate P lost in runoff and sediment, but at PAW1 the error is larger for sediment-bound P, while runoff P has the worse prediction at WIL2. Sediment-bound N is overestimated at both sites, but by a much larger factor at WIL2. On the other hand, runoff N is grossly underestimated at PAW1 but correctly predicted at WIL2.

Within a single site, the performance of the models varies over time. At PAW1, the predictions are much better for the relatively wet 2012 – 2013 period than for 2014 – 2015. Performance at WIL2 is more mixed, but APEX perhaps tends to do better in the first half of each year than the second. Overall the baseline models do not perform well, and they are not consistent in the magnitude, direction, and timing of their errors.

§2.3.1 and §2.3.2 identified some aspects of the models that may present opportunities to improve model performance through calibration. In particular, the crops absorbed fewer heat units than expected, were subject to high levels of N stress, and

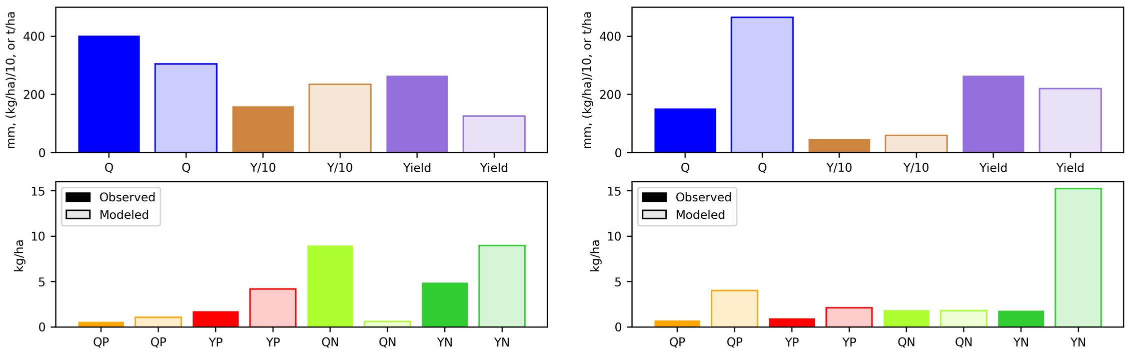

Figure 2.34: Total observed and modeled runoff, erosion, etc. for the duration of the APME project (Q=runoff, Y/10=sediment divided by 10, QP/N= runoff P/N, YP/N=sediment P/N). Left: PAW1. Right: WIL2.

the model predicted much higher levels of NH_3 volatilization than expected for these watersheds. Adjusting parameters related to heat unit uptake and volatilization may also improve crop yields. In addition numerous other options – e.g. runoff and erosion calculation methods – and parameters can be adjusted to improve the model outputs. Chapter 3 takes a closer look at that process and the results that can be achieved.

3 APEX Model Calibration

The "baseline" APEX models presented in Chapter 2 were based on real climate, management and agronomic records for the sites that were being modeled. However, most other model parameters were left at their default values. Those defaults may not be appropriate for the specific circumstances of dairy farms in Vermont, but it is not clear a priori what the optimal values might be. It should therefore be possible to improve the baseline models through a process of calibration.

Calibration refers to the action of refining the parameters of a model to improve its ability to reproduce a set of observations. In the simplest version, the value of a parameter is adjusted, the model re-run, and the new output compared against data such as the APME measurements described in Chapter 2. This is repeated for different values of the same parameter, and for other parameters, until the model output is optimized or until it is judged that a satisfactory outcome has been reached.

Sensitivity analysis is often carried out before calibration takes place. The number of APEX parameters is large, and it is not practical to search for the optimal value of every one of them even with automatic calibration tools. Various methods, including dedicated APEX support software (§3.1), are available to quantify whether a given parameter has a significant effect on the output variables that are of interest (e.g.

Wang et al., 2006). These methods allow the user to identify the important variables and select only those ones to take part in the calibration process.

This chapter describes the sensitivity analysis and calibration methods used to improve the PAW1 and WIL2 baseline models. Calibration requires a means of quantifying how the changes are affecting the model at each iteration, and how well the final version performs. Model performance measures are therefore also introduced, and used to evaluate the baseline, intermediate, and final, calibrated models.

3.1 Overview of the Calibration Process

Calibration can be done manually, automatically, or using a combination of the two approaches (e.g. using manual calibration to find an approximate solution followed by an automatic search to refine it). Manual calibration is a painstaking process in which parameters are adjusted by hand, one at a time, and the results used to decide upon the next step in the process. The time required for manual calibration means that many combinations of parameters will not be explored. Interactions between parameters can have significant effects on the model output (Wang et al., 2014), so a limited manual calibration could lead to useful models being overlooked.

Automatic calibration programs quickly search through many combinations of parameters, seeking to maximize or minimize a specified objective function. Automatic methods are capable of producing models with better performance statistics than manual calibration (Wang et al., 2014), and an extensive automatic search may identify useful solutions where coarser manual methods are unable to find acceptable models. However, Wang et al. (2014) suggest that this may be at the expense of misrepresenting parameters and processes that are not constrained by data.

The APEX-CUTE[1] program can perform automated calibration of APEX models (Wang et al., 2014; Wang and Jeong, 2016). However, to calculate model performance statistics, the program needs to compare daily, monthly, or annual APEX output with field measurements on the same timescales. Aside from crop yield, the APME data do not conform to any of those timescales; the runoff etc. measurements reflect events that may have a duration of <1 – several days (§2.4.1). This author's experiments with using APEX-CUTE to calibrate the PAW1 model were not successful, and the effort required to update the software to handle variable-timescale field data was judged to be beyond the scope of this thesis. A manual calibration approach was therefore adopted for this work.

During manual calibration, model processes are calibrated sequentially. Wang et al. (2012) recommend that runoff be calibrated first, together with or followed by crop yield, then sediment. Nutrient losses are calibrated only when satisfactory results have been achieved for the underlying runoff and erosion processes that drive them. This sequence of runoff – yield – sediment – nutrients is used for the APEX models in this work.

Sensitivity analysis for APEX can also be carried out with APEX-CUTE (Wang et al., 2014; Wang and Jeong, 2016), which uses the enhanced Morris method (Campolongo et al., 2007; Morris, 1991). The user specifies which parameters are to be included in the analysis, the range of values to be sampled for each one, and the number of points to be sampled within that range. APEX-CUTE runs APEX using

[1]https://epicapex.tamu.edu/apex-cute-4-3-download/

a starting set of values for each parameter, changes the value of one parameter, runs APEX again, and calculates the difference between the model outputs of interest from one run to the next. Then, APEX-CUTE selects another parameter, changes its value, runs the model again, and so on. The sensitivity measure for each parameter is the mean of the differences in model outputs for each run in which that variable is changed (Wang et al., 2006).

APEX-CUTE allows the user to select any of 100 PARMs (§2.2.6), many crop growth parameters (§2.2.4), and a handful of site, subarea, and soil parameters to take part in the sensitivity analysis. However, there are many parameters that are not included. They encompass, for example, parameters that specify the methods to be used by the model, such as the equation to be used for erosion calculations, or the method of deriving the curve number used to estimate runoff each day. The calibration was therefore carried out using a combination of user judgment to select potentially relevant parameters, and sensitivity analysis with APEX-CUTE on a subset of them.

In most hydrological modeling studies, the calibrated model is run on a set of "validation" or "test" data that were excluded from the calibration or "training" data set. This verifies that the model can produce an acceptable description of the data in general, and that it has not been so finely tuned that it reproduces the calibration data at the expense of its ability to generalize to other data sets. There are various ways of separating data into training and test sets (e.g. Daggupati et al., 2015; James et al., 2000), but many studies simply split the data set into a calibration period and a validation period.

In this thesis, no validation data set is used. This is partly because the number of data points is quite limited, particularly for water quality variables and especially for the WIL2 watershed. Also, for PAW1, no model was ever found that simultaneously fit both the high-and low-runoff years very well, despite much experimentation. Modeling these periods together gives a chance of finding a model that performs adequately at both times.

3.2 Model Performance Indicators

To assess whether parameter changes are improving the model, and to judge the quality of the final version, it is necessary to quantitatively evaluate the model's performance. Moriasi et al. (2007, 2015, hereafter M07, M15) provide guidance in this area. Those papers reviewed the hydrological modeling literature to identify and evaluate the ways in which researchers have presented and assessed their modeling results. They describe various graphical performance indicators (used in Ch. 2), identify the statistical performance indicators used in the literature, and discuss their pros and cons. Finally, the authors recommend a set of six statistical performance measures (PMs) that should be presented for any hydrological modeling study. This thesis largely follows their advice.

The recommendations of M07 and M15 are based on (a) the effectiveness of the PMs in indicating model performance, and (b) whether they are frequently used in the literature, thereby enabling useful inter-model comparisons. The PMs are discussed in detail in M07 and M15 and references therein. Definitions of, and brief comments on these statistics are presented below (O and P denote observed and predicted values, respectively).

1. Coefficient of determination (R^2)

$$R^2 = \left[\frac{\sum_{i=1}^{n} \left(O_i - \bar{O}\right)\left(P_i - \bar{P}\right)}{\sqrt{\sum_{i=1}^{n} \left(O_i - \bar{O}\right)^2} \sqrt{\sum_{i=1}^{n} \left(P_i - \bar{P}\right)^2}} \right]^2 \tag{3.1}$$

R^2 ranges from 0 to 1, with an optimal value of 1. R^2 is sensitive to outliers but additive and proportional differences between model and observed data do not reduce its value. The intercept and slope of the regression line should ideally also be reported and should be near 0 and 1 respectively for a good model fit.

2. Nash-Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(O_i - P_i\right)^2}{\sum_{i=1}^{n} \left(O_i - \bar{O}\right)^2} \tag{3.2}$$

NSE ranges from -∞ to 1, with an optimal value of 1. NSE <0 implies that the mean of the observed values is a better predictor than the model output for that point. NSE is sensitive to outliers and does not indicate whether a model is biased.

3. Index of agreement (d)

$$d = 1 - \frac{\sum_{i=1}^{n} \left(O_i - P_i\right)^2}{\sum_{i=1}^{n} \left(|P_i - \bar{O}| + |O_i - \bar{O}|\right)^2} \tag{3.3}$$

d ranges from 0 to 1, with an optimal value of 1. Like R^2, d is sensitive to extreme values. Unlike R^2, though, d is also sensitive to additive and proportional differences between observations and model. Various authors have reported finding high d values for poor model fits (Moriasi et al., 2015).

4. Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(O_i - P_i\right)^2} \tag{3.4}$$

RMSE ranges from 0 to ∞, with an optimal value of 0. Unlike the dimensionless statistics discussed so far, RMSE is in the same units as the observed/modeled quantity being evaluated, which can be helpful for interpretation.

5. Ratio of RMSE to the standard deviation of the observations (RSR)

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum_{i=1}^{n}\left(O_i - P_i\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(O_i - \bar{P}\right)^2}} \tag{3.5}$$

RSR ranges from 0 to ∞, with an optimal value of 0.

6. Percent Bias (PBIAS)

$$PBIAS = \frac{\sum_{i=1}^{n}\left(O_i - P_i\right)}{\sum_{i=1}^{n}\left(O_i\right)} \times 100 \tag{3.6}$$

PBIAS ranges from $-\infty$ to ∞, with an optimal value of 0. PBIAS measures the average tendency of a model to over- or under-estimate the observed values. PBIAS can be close to zero for a model that both under- and over-predicts, and it says little about how well the model reproduces trends in the observations.

In addition to the PMs recommended by Moriasi et al. (2007, 2015), this thesis also reports the *p* value derived from the two-sample Kolmogorov-Smirnoff test. A p-value near zero indicates that the model and data are drawn from different distributions; that is, that the model is a poor representation of the data.

Each of these PMs is capable of identifying some aspects of model performance but is blind to others, and many of them can be disproportionately affected by extreme values. It is therefore usually advisable to use more than one PM to evaluate the performance of a model. In the following sections, NSE, and PBIAS are used to illustrate the progress of the calibration. PBIAS relates to how well the model predicts the mean and sum of the observed values, while NSE gives some indication of how well the model fits the pattern of the observed points. In the interest of enabling future comparisons between these models and others in the literature, all of the M07 and M15 recommended statistics are presented for the initial baseline and final calibrated models[2].

Moriasi et al. (2007, 2015) also present numerical criteria by which a hydrological model can be judged to be "very good", "good", "satisfactory", or "unacceptable". For example, M07 suggest that in general, a model can be considered "satisfactory" if it has $0.60 < \text{RSR} \leq 0.7$, $0.50 < \text{NSE} \leq 0.60$, $\pm 30\% < \text{PBIAS} \leq \pm 55\%$ for sediment and $\pm 40\% < \text{PBIAS} \leq \pm 70\%$ for nutrients. An important point that is mentioned only in passing is that their criteria define how a model performs *relative to other models*. The standards are based on how the model's PMs compare to their compilations of published PMs for that model, and a model that is judged to be "very good" or

[2]Calculated using the "spotpy" package (Houska et al., 2015).

"unacceptable" by those standards may or may not be useful for the intended purpose of the modeling[3]. The M07 and M15 criteria, along with some guidelines given by Wang et al. (2012), will be used to give context to the performance of the APEX models in this thesis, but will not be used to establish the utility of the models.

3.3 Model Calibration: PAW1

This section records the steps taken to calibrate the PAW1 model for runoff, silage yield, erosion, and nutrient losses. The progress of the calibration is shown in a set of plots that record performance statistics at each step (e.g. Figure 3.2).

3.3.1 Runoff

The steps involved in the runoff calibration of the PAW1 model are illustrated in Figure 3.1. The next sections follow the sequence depicted in the figure, explaining the reasoning behind the choices of processes and parameters tested, and the results obtained at each step. Figure 3.2 shows how the model NSE and PBIAS change as the calibration progresses.

Runoff calculation method

Two runoff calculation methods are available in APEX: the SCS curve number (CN) method (USDA-SCS, 1972) used for the baseline models (INFL=0 in the Control file),

[3]M15 also acknowledge several other limitations: that the number of published performance statistics for APEX is small, that model performance varies when compared with data on different time scales, and that a bias against publishing poor model fits may mean that the published statistics do not give an accurate picture of overall model performance.

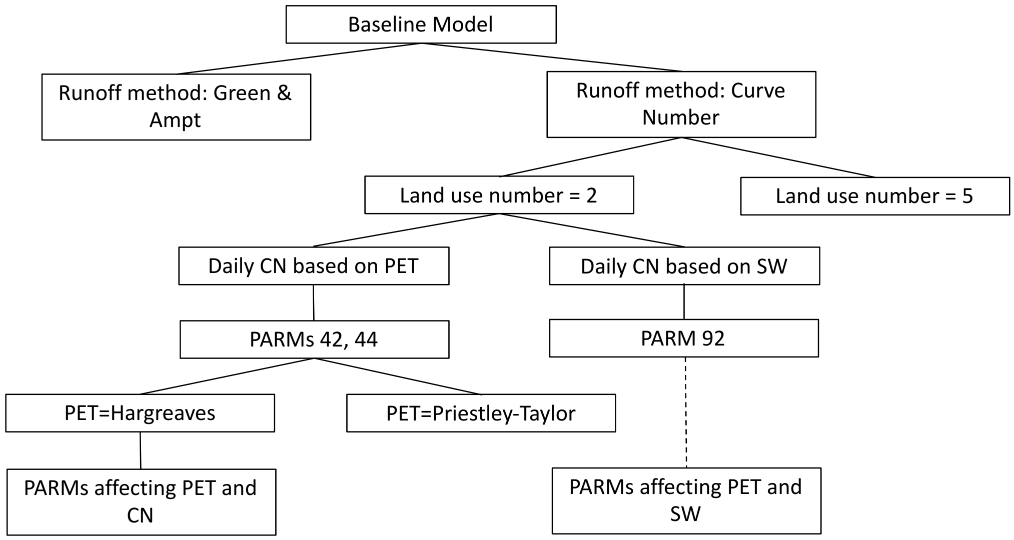

Figure 3.1: Sequence of processes/parameters tested during the calibration of the PAW1 model for runoff (see text for details). CN = curve number, PET = potential evapotranspiration, SW = soil water.

and the Green & Ampt (G&A) method (INFL=1,2,3,4; Green and Ampt, 1911). With G&A, several options are available depending on the form assumed for the rainfall distribution within a rain event (exponential or uniform), and whether intra-storm rainfall and peak rainfall data are available. Although 15-minute rainfall quantities are available for the APME project watersheds, using APEX to model future climate scenarios will require rainfall data to be simulated. Therefore only the INFL=1: "Rainfall Exponential Distribution, Peak Rainfall Rate Simulated" option was tested. Setting INFL=1 caused the runoff to drop almost to zero, leading to PBIAS=-99% and NSE=-0.60. This is clearly a much poorer result than for the default INFL=0 (Figure 3.2), so the curve number method was retained for the remainder of the calibration.

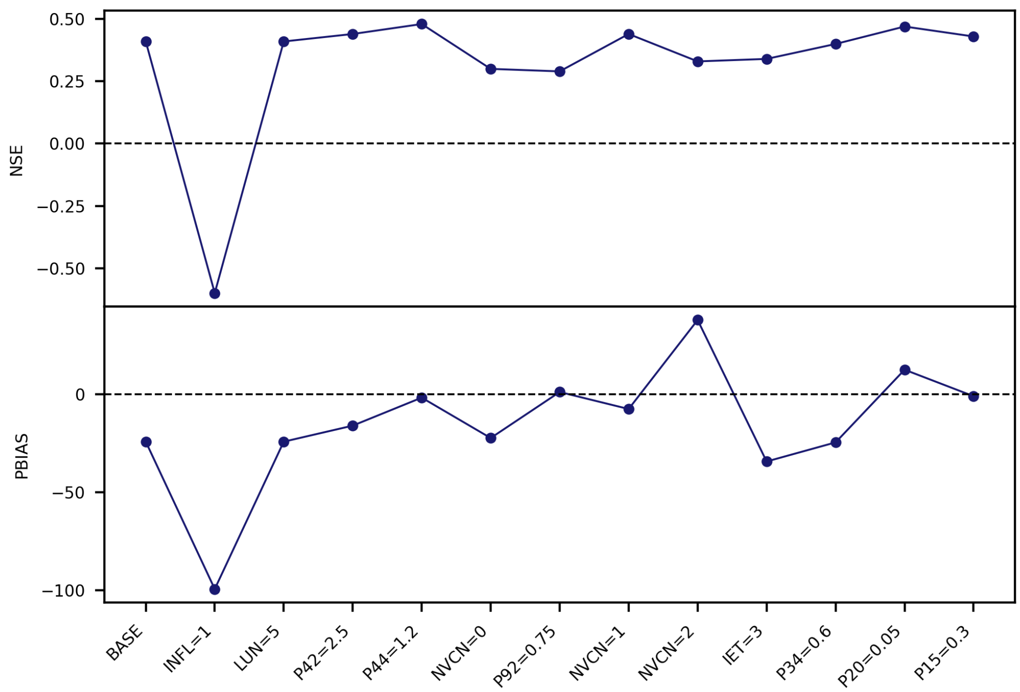

Figure 3.2: Changes in performance statistics as the PAW1 baseline model is calibrated for runoff. X-axis labels relate to parameter changes; see text for full details.

Land use number

With the curve number method, runoff is predicted according to:

$$Q = \frac{(RFV - 0.2s)^2}{RFV + 0.8s};\ RFV > 0.2s \tag{3.7}$$

$$Q = 0;\ RFV \leq 0.2s \tag{3.8}$$

where RFV is the daily rainfall and s is the potential retention after runoff begins. The retention parameter (in mm) is derived from the curve number:

$$s = 254 \times \frac{100}{CN - 1} \tag{3.9}$$

The curve number itself is found from standard lookup tables that relate the CN in average moisture conditions to land use (e.g. fallow, row crops, small grains), land cover (straight row, contoured, or terraced), soil condition (good or poor infiltration) and soil hydrologic group. The APEX parameter that encapsulates this information and determines which CN is used is the Land Use Number, LUN, in the OPS file.

In the baseline model, LUN is set to LUN=2 ("Straight Row, Poor Infiltration"). The model underestimates runoff in 2012–2013 (a relatively wet period), overestimates in 2014 – 2015 (a relatively dry period), and results in too little total runoff. LUN=2 is the value most conducive to runoff in a row cropping situation, so changing LUN would most likely exacerbate the overall deficit of runoff in the model. LUN=5 ("Contoured, Good Infiltration") was tested during the PAW1 calibration, but perhaps surprisingly, it made no difference to the runoff statistics. LUN=2 was used for the remaining calibration steps.

Daily CN calculation

Once APEX has established the average curve number according to the equation above, the code recalculates a new CN every day depending on how wet or dry the soil is that day. There are four relevant[4] ways of calculating the daily curve number, three of which relate CN directly to the soil water content and one that uses an indirect "soil moisture index" instead. The method to be used is specified by the

NVCN variable:

- NVCN=0: "Variable daily CN nonlinear CN/SW with depth soil water weighting". The daily curve number is related to soil water content, field capacity, and wilting point in a nonlinear way, and is weighted by depth.
- NVCN=1: "Variable daily CN nonlinear CN/SW without depth weighting".
- NVCN=2: "Variable daily CN linear CN/SW no depth weighting".
- NVCN=4: "Variable daily CN SMI (soil moisture index)". This method relates the daily curve number to rainfall and evapotranspiration instead of directly to soil water content, field capacity, and wilting point.

The default value, NVCN=4 was used for the baseline model. The equation describing this option has two parameters, PARM42 and PARM44, that can be used to tune the results:

$$s = s_o + PET \times e^{-PARM_{42} \times s_o / s_1} - RFV + Q \tag{3.10}$$

$$s < PARM_{44} \times s_1 \tag{3.11}$$

where s_1 is the retention parameter associated with dry conditions. Wang et al. (2012) suggest using PARM42 as a calibration parameter, and Williams et al. (2012) note that PARM42 and PARM44 are "convenient for calibration". Increasing PARM42 from 1.0 to 2.5 (maximum) generally increases runoff in the PAW1 model. This negatively affects the match between observed and model runoff in 2015, where the model

[4]NVCN=3, "Non-varying CN – CN2 used for all storms", is intended for feedlot-type situations and is not relevant here.

already predicted too much runoff, but overall the performance statistics are slightly improved (Figure 3.2).

With PARM42=2.5, decreasing PARM44 from 1.5 to 1.2 gets the total runoff almost exactly right and also has the highest NSE so far. Broadly speaking, the overall effect is to increase runoff in all events. Again, this improves the model in 2012 – 2013 at the expense of 2014 – 2015.

The remaining NVCN options relate the daily curve number directly to soil water content, so they could potentially result in a different pattern of runoff event magnitudes that better matches the field data. The statistics obtained using NVCN=0, 1, and 2 are shown in Figure 3.2. PARM92 adjusts the results of the NVCN=0 calculation, analogous to PARM42 for NVCN=0 (Steglich et al., 2016), so results from changing PARM92 are also shown.

With NVCN=0 and PARM92=0.75, APEX is able to closely match the total measured runoff. The pattern of runoff event magnitudes is indeed somewhat different from the NVCN=4 models: the new model comes closer to reproducing events in late 2013 for which the NVCN=4 models predicted little to no runoff (Figure 3.3). This is balanced by difficulties matching some other points, however, and the NSE of this model is slightly lower than that of the baseline model. The NSE of the NVCN=1 model is higher, but so far the best combination of all the statistics in Figure 3.2 comes from the model with NVCN=4, PARM42=2.5, and PARM44=1.2.

The NSE and PBIAS for the best model so far are 0.48, and -1.6% respectively. Wang et al. (2012) suggests aiming for NSE $\geq$0.55, and PBIAS within 20% when using APEX to model runoff. The PAW1 model meets the suggested criterion for PBIAS, but the NSE is slightly lower. The underlying issue is that the model tends

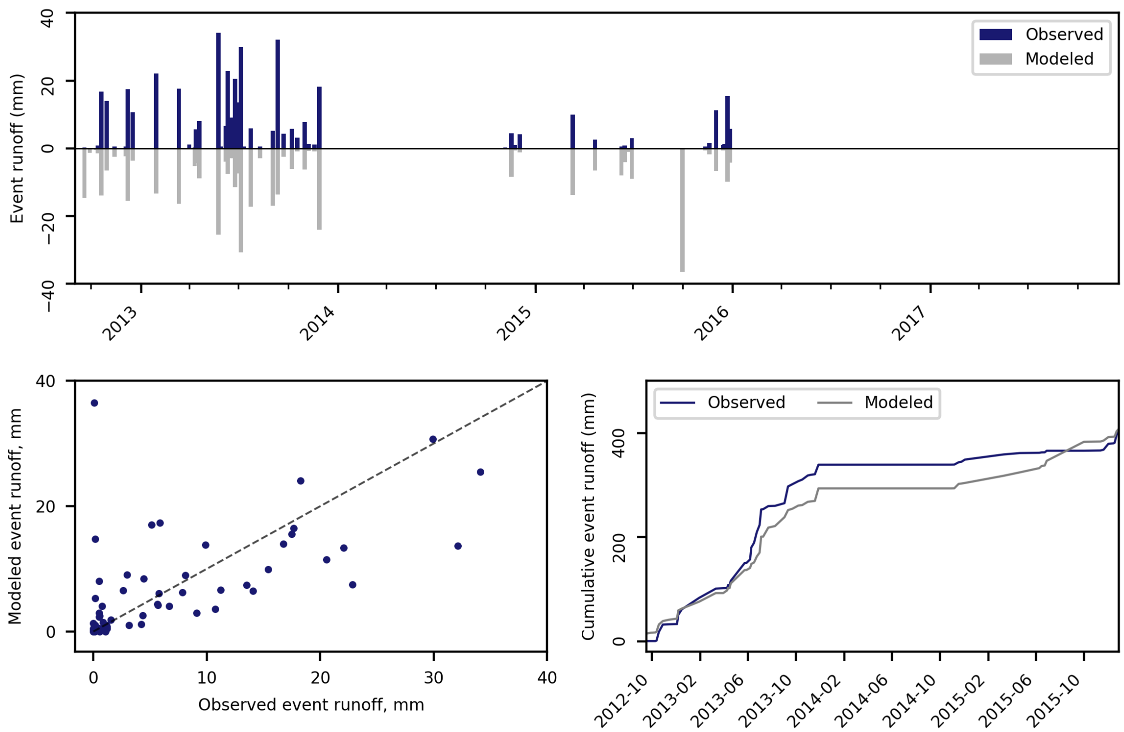

Figure 3.3: The PAW1 model with NVCN=0 and PARM92=0.75. This model is better at reproducing some of the points in late 2013, at the expense of other points (especially in 2015). This is about the largest change in the pattern of model runoff found during the calibration process. Most parameter changes simply increased or decreased the overall level of runoff.

to predict more runoff, relative to the field measurements, in 2014 – 2015 than in 2012 – 2013. Improving the model's performance would require identifying parameters that cause substantial differences in the pattern of the model runoff rather than simply increasing or decreasing runoff overall. Such parameters may or may not exist in APEX.

A search for parameters having the desired effects could be done in two ways. The NVCN=0 – 2 models are formally slightly inferior to the model with NVCN=4, PARM42=2.5, and PARM44=1.2. However, with NVCN=0 – 2, APEX relates the daily curve number directly to soil water. Many parameters in APEX are likely to directly or indirectly affect soil water content, so further calibration of one of more of the NVCN=0 – 2 models may offer more flexibility. On the other hand, the parameter space that could be explored is very large, and higher flexibility also implies a higher likelihood of overfitting the model. Given the limited resources available for this project, and the possibility that no parameter combination is capable of fitting all the PAW1 runoff measurements simultaneously, the following sections present a cursory examination of the effect of some parameters that affect PET and the curve number itself, on the NVCN=4, PARM42=2.5, and PARM44=1.2 model.

Potential evapotranspiration method

With NVCN=4, runoff is a function of potential evapotranspiration (PET, Eq. 3.10). APEX contains several options for estimating PET. The Hargreaves method (IET=4) was chosen for the baseline models because it does not require wind speed and humidity data and is recommended by APEX support staff. However, the Priestly-Taylor method (IET=3) can also be used when limited weather data are available. Using this option turns out to preserve the overall pattern of the model runoff while decreasing its magnitude (PBIAS=-35%). It may be possible to change the overall magnitude using PARM42 and PARM44, but there would be no obvious benefit to doing that.

PARMs

An APEX-CUTE sensitivity analysis was carried out to identify the PARMs likely to have the greatest effect on runoff calculated using NVCN=4. Table 3.1 lists the PARMs that were judged to be potentially relevant and therefore included in the

Table 3.1: APEX parameters used for runoff calibration/sensitivity analysis

Parameter	Definition[a]
PARM 15	Runoff CN residue adjustment factor
PARM 16	Expands CN retention parameter
PARM 20	Runoff curve number initial abstraction
PARM 22	Reduces runoff CN retention parameter for frozen soil
PARM 23[b]	Hargreaves PET equation coefficient
PARM 25	Exponential coefficient used to account for rainfall intensity on CN
PARM 34	Hargreaves PET equation exponent
PARM 42	CN index coefficient
PARM 44	Upper limit of CN retention parameter
PARM 49	Maximum rainfall interception by plant canopy
PARM 50	Rainfall interception coefficient

[a] APEX User's Guide Section 2.21
[b] Repeatedly caused APEX-CUTE to crash, therefore not included in the final analysis.

analysis. Although PARM42 and PARM44 were explored above, they are included so that their influence can be compared with that obtained for other PARMs.

Figure 3.4 shows the results of the analysis. The most sensitive parameters are PARM34 (the exponent in the Hargreaves PET equation), PARM20 (the initial abstraction parameter, i.e., the factor of 0.2 in Eq. 3.7), and PARM15 (which adjusts the curve number according to how much surface crop residue is present). Perhaps unexpectedly, the Morris index plot suggests that PARM42 and PARM44 should have relatively little effect on runoff.

Changing PARM34 from 0.5 (min) to 0.6 (max) decreases total runoff without having a major effect on the distribution of runoff volume across events. Changing PARM20 from 0.2 to 0.05 (min) has a similar effect in the opposite direction (increasing runoff). Increasing PARM15 from 0.0 (min) to 0.3 (max) does not change the total runoff and does redistribute the runoff somewhat differently between events. However, this is not on a scale that makes any significant difference to the model

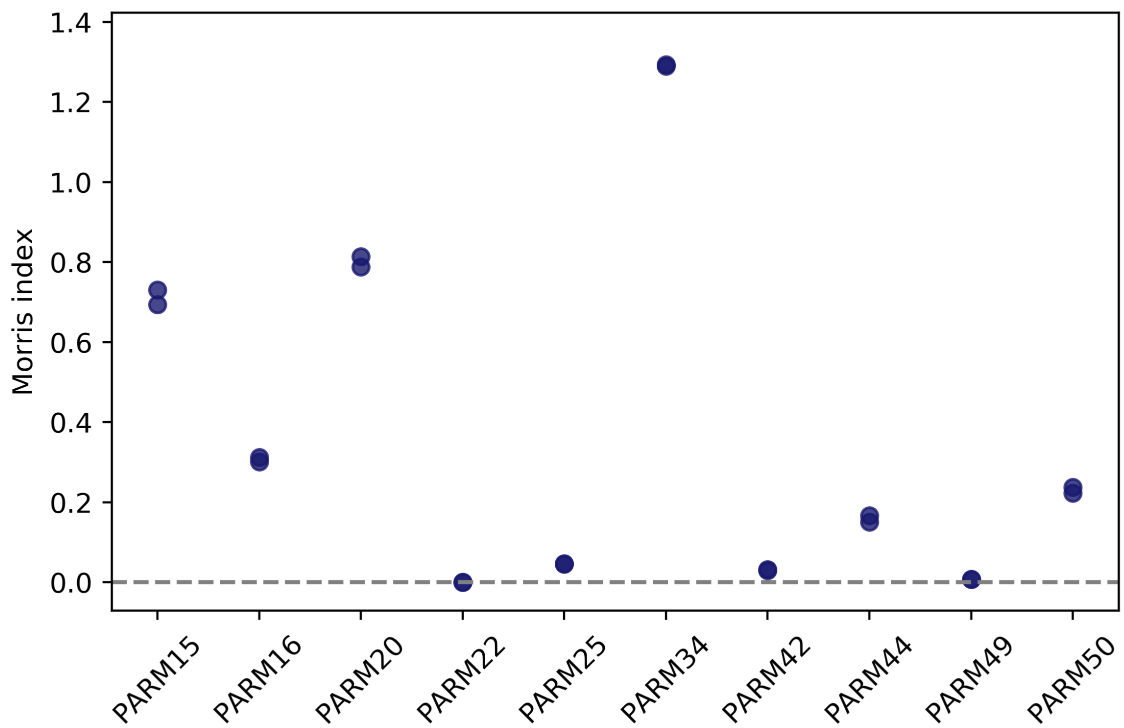

Figure 3.4: Results of APEX-CUTE sensitivity analysis for the PARMs listed in Table 3.1. The Morris method involves randomly sampling within a range of possible values for each parameter, so two runs were performed to verify that consistent Morris indices were obtained.

performance statistics.

Runoff calibration summary

Changing two parameters, PARM42 and PARM44, was sufficient to achieve a close match between the model and observed total runoff at PAW1 (and a correspondingly low PBIAS). Even the best model fell short of achieving an NSE that is considered good in the literature, largely because of its inability to simultaneously give good results for both the 2012 – 2013 and 2014 – 2015 monitoring periods. As well as

PARM42 and PARM44, several other parameters could probably have been used to achieve similar performance statistics. The main effect of them was simply to increase or decrease the volume of runoff over all events; the pattern of event sizes never changed enough to correct the general overestimate of runoff in 2014 – 2015 while retaining good results for 2012 – 2013.

At this point, the model with NVCN=0, PARM42=2.5 and PARM44=1.2 will be calibrated for crop yield. None of the parameters tested during the runoff calibration will be varied during the yield calibration stage, and the runoff statistics will be monitored as the crop calibration progresses.

3.3.2 Crop Yield

The mean forage yield in the baseline model was 6.0 t ha^{-1} dry matter compared to the statewide historical mean value of 12.6 t ha^{-1}. At the end of the runoff calibration, the mean yield is essentially the same (6.1 t ha^{-1}). In §2.3.1 it was suggested that the low yields in the baseline model are likely to be due to N stress caused by high volatilization. Insufficient heat unit uptake may also play a role. Both of these factors are still present in the runoff-calibrated model.

The first step in calibrating the crop yield was to run a sensitivity analysis on parameters that could affect N availability in general and volatilization in particular. The parameters included in the analysis are given in Table 3.2. Most of them bear directly on the various forms taken by N in the soil, but the list also contains some parameters to do with organic matter turnover and mineralization. Because the APEX manual cautions against changing crop parameters (such as maximum leaf area index, plant N concentration, radiation use efficiency, etc.) without good reason (§2.2.4),

Table 3.2: APEX parameters used for yield sensitivity analysis

Parameter	Definition[a]
PARM4	Water storage N leaching
PARM14	Nitrate leaching ratio
PARM27	CEC effect on nitrification and volatilization
PARM29	Biological mixing efficiency
PARM31	Maximum depth for biological mixing
PARM35	Denitrification soil water threshold
PARM36	Upper limit of daily denitrification rate
PARM69	Adjusts microbial activity in top soil layer
PARM72	Volatilization/nitrification partitioning coefficient
PARM80	Upper limit of nitrification – volatilization
PARM86	N and salt upward movement by evaporation
PARM100[b]	Century slow humus transformation coefficient

[a] APEX User's Guide Section 2.21
[b] Repeatedly caused APEX-CUTE to crash, therefore not included in the final analysis. Also, parameters after PARM100 which are relevant to organic matter decomposition are not included in APEX-CUTE.

parameters from the CROP file were not included in the analysis or calibration.

Figure 3.5 shows the results of the sensitivity analysis. PARM72 is clearly the most influential parameter. This parameter governs how N is partitioned between nitrification and volatilization, so it makes sense that it would have a strong effect on yield in a simulation that has such high volatilization losses. PARMs 4, 14, and 80, which are also to do with N leaching and volatilization, are of secondary importance.

Decreasing PARM72 from 0.4 to 0.05 (min) has a clear effect on the yield, raising the mean from 6.0 t ha^{-1} to 11.6 t ha^{-1}. The mean NH_3 volatilization is now 24 kg ha^{-1} yr $^{-1}$, reduced from 115 kg ha^{-1} yr $^{-1}$ in the runoff-calibrated model. The mean N stress also decreases from 46 days yr^{-1} to just 7.5 days yr^{-1}, and the predominant crop stress is now temperature stress.

With PARM72=0.05, changing PARM4 from 0.5 to 1.0 (max) has only a mini-

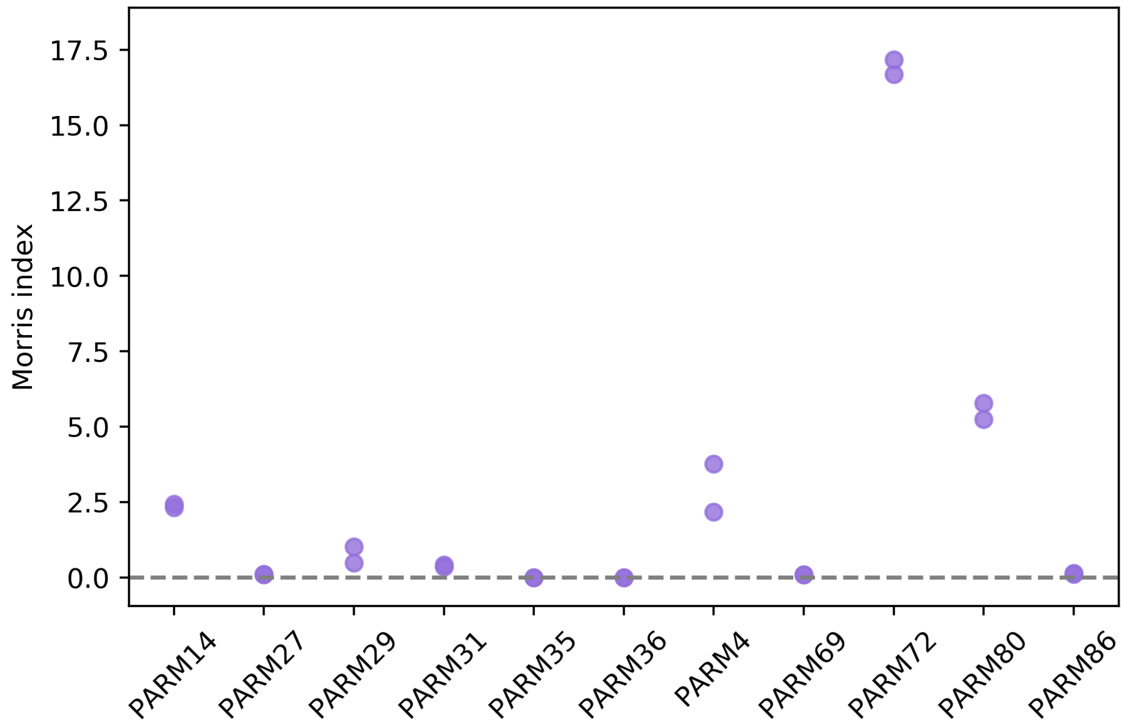

Figure 3.5: Results of APEX-CUTE sensitivity analysis for the PARMs listed in Table 3.2.

mal effect, increasing the mean yield to 11.8 t ha^{-1}. PARM4, which affects N lost via percolation, will therefore be left unchanged for crop yield and revisited during nutrient calibration. A final adjustment, decreasing each year's potential heat units parameter from 1333 to 1050, brings the yield to 12.3 t ha^{-1}. However, for reasons that are not clear, this raises the total runoff by about 57 mm. Rather than revisit the runoff calibration, PHU will be left unchanged.

Although the mean yield is now close to the statewide historical average and PBIAS=-6%, the NSE is a very poor -1.59. Given that the state yield records average over a wide range of soils, climate conditions, and farm management, it is to be expected that a model for a single site will not accurately capture the year-to-year variations in the records. Calibrating APEX so that the average yield approximately matches the statewide mean yield hopefully at least makes it likely that soil and crop processes are represented more accurately than they would have been had no yield calibration been attempted.

3.3.3 Erosion/Sediment

APEX calculates erosion using the Universal Soil Loss Equation (USLE) or one of several variants. These equations are of the form:

$$Y = X \times EK \times CVF \times PE \times SL \times ROKF \tag{3.12}$$

where X is the erosive energy, EK the soil erodibility factor, CVF the crop management factor, PE the erosion control practice factor, SL the slope length/steepness factor, and ROKF the coarse fragment factor. The erosion methods differ in the form that is taken by X. The USLE and RUSLE2 equations base X on rainfall, while the others (MUSS, MUSLE, etc.) relate X to runoff rate and volume in different ways. In addition there are two methods available for calculating SL.

Of the remaining factors in Eq 3.12, EK is related to soil texture and organic matter, CVF is based on residue cover, live biomass, and soil roughness, and PE represents the effectiveness of erosion control practices applied on the field. Some of the properties involved in EK and CVF may change over time, so could potentially alter the relative magnitudes of erosion events in ways that better reproduce the observed events. However, the PAW1 model was set up using soil test data and the crop yield has been calibrated, so EK and CVF should already have reasonable values (and the number of parameters that could affect EK and CVF is huge). Instead, the PE factor (corresponding to APEX's PEC parameter) can be used as a calibration parameter to generally increase or decrease sediment loss.

APEX allows the user to select the equation that is used to calculate the erosion that interacts with the other model components (DRV parameter), while also calculating and printing erosion for the remaining models. The MUSS equation was used for the baseline models and, now that the PAW1 model has been calibrated for runoff and crop yield, the total MUSS erosion is 2432 kg ha^{-1}, compared to the 1590 kg ha^{-1} measured at the site. The MUSLE model, which also bases X on runoff, estimates 5445 kg ha^{-1}. The erosion predicted by the models that base X on rainfall is much higher, ranging from 24,000 to 36,000 kg ha^{-1}. Therefore DRV=MUSS will continue to be used.

With DRV=MUSS, changing PEC from 1.0 to 0.66 brings the total erosion down to 1593 kg ha^{-1}, very close to the observed value. At this point NSE=0.22 and PBIAS=0.2%. This is well below the NSE$\geq$0.5 suggested by Wang et al. (2012), but there is no obvious way to improve those statistics. The erosion calibration had negligible effects on runoff and crop yield.

3.3.4 Nutrients

After calibration for runoff, crop yield, and erosion, the nutrient losses predicted by the model have changed substantially from the baseline values. As Figure 3.6 shows, P losses rose as the runoff was calibrated (i.e., as total runoff increased), then fell as crop yield was calibrated (i.e., as mean yield increased), then fell again or remained stable as erosion was calibrated (i.e., as total sediment loss decreased). Sediment N loss rose until yields had been calibrated, and then dropped slightly as sediment loss was decreased. Runoff N has risen slightly, but has remained well below the observed value throughout the process. At this point sediment-bound nutrients are lost in excess, while runoff nutrient losses are smaller than observed.

Partly, the trends in Figure 3.6 simply reflect changes in runoff and erosion; more runoff led to both more sediment loss and more of each form of nutrient loss. Decreasing erosion by reducing the PEC parameter reduced the amount of sediment-borne nutrient loss, but had little effect on losses in runoff. However, the crop yield calibration also had significant effects on nutrient losses. Although increasing the crop yield did decrease erosion from 3236 kg ha $^{-1}$ to 2430 kg ha $^{-1}$, this cannot fully explain the changes in N and P loss.

At least two underlying factors appear to be at work. First, the yield calibration greatly reduced the soluble P concentration in the top soil layer. This may be due to increased plant P uptake, and probably accounts for the large reduction in P in runoff at that time.

Second, the crop calibration changed the relationship between sediment-bound N (YN) and sediment (Y). Previously, APEX estimated a similar amount of N loss per unit Y as is found in the field data (Figure 2.20). After reducing NH_3 volatilization, decreasing N stress, and increasing biomass production, the model now calculates that more N is lost per unit Y than before.

This probably means that there is more organic N in the soil that can be lost as erosion occurs. In APEX, only organic (as opposed to mineral) N and P are lost with sediment, and the equation for sediment-borne nutrient loss is:

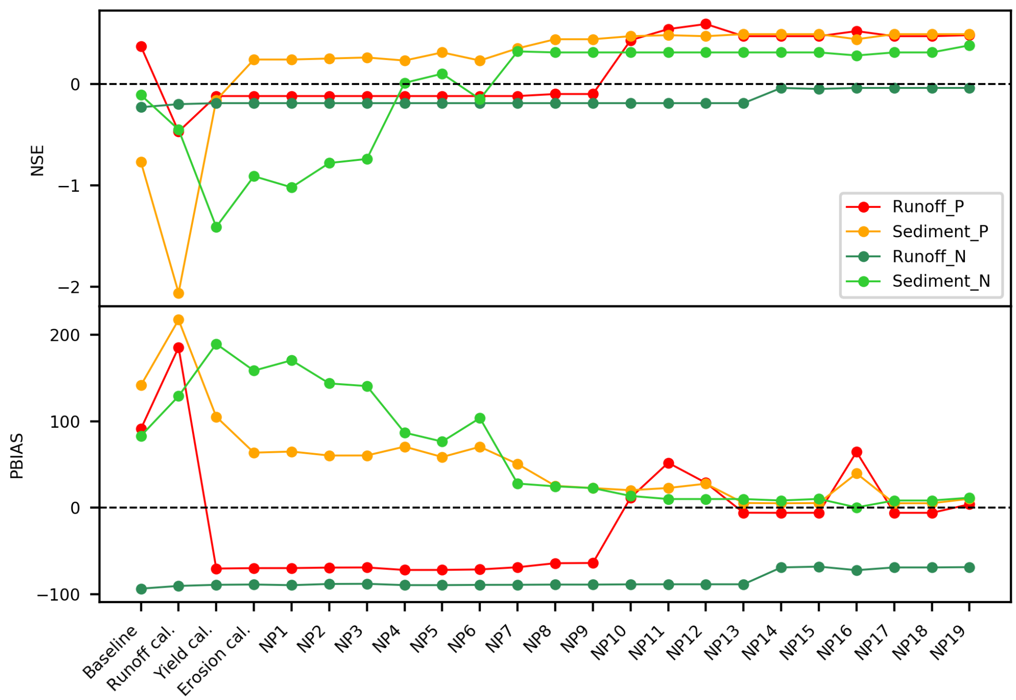

Figure 3.6: Changes in the ratio of model:observed total N and P losses as the PAW1 baseline model is calibrated for runoff, crop yield, sediment, and nutrient losses. X-axis labels relate to calibration steps; see text and Table 3.4 for full details.

$$YON = 0.001 \times Y \times CON \times ER \tag{3.13}$$

$$YP = 0.001 \times Y \times CP \times ER \tag{3.14}$$

where YON (YP) is the organic N (P) transported by sediment, Y is the sediment loss, CON (CP) is the concentration of organic N (P) in the top soil layer, and ER is the "enrichment ratio" (Williams et al., 2012). The enrichment ratio is the concentration of the nutrient in sediment vs that in the soil, and it is a function of rainfall and runoff. Now that sediment loss and runoff have been calibrated, Y and

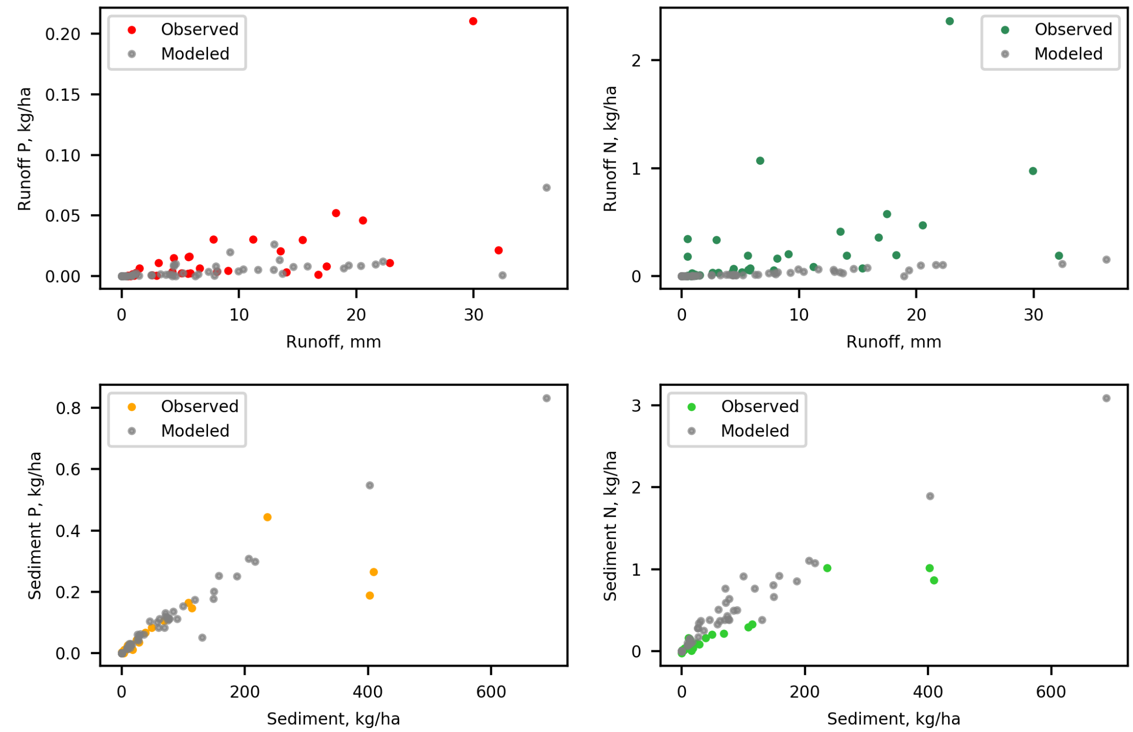

Figure 3.7: Model and observed relationships between nutrient losses and runoff (top) and sediment (bottom), on the PAW1 watershed after calibration for crop yield. Compare to Figure 2.20 for the baseline model.

ER are fixed. The remaining quantity affecting YN (YP) is CON (CP), which must be a function of the amount of organic matter in the top soil layer and the concentration of N and P within that organic matter. If increased crop growth leads to more crop residue and more organic matter production, for example, higher N (and, presumably, P) loss in sediment could make sense.

Examination of the model outputs reveals that organic N (the WON variable in the .ACN file) now builds up over time instead of decreasing (as in the baseline model; Figure 2.4). Probably more importantly, the amount of organic N *in the top 1 cm of the soil* – from where material is eroded – increases, both in absolute terms and as a fraction of the organic N in the top ~20 cm.

The situation for YP seems somewhat different. The crop calibration actually slightly improved the relationship between YP and Y compared to the baseline model. APEX does not report annual soil organic P masses, but initial and final N and P concentrations are recorded in the model's main log (.OUT) file. While the end-of-simulation organic N in the top 1cm increases by a factor of 1.6 from the baseline model to the yield-calibrated model, the organic P increases by just a factor of 1.2. Nitrogen and phosphorus cycling are calculated independently in APEX (Williams et al., 2012), even if they are coupled via several plant and soil processes, so this behavior is perhaps not surprising.

The above analysis suggests that parameters to do with N mineralization and immobilization and mixing of soil layers may improve the calibration for YN. Other parameters, such as the initial soluble P concentration, may also be relevant. Organic matter cycling and biological processes were chosen as the starting point for the nutrient calibration, focusing first on organic N and sediment N losses.

Again, an APEX-CUTE sensitivity analysis was carried out in order to identify parameters to which organic N is sensitive. The value of PARM72 was fixed by the crop calibration, but it is included in the sensitivity analysis so it can be used to gauge the relative sensitivities of other parameters. Table 3.3 lists the parameters included in the analysis, and Figure 3.8 shows the results. As described below, two additional parameters recommended by others were also used in the YN calibration, and parameters for calibrating YP, QN, and QP were selected without prior sensitivity analysis[5].

[5] APEX-CUTE crashes if sensitivity analysis for organic P is attempted. Also, although the software can calculate Morris Indices for "total" and "mineral" N and P, it does not discriminate

Table 3.3: APEX parameters used for organic N sensitivity analysis

Parameter	Definition[a]
PARM29	Biological mixing efficiency
PARM31	Maximum depth for biological mixing
PARM51	Water stored in litter (residue) coefficient
PARM52	Coefficient re. tillage effect on residue decay
PARM53	Coefficient modifying microbial activity with soil depth
PARM69	Adjusts microbial activity in top soil layer
PARM70	Microbial decay rate coefficient
PARM72	Volatilization/nitrification partitioning coefficient
PARM76	Standing dead fall rate coefficient
PARM100[b]	Century slow humus transformation coefficient
PEC[c]	Erosion control practice factor

[a] APEX User's Guide Section 2.21

[b] Repeatedly caused APEX-CUTE to crash, therefore not included in the final analysis. Also, PARM101 – 107, which are relevant to organic matter decomposition are not included in APEX-CUTE.

[c] PEC is included for the following reason. The Morris indices calculated by APEX-CUTE are in the units of the model output being evaluated, which means that the indices for nutrient loss are much smaller than those for runoff and biomass. The program only reports the indices to three decimal places, however, so indices for organic N tend to be rounded to either 0.001 or 0.000. Including the PEC parameter and allowing it to vary within a very narrow range near its maximum value (PEC=10) effectively scales up the sediment loss and therefore the amount of N lost in sediment. The resulting Morris indices for organic N are an order of magnitude larger and contain two significant figures' worth of information.

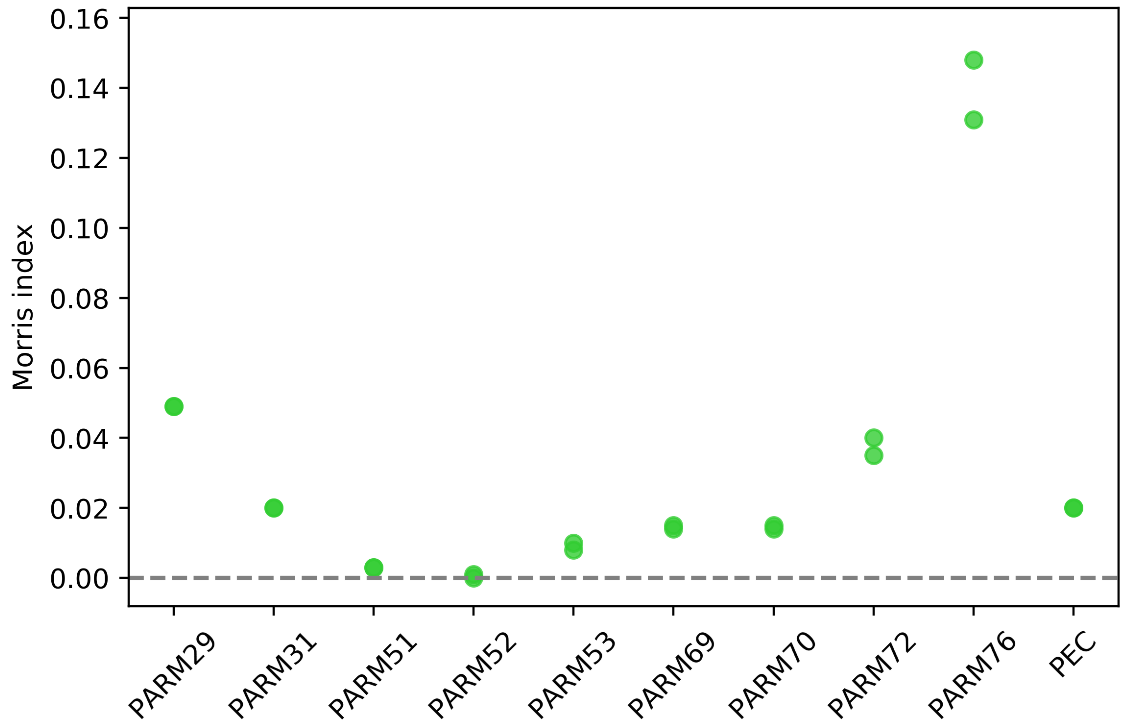

Figure 3.8: Results of APEX-CUTE sensitivity analysis for the PARMs listed in Table 3.3.

The sequence of calibration steps is laid out in Table 3.4 and illustrated in Figure 3.6. The apparently most influential parameter, PARM76 (steps NP1–3), was tested first but was found to have only minor effects on nutrient losses. Adjusting the biological mixing efficiency (PARM29, step NP4) had a larger effect on YN, presumably because it reduces the amount of N in the top 1cm of the soil. Increasing this parameter increased erosion by about 14%, which was corrected using the PEC parameter (§3.3.3). Changing the microbial activity function (PARM69, step NP7) also had a beneficial effect on YN.

between leaching, runoff, etc. Fortunately the number of parameters that appear directly relevant to QN and QP is fairly small.

The first parameter to result in noticeable improvement to YP was PARM70 (step NP8), the microbial decay rate coefficient. Increasing PARM70 also slightly increased the crop yield. However, raising PARM70 led to increased N leaching and denitrification (DN), perhaps related to increased mineralization of organic matter.

Increasing the slow humus turnover rate (PARM100, step NP9)) had little effect, but decreasing RTN1 from 150 to 0 (step NP10) brought the total QP much closer to the observed value and had a small positive effect on YN as well. RTN1 refers to the number of years of cultivation prior to the start of the simulation, and reducing it speeds up the mineralization rate. RTN1 is suggested by APEX staff as a possible calibration parameter for nutrient losses[6]. The soil soluble P value, which was much reduced after the crop calibration, increased to close to the expected value on setting RTN1=0. However, this change caused a further increase in N leaching and denitrification losses.

Wang et al. (2012) note that FHP, the fraction of humus in the passive pool, can be useful for calibrating nutrient losses. Changing it from its near-maximum default value to its minimum (step NP11) slightly improved YN while worsening QP. PARM8 sets the concentration of P in runoff relative to sediment, so this parameter was then used to bring QP back closer to the observed value (step NP12). A further reduction in both QP and YP was achieved by changing the parameter describing the upwards movement of P by evaporation (PARM59, step NP13).

At this point, the total YP, YN, and QN were close to the APME values. QN, however, was still very low, and N losses via leaching and denitrification had increased greatly. The mean DN rate in the NP13 model is 10.6 kg ha^{-1} yr^{-1}, which is 20% of

[6]https://groups.google.com/d/msg/agriliferesearchmodeling/Fw5jX7ZRK8o/vZKVkXkcCQAJ

the volatilization rate of 52 kg ha^{-1} yr^{-1} (which has itself more than doubled since the crop yield calibration). Given that Duncan et al. (2017) find that NO_3-N losses are at least 3 orders of magnitude below NH_3 losses in a manure injection system where DN rates are high, the denitrification predicted by APEX at this stage seems excessive. It appears that the model is directing too much mineral N into leaching, volatilization and denitrification, and not enough into runoff.

In an attempt to correct that, PARM14, which sets the relative N losses in percolate and runoff, was set to its maximum value (step NP14). This caused the total QN to increase by almost a factor of 3. It remained, though, more than a factor of 3 below the target value. PARM4 governs the amount of N that is lost in leachate vs that remaining in soil pores, but changing it had little effect on QN (steps NP15-16).

Finally, two parameters related to denitrification were tested. PARM35 sets the soil water content needed to trigger denitrification, and PARM36 limits the daily denitrification rate. Neither of these parameters had any effect on QN (steps NP17-18). It is certainly possible that there exist combinations of parameters that are able to simultaneously predict realistic losses of all nutrients at PAW1, but at this point it was decided that climate simulations should be prioritized over full nutrient calibration in the time remaining to the project.

The parameters for the final, "calibrated" model (step NP19) were chosen from the parameters that had generally beneficial effects on the total YN, YP, QN, and QP regardless of their effects on the underlying N-related processes. The parameter changes between the baseline and calibrated models, and performance statistics for the baseline and final models, are summarized in §3.3.5. Briefly, for YN, YP, and QP, PBIAS $\leq$12% and $0.38 < NSE < 0.49$. For QN, PBIAS=-68% and NSE=-0.04.

Statistics for runoff and erosion have changed very little, and the mean crop yield is now slightly closer to the statewide average.

Table 3.4: PAW1 nutrient calibration process

Step	Parameter Change[a]	Comments
NP1	P76 0.01→0.1 (max)	
NP2	P76 0.01→0.001	
NP3	P76 0.01→0.0001 (min)	Some improvement in YN
NP4	P29 0.1 (min)→0.5 (max)	Larger improvement in YN
NP5	As for NP4 plus PEC 0.66→0.58	Erosion had risen slightly; correcting
NP6	As for NP5 plus P31 0.3 (max)→0.1 (min)	
NP7	As for NP5 plus P69 1.0 (max) →0.1 (min)	YN improved
NP8	As for NP7 plus P70 0.5 (min)→1.5 (max)	Slight improvement in YP and QP
		N leaching and denitrification increased
		Crop yield slightly increased
NP9	As for NP8 plus P100 0.000548→0.00068 (max)	
NP10	As for NP9 plus RTN1 150→0 (min)	QP much improved
		Large increase in soil soluble P
		N leaching and denitrification increased again
NP11	As for NP10 plus FHP 0.695 (max)→0.3 (min)	QP NSE slightly better, PBIAS worse
NP12	As for NP11 plus P8 15→20 (max)	Improves QP, YP slightly worse
NP13	As for NP11 plus P59 10→3	YN, YP, QP now quite good, QN still v. low
NP14	As for NP13 plus P14 0.2→1.0 (max)	Some improvement in QN
NP15	As for NP14 plus P4 0.5→1.0 (max?)	
NP16	As for NP14 plus P4 0.5→0.0 (min)	
NP17	As for NP14 plus P36 0.001→0.0001 (min)	
NP18	As for NP17 plus P35 1.01→1.10 (max)	
NP19	Erosion-calibrated model plus P8=10, P14=1, P29=0.5, P59=3, P69=0.1, P70=1.5, P76=0.0001, RTN1=0, PEC=0.62	Final model

[a] See text and Table 3.3 for parameter definitions.

Table 3.5: Differences between PAW1 baseline and calibrated models

Process	Variable	Description	Base.	Cal.
Runoff	PARM42	CN index coefficient	1.0	2.5
	PARM44	Upper limit of CN retention parameter	1.5	1.2
Yield	PARM72	Volatilization/nitrification partition coeff.	0.4	0.05
Sediment	PEC	Erosion control practice factor	1.0	0.66[a]
Nutrients	PARM8	Soluble P runoff coeff.	15	10
	PARM14	Nitrate leaching ratio	0.2	1.0
	PARM29	Biological mixing efficiency	0.1	0.5
	PARM59	P movement by evaporation coeff.	10	3
	PARM69	Microbial activity coeff.	1.0	0.1
	PARM70	Microbial decay rate coeff.	0.5	1.5
	PARM76	Standing dead fall rate coeff.	0.01	0.0001
	RTN1	Years of cultivation at simulation start	150	0
	PEC	Erosion control practice factor	1.0	0.62

[a] Further adjustment made during nutrient calibration

3.3.5 Final Model

Table 3.5 lists the differences between the baseline and calibrated PAW1 models, and Table 3.6 gives the full set of performance metrics for both models. Updated versions of the model – data comparison plots shown in Ch. 2 are shown in Figures 3.9 – 3.18.

Overall the statistics have improved. This is not surprising, as the calibration procedure was based on monitoring the PMs and totals as parameters were changed. Figures 3.9 and 3.10 show that the model relationships between sediment and runoff, and nutrient losses and sediment/runoff, are now closer to those observed in the field data. This generally positive. However, it should be noted that all model outputs except runoff systematically underestimate the magnitude of large events (Figures 3.12 – 3.16). This may be relevant in interpreting the results of the climate scenario modeling in Chapter 4.

Table 3.6: Model performance statistics for the PAW1 initial baseline and final calibrated models[a]

Model	Component	R^2	d	RMSE	RSR	NSE	PBIAS	p	Mean		Sum	
									Obs	Mod	Obs	Mod
Baseline	Runoff	0.47	0.81	6.71	0.77	0.41	-24.1	0.15	6.93	5.26	402	305
	Yield	0.02	0.24	6.78	5.23	-26.4	-52.6	0.00	12.6	6.0	265	126
	Sediment	0.17	0.61	97.8	1.02	-0.05	48.0	0.04	40.8	60.4	1590	2354
	Runoff P	0.64	0.86	0.03	0.79	0.37	91.1	0.07	0.01	0.03	0.57	1.09
	Sediment P	0.53	0.75	0.12	1.33	-0.77	141.4	0.01	0.04	0.11	1.74	4.19
	Runoff N	0.08	0.33	0.47	1.11	-0.23	-93.3	0.00	0.23	0.02	8.97	0.60
	Sediment N	0.30	0.70	0.27	1.05	-0.11	83.1	0.01	0.13	0.23	4.90	8.97
Calibrated	Runoff	0.52	0.85	6.35	0.73	0.47	-1.23	0.90	6.93	6.85	402	397
	Yield	0.01	0.35	2.22	1.72	-1.95	-2.30	0.30	12.6	12.3	265	259
	Sediment	0.23	0.60	84.3	0.88	0.22	-3.85	0.07	40.8	39.2	1590	1529
	Runoff P	0.51	0.77	0.02	0.72	0.48	4.09	0.22	0.01	0.02	0.57	0.59
	Sediment P	0.49	0.81	0.06	0.72	0.49	10.5	0.07	0.04	0.05	1.74	1.92
	Runoff N	0.13	0.34	0.43	1.02	-0.04	-68.5	0.13	0.23	0.07	8.97	2.83
	Sediment N	0.39	0.74	0.20	0.79	0.38	11.6	0.04	0.13	0.14	4.90	5.47

[a] Units of RMSE, Mean, and σ are mm; others are dimensionless.

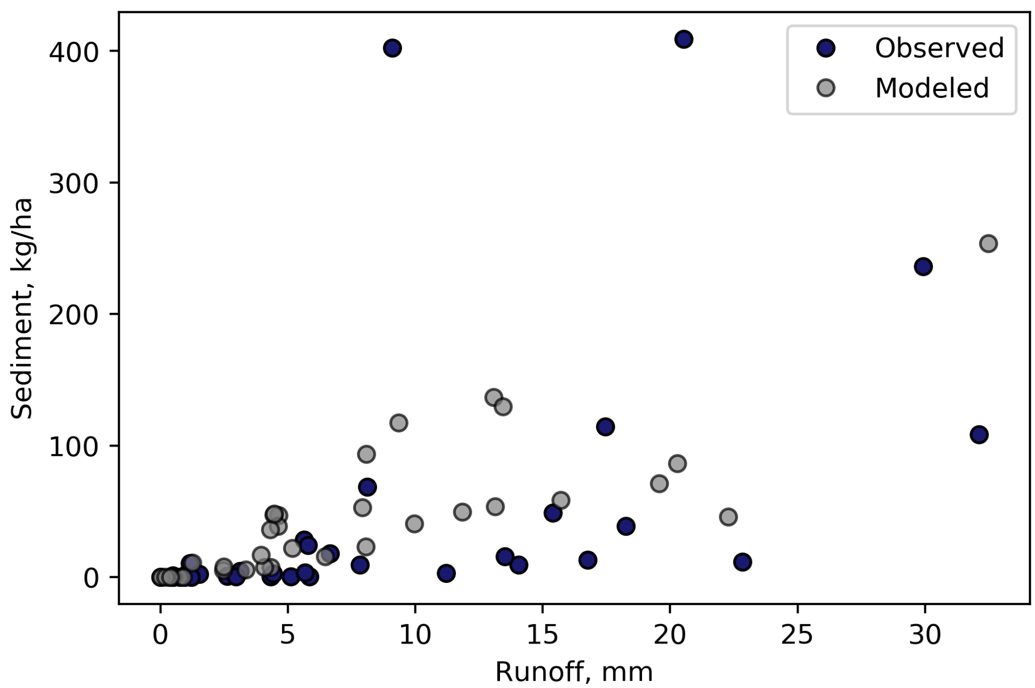

Figure 3.9: Model and observed relationships between sediment and runoff in the PAW1 final model. Compare to Figure 2.18 for the baseline model.

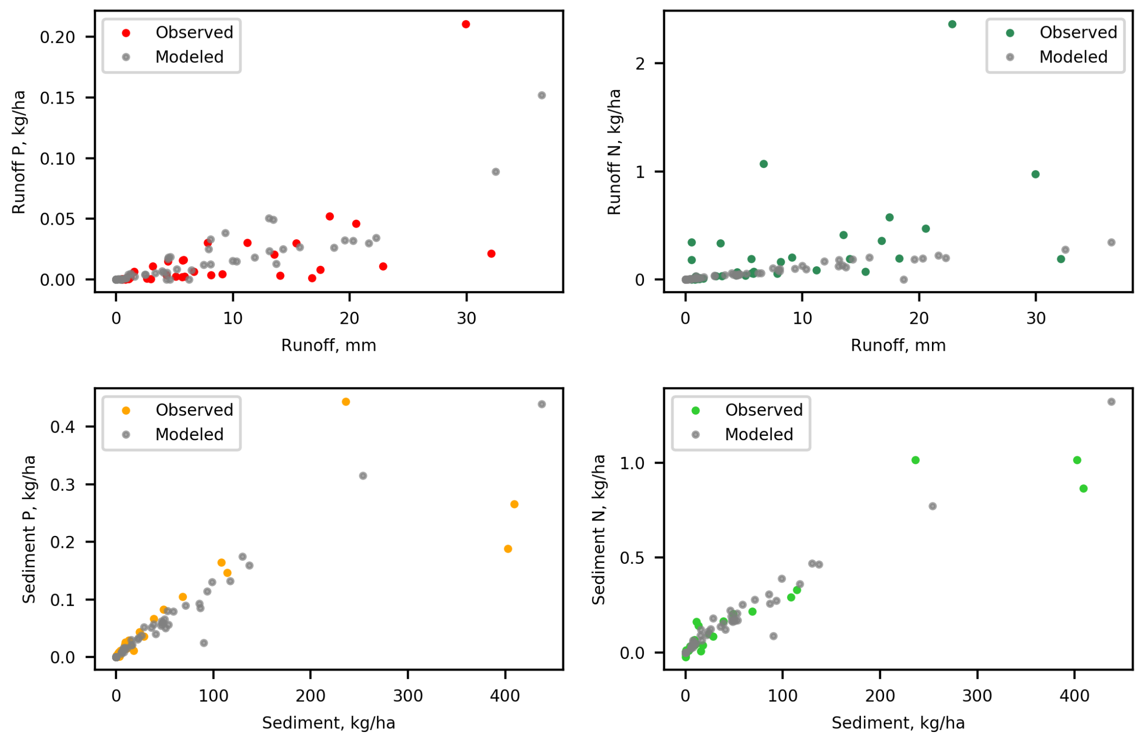

Figure 3.10: Model and observed relationships between nutrient losses and runoff (top) and sediment (bottom) in the PAW1 final model. Compare to Figure 2.20 for the baseline model.

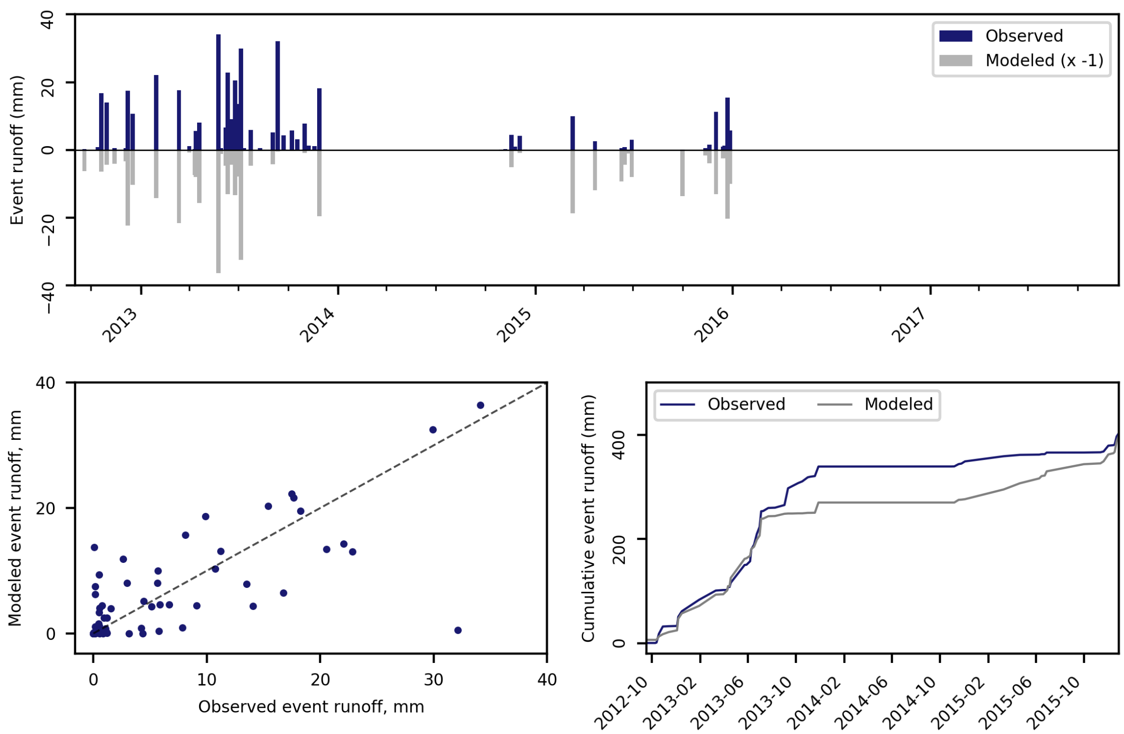

Figure 3.11: Runoff in the PAW1 calibrated model. Compare to the baseline version in Figure 2.16.

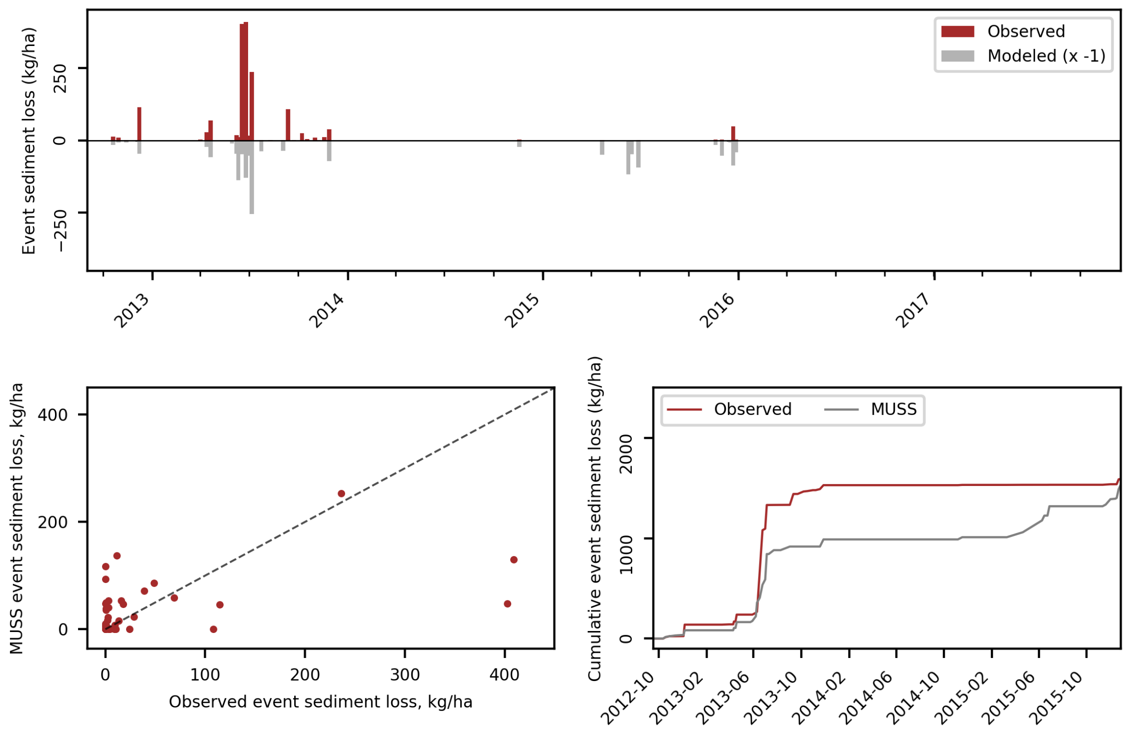

Figure 3.12: Erosion in the PAW1 calibrated model. Compare to the baseline version in Figure 2.17

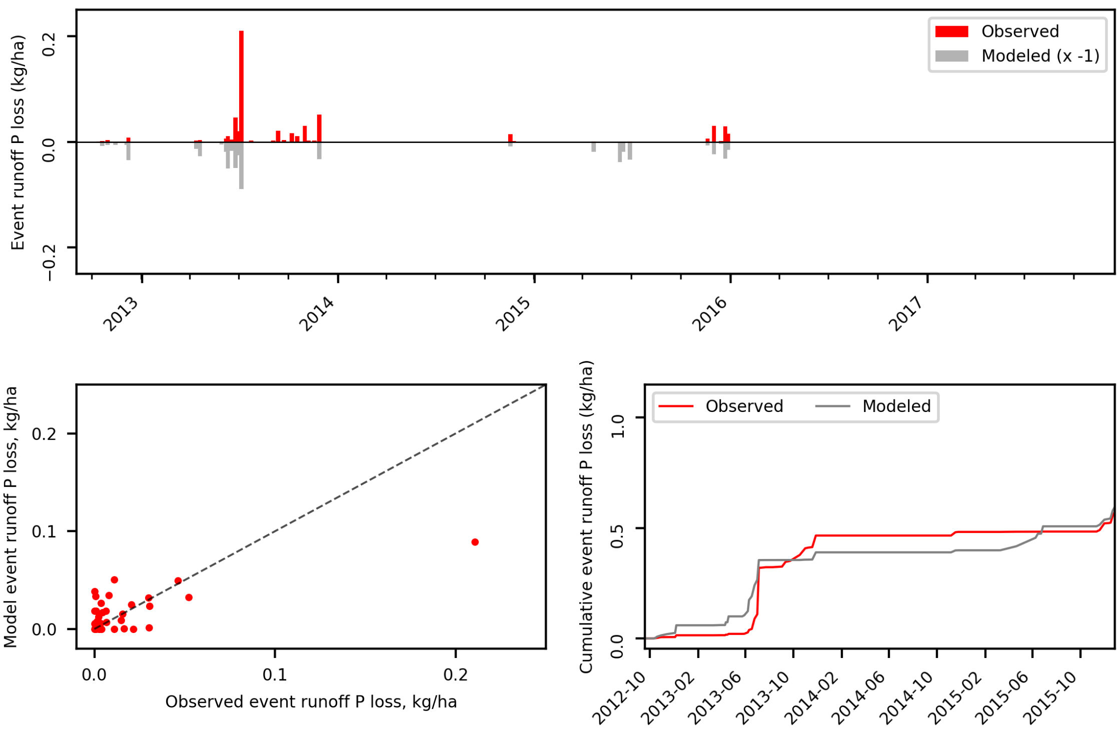

Figure 3.13: Runoff P in the PAW1 calibrated model. Compare to the baseline version in Figure 2.19

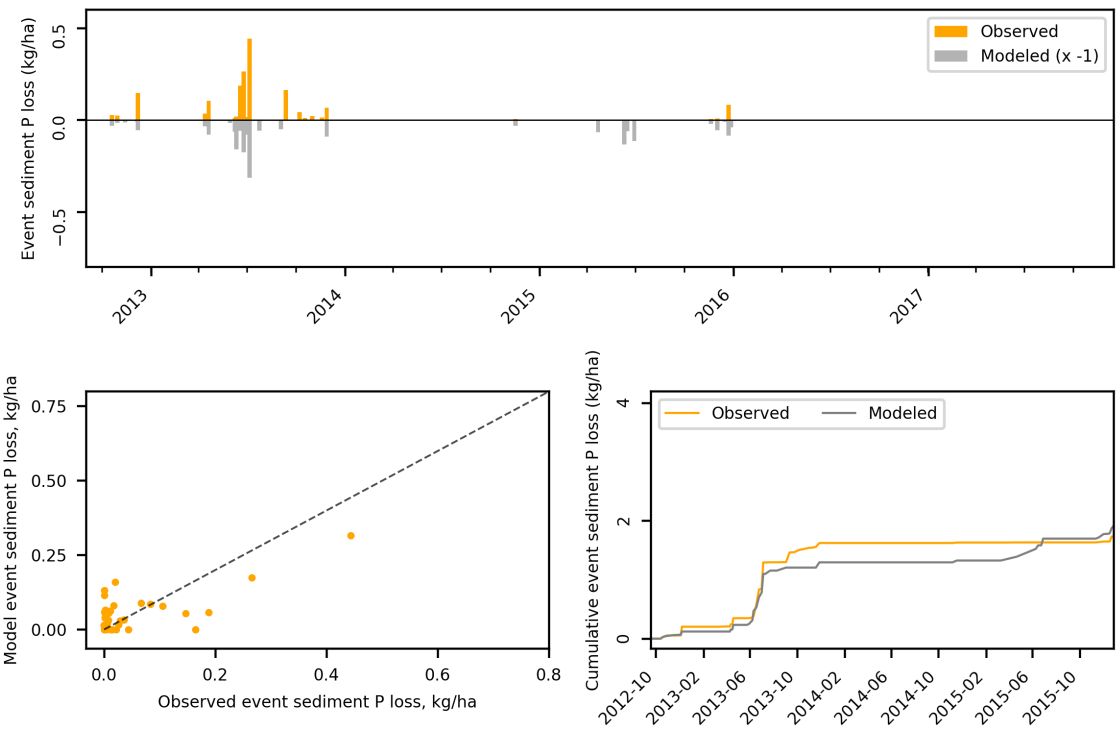

Figure 3.14: Sediment P in the PAW1 calibrated model. Compare to the baseline version in Figure 2.21

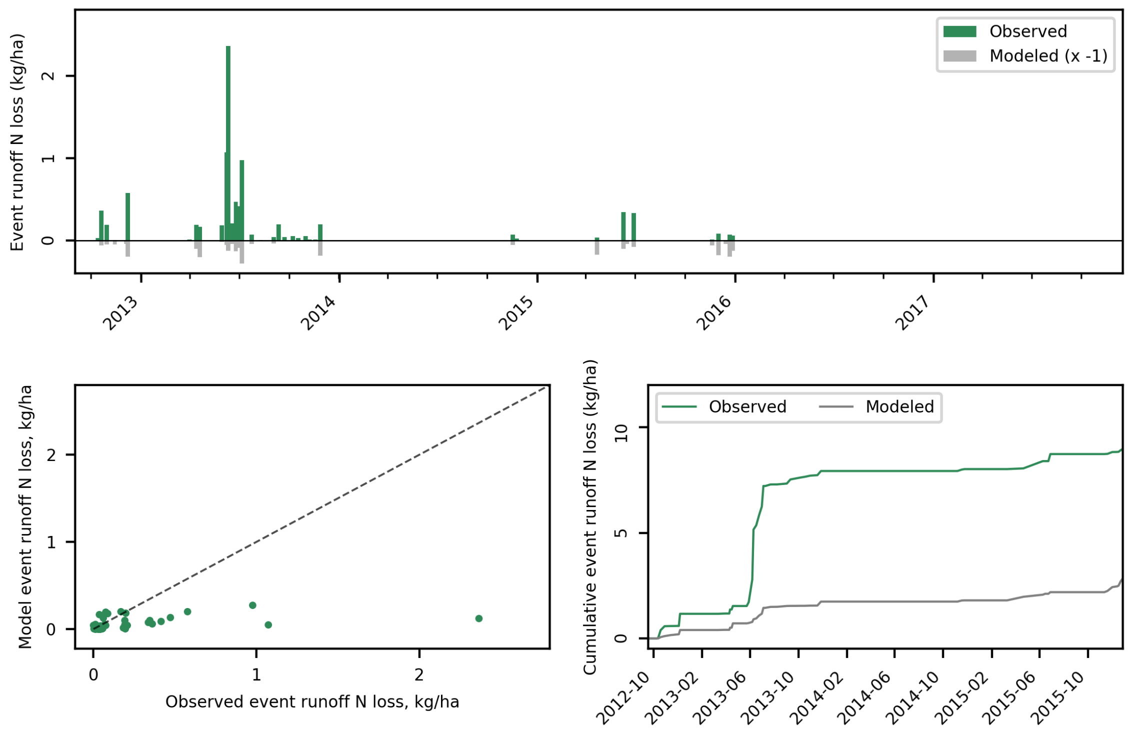

Figure 3.15: Runoff N in the PAW1 calibrated model. Compare to the baseline version in Figure 2.22

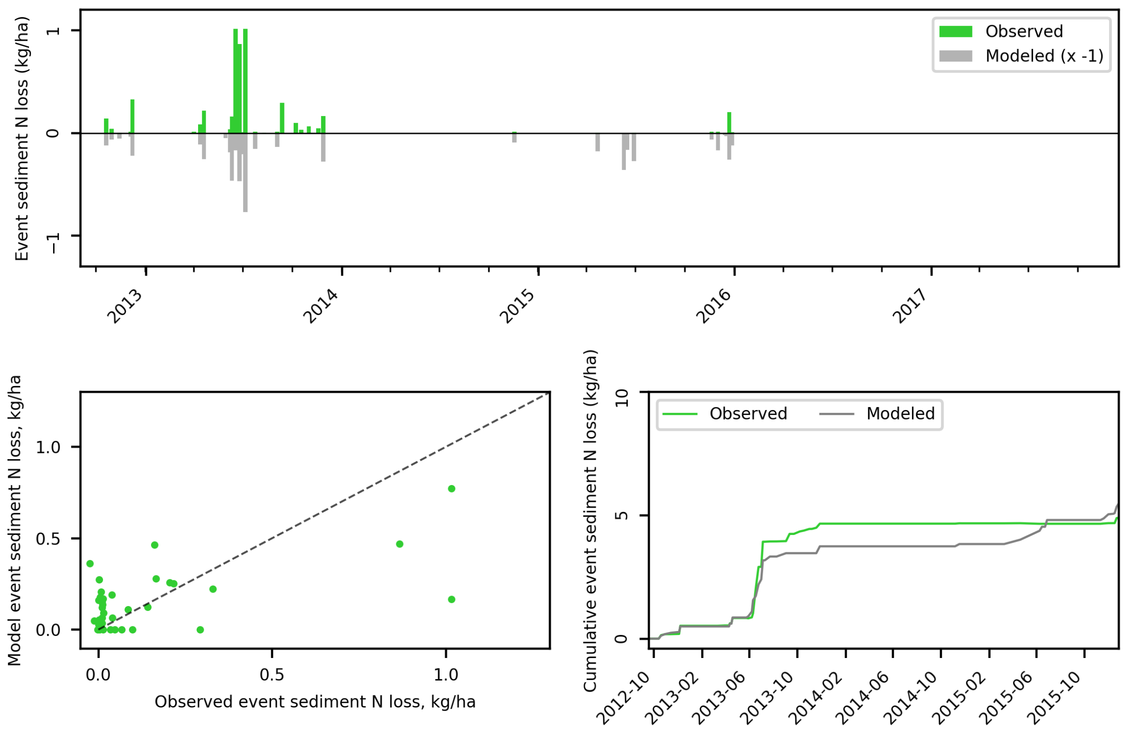

Figure 3.16: Sediment N in the PAW1 calibrated model. Compare to the baseline version in Figure 2.23

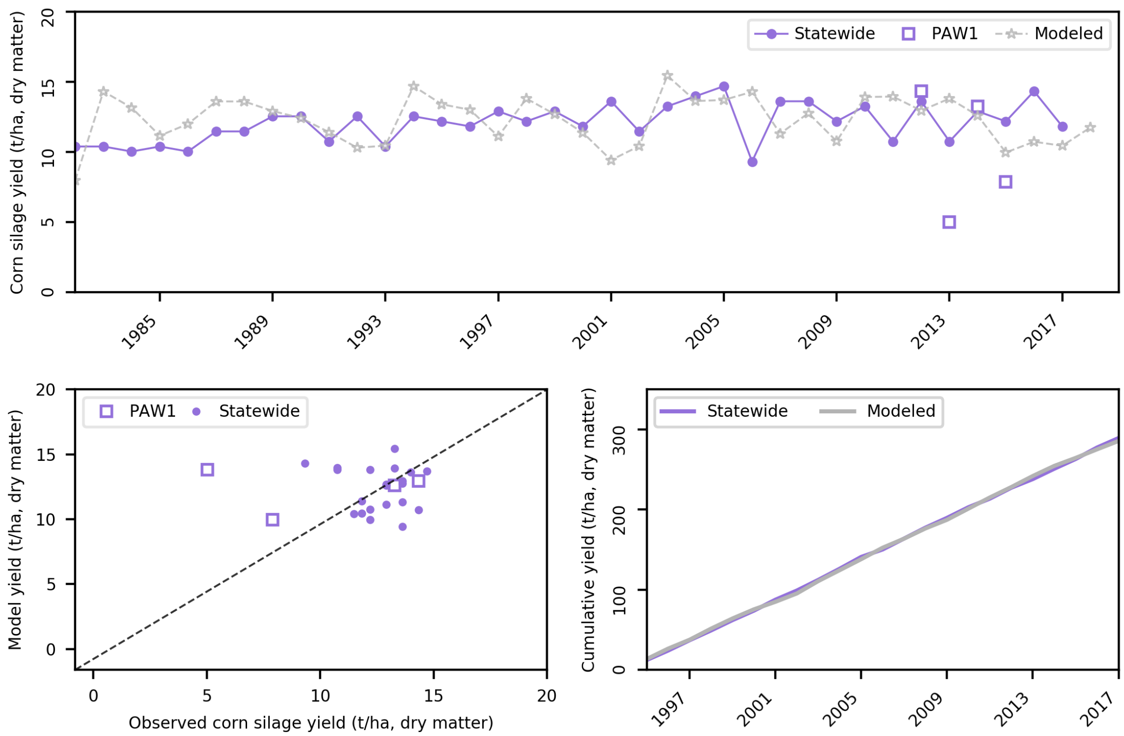

Figure 3.17: Forage yield in the PAW1 calibrated model. Compare to the baseline version in Figure 2.24

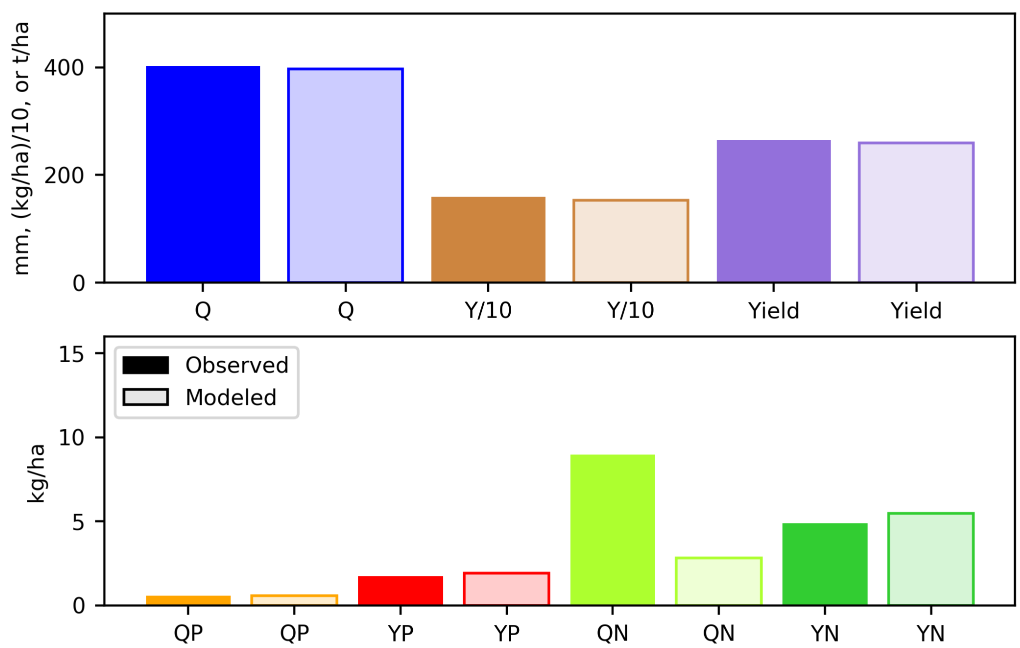

Figure 3.18: Total observed and modeled runoff, erosion, etc. for the duration of the APME project at PAW1 (Q=runoff, Y/10=sediment divided by 10, QP/N= runoff P/N, YP/N=sediment P/N). Compare to the baseline version in Figure 2.34.

3.4 Model Calibration: WIL2

The WIL2 calibration followed the same basic path as the PAW1 calibration. The first difference was that runoff needed to be reduced overall, rather than increased. In the PAW1 model, adjusting PARM42 and PARM44 with NVCN=4 (the baseline value for the daily curve number method) was sufficient to raise the total runoff to match the field measurement (§3.3.1). At WIL2, runoff could not be lowered sufficiently with those parameters alone (Figure 3.19). What reductions were achieved came partly from decreasing the model runoff in 2015 and 2016 essentially to zero, worsening the match between model and data in those years. Experience with the PAW1 runoff calibration suggested that manipulating further parameters with NVCN=4 would be unlikely to fix this problem.

With NVCN=0 and PARM92=1.3, the model was able to give the correct total runoff while preserving some runoff in 2015 – 2016. For that model, PBIAS=0.3% and NSE=0.40. Those parameters were adopted for the remainder of the calibration.

As with PAW1, crop yields in the WIL2 baseline model were low and N stress and volatilization were high. Again, changing PARM72 reduced volatilization and N stress and brought the mean yield to match the statewide annual average. After reducing runoff and increasing yields, the model erosion had decreased relative to the baseline model, and had in fact become lower than the observed value. Three ways of bringing the total erosion to the observed value were identified:

1. With DRV=MUSS, raise PEC from 1.0 to 2.12.
    - NSE=0.77, PBIAS=-0.9%

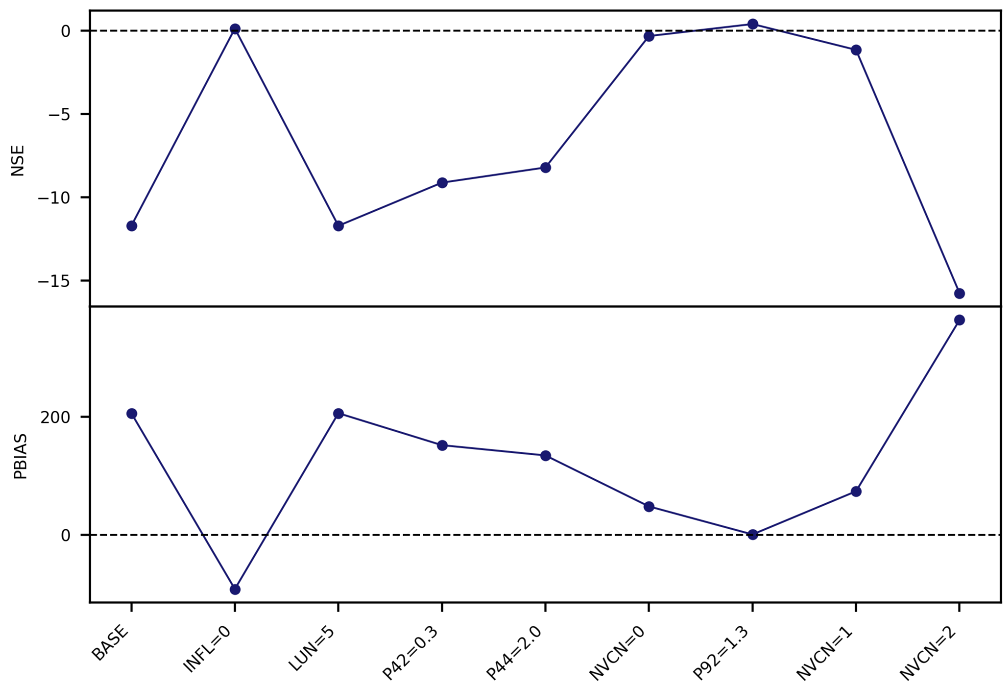

Figure 3.19: Changes in performance statistics as the WIL2 baseline model is calibrated for runoff. X-axis labels relate to parameter changes; see text for full details.

2. With DRV=MUSLE, change ISLF from 0 (RUSLE SL factor) to 1 (MUSLE SL factor) and PEC from 1.0 to 0.77
   - NSE=0.78, PBIAS=0.9%
3. With DRV=RUSLE2, change PEC from 1.0 to 0.035
   - NSE=0.65, PBIAS=3%

The first two methods give essentially the same outcome, while the third results in a slightly poorer pattern of erosion event magnitudes. While all three methods are somewhat arbitrary and arguably unphysical, method (1) was adopted for the erosion-calibrated model.

WIL2 nutrient calibration was attempted using the parameters that had been beneficial for nutrient calibration at PAW1 (Table 3.4). Statistics for YN remained extremely poor through the process (Figure 3.20). PBIAS for QN and QP could be brought close to zero (Figure 3.20), but no model with NSE >0.1 could be identified for QN, and NSE was always negative for QP. Calibration of YP was relatively successful, with NSE=0.82 and PBIAS=8%. However, given the uncertainty in underlying soil processes suggested by the poor model performance for the other nutrient losses, further investigation is recommended before using this model to make predictions for sediment-bound P.

3.4.1 Final Model

Table 3.7 lists the differences between the baseline and calibrated WIL2 models, and Table 3.8 gives the full set of performance metrics for both models. Updated versions of the model – data comparison plots shown in Chapter 2 are shown in Figures 3.21 – 3.30.

As with the PAW1 model, the statistics for the model components that are considered to have been successfully calibrated (runoff, sediment, yield, and sediment-bound P) have improved. The relationships between runoff and sediment, and nutrients and runoff/sediment, also look better (Figures 3.21, 3.22). The WIL2 model is less prone to underestimating large events than the PAW1 model, although the smaller number of data points makes it difficult to draw firm conclusions.

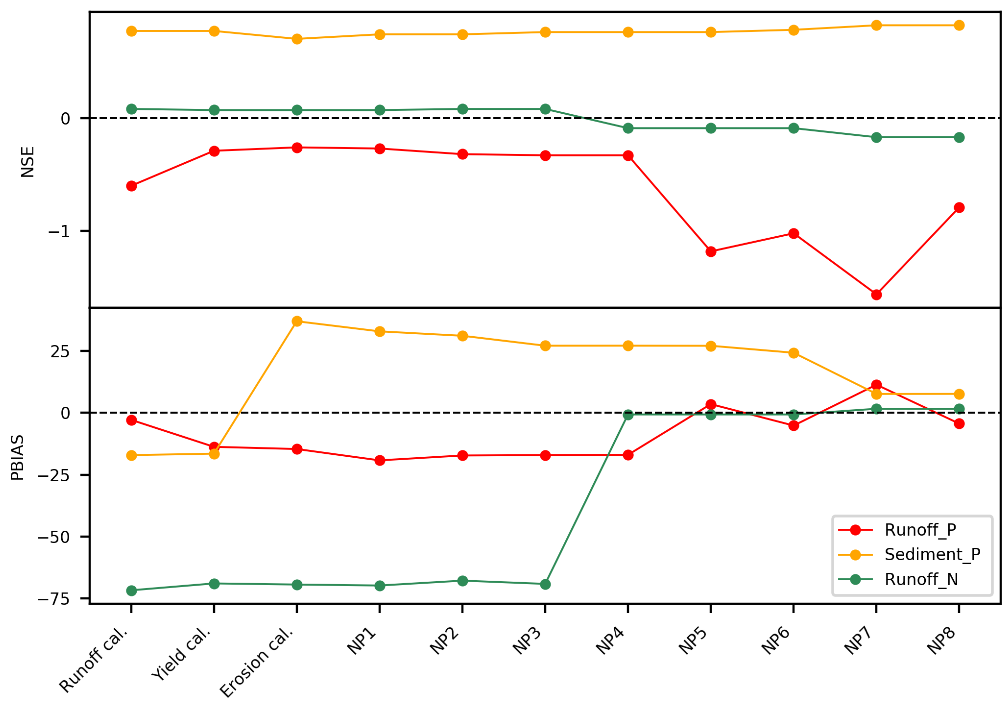

Figure 3.20: NSE and PBIAS for runoff N and P and sediment P during nutrient calibration at WIL2. Nutrient loss statistics were very poor for the baseline model and for YN for all models, so they have been omitted for clarity.

Table 3.7: Differences between WIL2 baseline and calibrated models

Process	Variable	Description	Baseline	Cal
Runoff	NVCN	Daily curve number calc. method	4	0
Runoff	PARM92	Runoff volume adjustment	1.0	1.3
Yield	PARM72	Volatilization/nitrification partition coeff.	0.4	0.1
Sediment	PEC	Erosion control practice factor	1.0	2.12
Nutrients	PARM8	Soluble P runoff coeff.	15	14
	PARM14	Nitrate leaching ratio	0.2	0.7
	PARM29	Biological mixing efficiency	0.1	0.5
	PARM59	P movement by evaporation coeff.	10	8
	PARM69	Microbial activity coeff.	1.0	0.1
	PARM70	Microbial decay rate coeff.	0.5	1.5
	PARM76	Standing dead fall rate coeff.	0.01	0.0001

Table 3.8: Model performance statistics for the WIL2 initial baseline and final calibrated models[a]

Model	Component	R^2	d	RMSE	RSR	NSE	PBIAS	p	Mean		Sum	
									Obs	Mod	Obs	Mod
Baseline	Runoff	0.73	0.54	20.0	3.56	-11.7	206	0.00	2.98	9.12	152	465
	Yield	0.01	0.32	3.24	2.5	-5.25	-17.0	0.00	12.6	10.5	265	220
	Sediment	0.83	0.94	23.5	0.57	0.68	26.2	0.00	17.4	21.9	469	591
	Runoff P	0.52	0.30	0.32	7.49	-55.1	450	0.00	0.02	0.13	0.73	4.02
	Sediment P	0.79	0.83	0.10	1.20	-0.45	123	0.00	0.03	0.07	0.95	2.12
	Runoff N	0.70	0.89	0.09	0.78	0.39	-3.46	0.00	0.06	0.06	1.87	1.81
	Sediment N	0.79	0.30	1.21	8.80	-76.4	738	0.00	0.06	0.51	1.82	15.3
Calibrated	Runoff	0.70	0.89	4.37	0.78	0.39	0.89	0.00	2.98	3.01	152	153
	Yield	0.11	0.57	1.83	1.41	-0.99	3.30	0.53	12.6	13.1	265	274
	Sediment	0.80	0.95	19.6	0.47	0.78	-1.65	0.00	17.4	17.1	469	461
	Runoff P	0.48	0.75	0.06	1.34	-0.79	-4.38	0.00	0.02	0.02	0.73	0.70
	Sediment P	0.82	0.95	0.04	0.42	0.82	7.63	0.00	0.03	0.03	0.95	1.02
	Runoff N	0.18	0.65	0.13	1.08	-0.17	1.62	0.01	0.06	0.06	1.87	1.90
	Sediment N	0.76	0.44	0.69	5.04	-24.4	412	0.00	0.06	0.31	1.82	9.31

[a] Units of RMSE, Mean, and σ are mm; others are dimensionless.

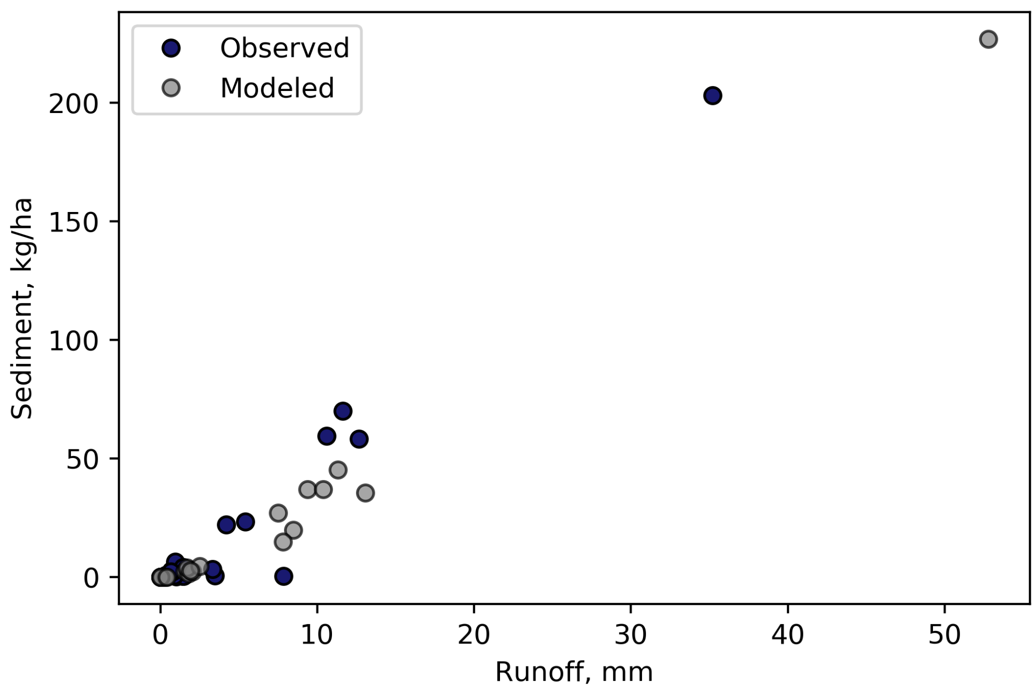

Figure 3.21: Model and observed relationship between sediment and runoff in the WIL2 final model. Compare to Figure 2.26 for the baseline model.

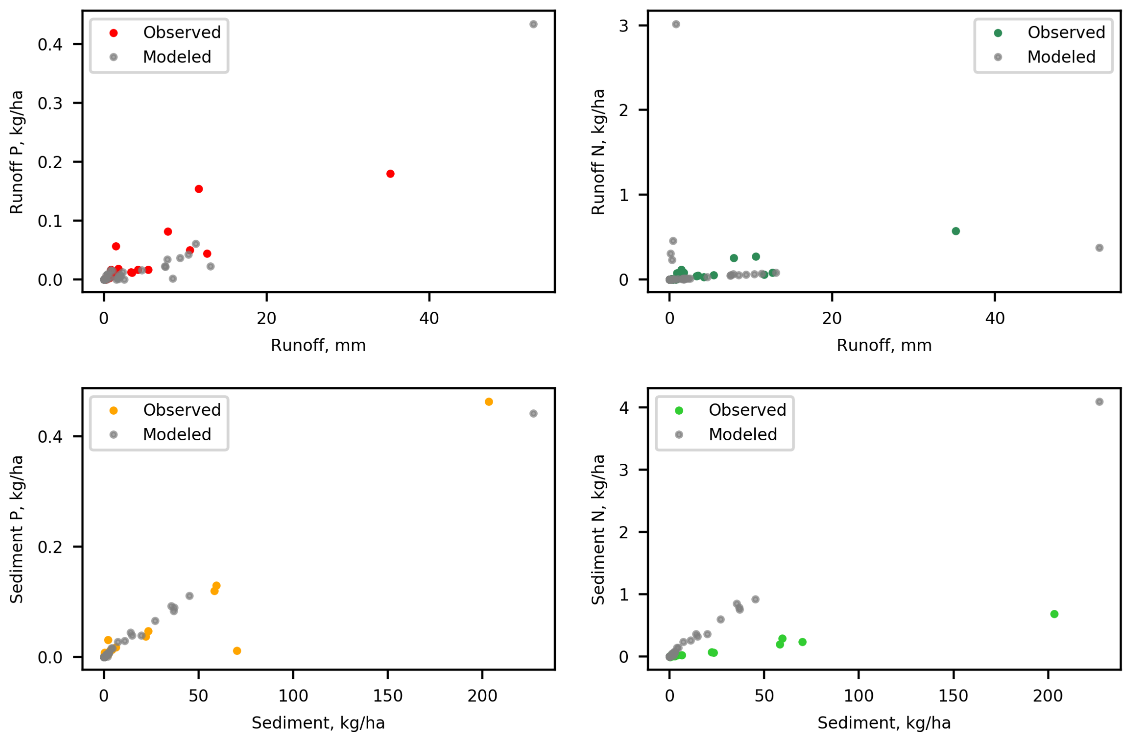

Figure 3.22: Model and observed relationships between nutrient losses and runoff (top) and sediment (bottom) in the WIL2 final model. Compare to Figure 2.28 for the baseline model.

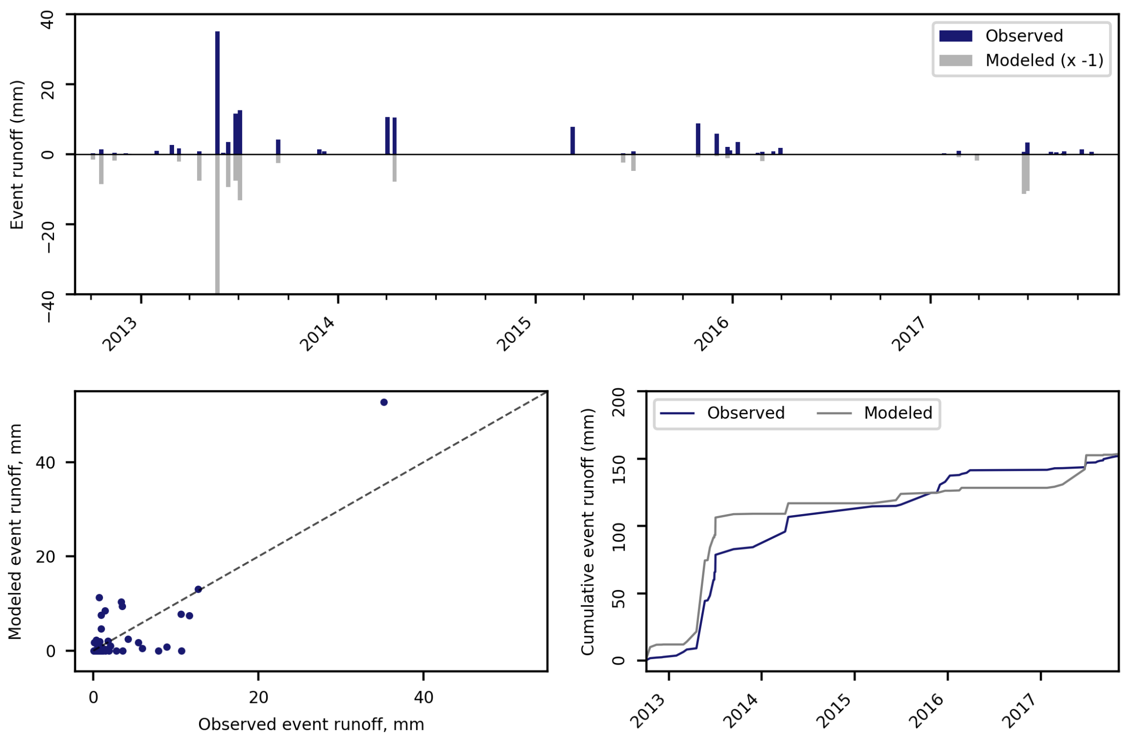

Figure 3.23: Runoff in the WIL2 calibrated model. Compare to the baseline version in Figure 2.25.

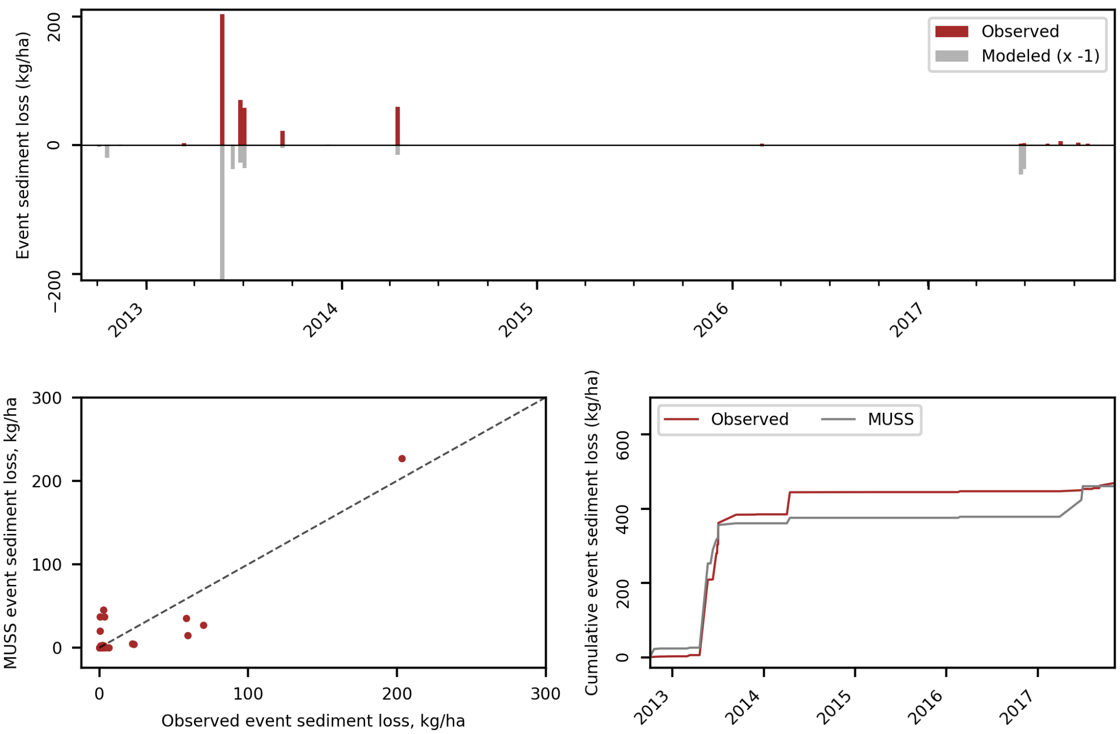

Figure 3.24: Erosion in the WIL2 calibrated model. Compare to the baseline version in Figure 2.27.

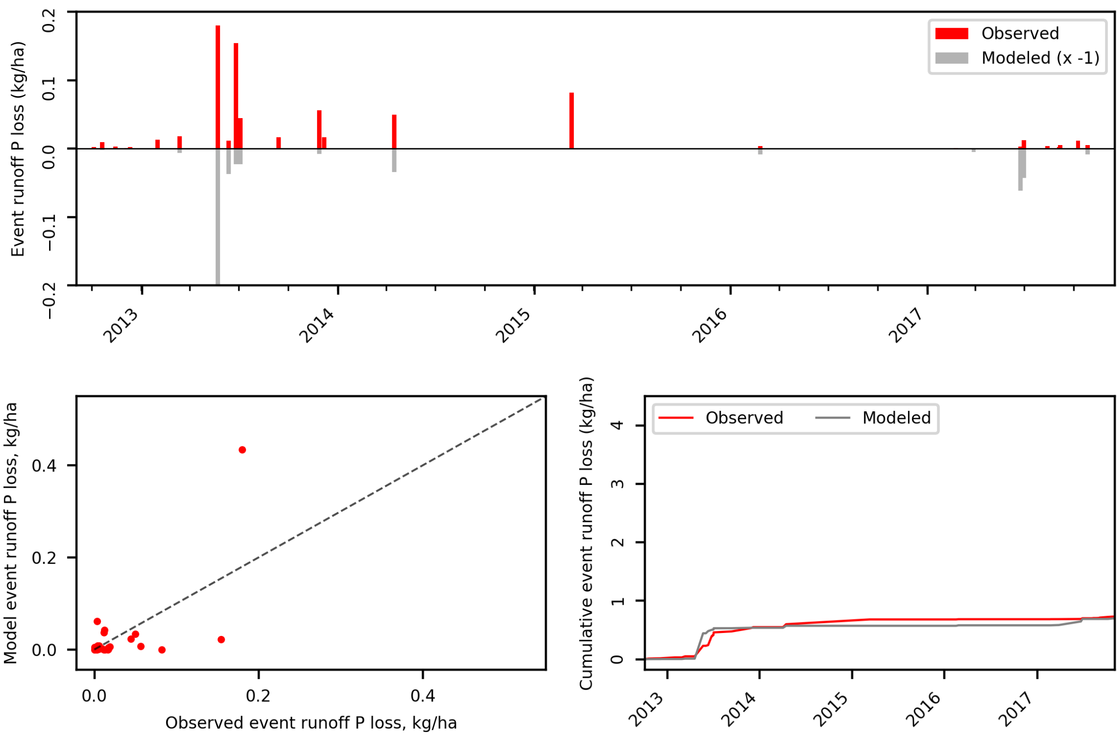

Figure 3.25: Runoff P in the WIL2 calibrated model. Compare to the baseline version in Figure 2.29.

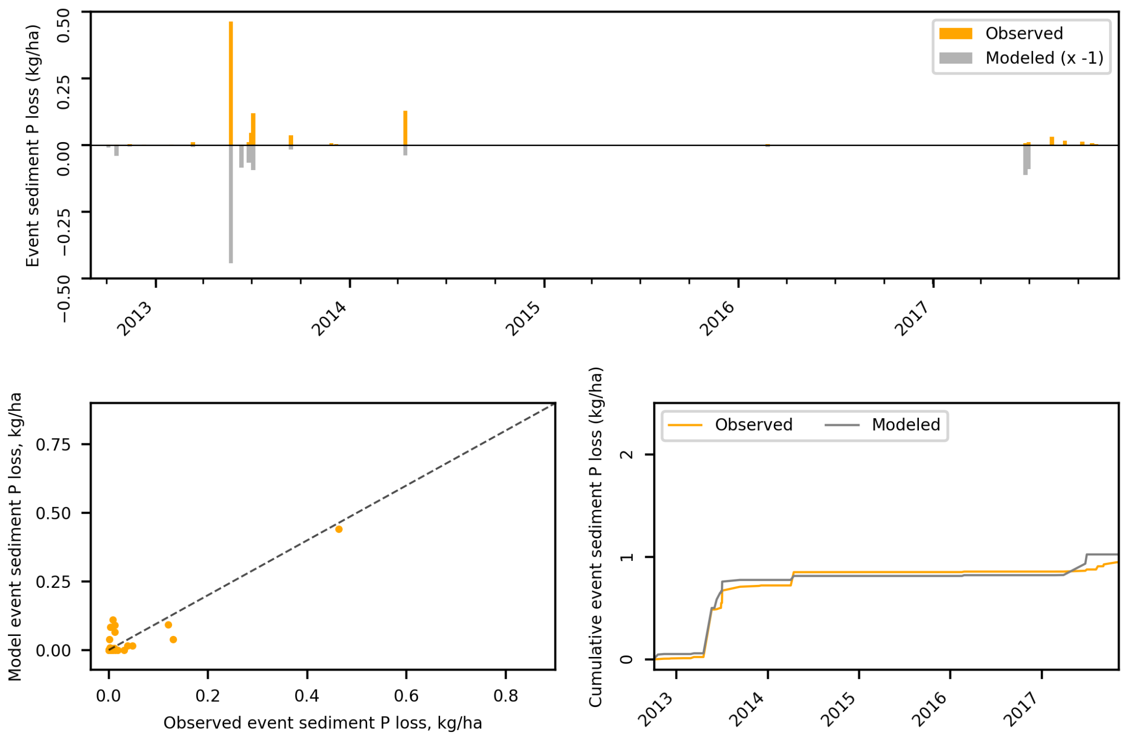

Figure 3.26: Sediment P in the WIL2 calibrated model. Compare to the baseline version in Figure 2.30.

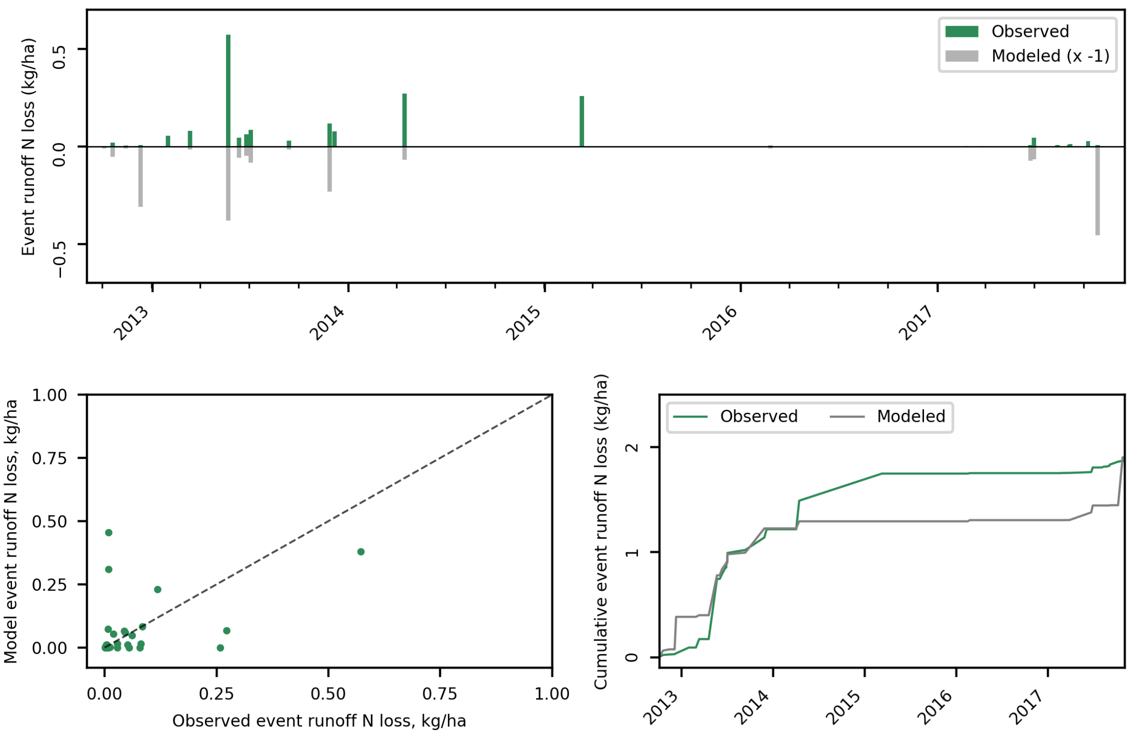

Figure 3.27: Runoff N in the WIL2 calibrated model. Compare to the baseline version in Figure 2.31.

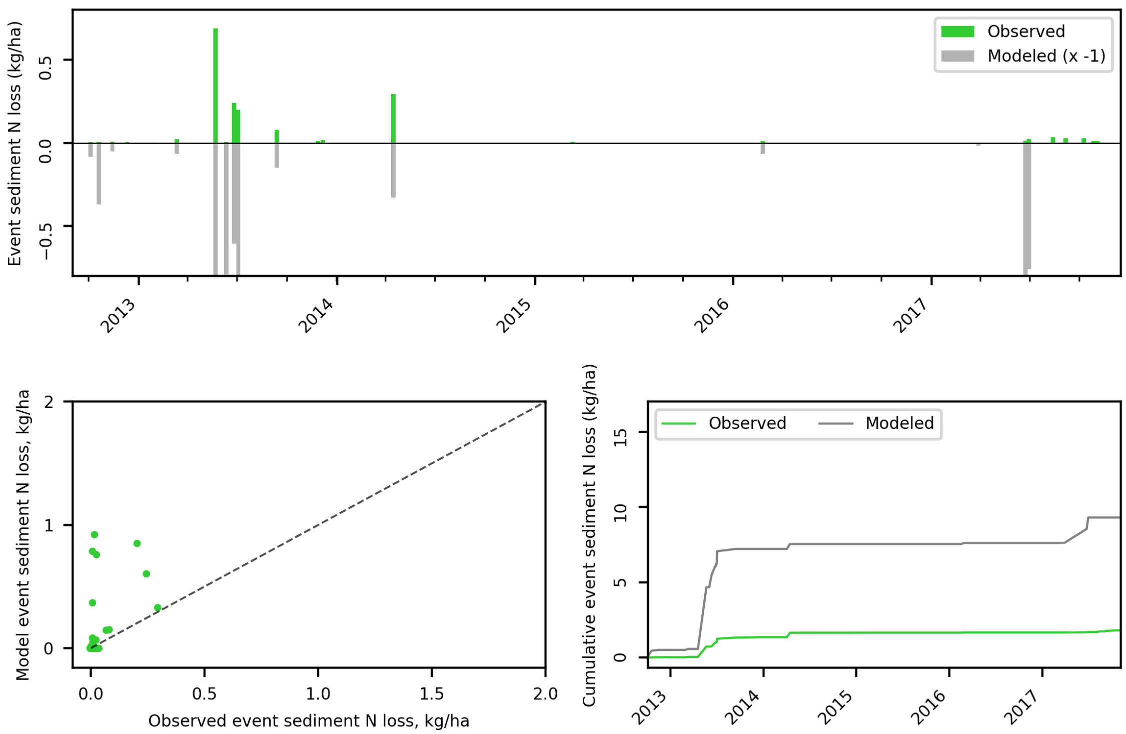

Figure 3.28: Sediment N in the WIL2 calibrated model. Compare to the baseline version in Figure 2.32.

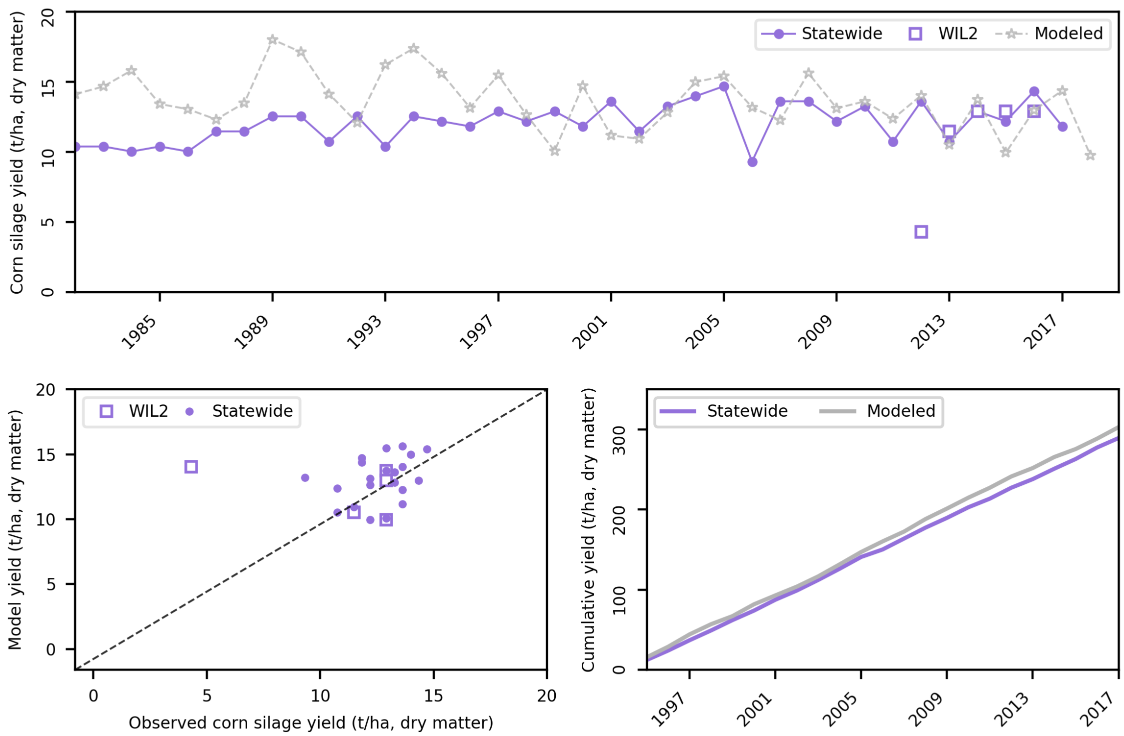

Figure 3.29: Forage yield in the WIL2 calibrated model. Compare to the baseline version in Figure 2.33.

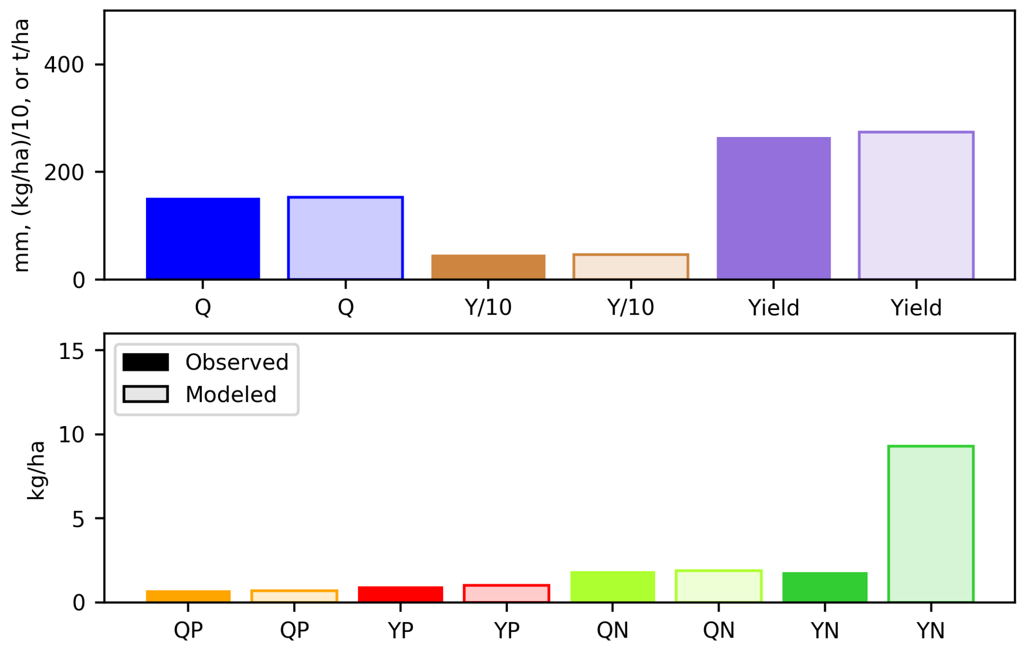

Figure 3.30: Total observed and modeled runoff, erosion, etc. for the duration of the APME project at WIL2 (Q=runoff, Y/10=sediment divided by 10, QP/N= runoff P/N, YP/N=sediment P/N). Compare to the baseline version in Figure 2.34.

3.5 Discussion

For PAW1, the calibration process led to low values of PBIAS for all of the model outputs except QN. That is, it was usually possible to obtain a model that is not biased towards high or low values and accurately reproduces the mean/total output. The same was true for WIL2 for everything but YN.

It was much more difficult to attain high values of NSE, which would indicate that the model does a good job of matching the pattern of the runoff events and sediment/nutrient losses. Aside from yield (for reasons discussed in §3.3.2) and QN, NSE > 0 was achieved for all outputs in the PAW1 model, meaning that the model is a better predictor than the mean of the observed data. At WIL2, NSE >0 was obtained only for runoff, sediment, and YP. The WIL2 sediment and Y components had very high NSE values (NSE=0.78 and NSE=0.82, respectively). For PAW1 and the other WIL2 outputs, though, NSE fell somewhat short of the NSE>0.5 that Moriasi et al. (2007) and Wang et al. (2012) suggest characterize an acceptable model.

Improving the model fits for runoff, for example, would have required parameter adjustments that resulted in major changes to the relative magnitudes of the runoff events. However, there seemed to be little scope for doing this. The parameters that were tested raised or lowered the runoff overall while making only small changes in the pattern of runoff. As runoff is a driving factor in sediment and nutrient losses, this affects how well the model can predict those quantities as well.

At the same time, relatively few parameter changes were necessary to achieve the PBIAS and NSE values in Tables 3.6 and 3.8 for runoff, sediment, and yield. For runoff at PAW1 it was sufficient to adjust two parameters that alter the daily soil moisture index calculation, while WIL2 required a different method of calculating the daily runoff. Especially at PAW1, there was little difference between the PMs for many of the model variants tested during runoff calibration (Figure 3.2), suggesting that many parameter combinations are capable of performing equally well. Both models were calibrated for erosion by adjusting the erosion control practice factor, PEC (albeit in different directions), and for yield by using PARM72 to reduce the NH_3 volatilization rate.

Nutrient calibration required changes to many more parameters and was ultimately only partially successful. Calibrating the crop yield at PAW1 by reducing volatilization and N stress had significant effects on soil N, P, and organic matter that worsened the model statistics for nutrient losses. Some of those effects were rectified by changes related to organic matter cycling and mineral P behavior. However, N cycling remained problematic, with the model appearing to direct too much N towards leaching, volatilization, and denitrification, and too little into runoff.

Because of the difficulty of calibrating the PAW1 model for nutrient losses, a less extensive search was carried out for the WIL2 model. Satisfactory (NSE>0) results at WIL2 were obtained only for YP. It may be possible to attain better results for both models through a wider search of parameter space, a task that would be aided by automatic calibration tools that can handle comparison data on arbitrary timescales.

Both before and after calibration, the PAW1 model performed quite well in the high-runoff 2012 – 2013 monitoring period, but generally overestimated runoff and erosion in 2014 – 2015. Annual or seasonal variations in the performance of the WIL2 model are less clear, especially in the case of sediment and nutrients for which few measurements are available. However, there may be a tendency for the model to predict too much runoff in spring/summer events and too little in autumn/winter.

Improving the calibration statistics for PAW1 also came at the expense of predictions for particularly large events. The calibrated model tends to underestimate the magnitude of large sediment and nutrient loss events, and also sometimes overestimates small ones. This tendency is less noticeable in the WIL2 model.

Differential APEX performance during modeling periods has been reported in the literature. Wang et al. (2014) calibrated APEX for streamflow and sediment in a large Chinese watershed using data from relatively wet conditions and found that it performed poorly in the drier validation period. Anomaa Senaviratne et al. (2013) found that APEX underestimated sediment losses for small-to-medium rainfall events in their corn-soybean rotations on claypan soils, while Nearing (1998) show that in general soil erosion models tend to over-predict losses for small events and under-predict for large ones. Given the variety of reported results, it is not surprising that the APEX models for Vermont do not perform consistently over time. A literature review of how APEX performs in different circumstances, and why, would be useful guidance for model users.

The temporal pattern in the PAW1 model suggests that using it to calculate and compare runoff etc. between dry and wet climates will result in underestimating the magnitude of potential changes. It may also suggest that simulations of runoff etc. in even wetter climates could work quite well – provided they do not generate too many large events. However, this work has not investigated the mechanism underlying this behavior, which could be to do with total precipitation, relative timing of precipitation events, or something else entirely. This means that the circumstances in which the model will give accurate results are not well understood.

In fact, the errors in these models in general are not well quantified. The PBIAS and NSE statistics show how well the models perform compared to others in the literature, but they do not provide confidence intervals on any of the model outputs. Deriving such information is well beyond the scope of this thesis. The calibrated models will simply be regarded as useful for making predictions for runoff, sediment, crop yield, and YP, YN, QP (PAW1) and YP (WIL2) in a range of climates in Ch. 4.

4 Farming in Vermont's Changing Climate

One of the aims of this thesis is to simulate agricultural outcomes – runoff, erosion, nutrient losses, crop yields, and delays to farm operations – in a selection of hypothetical climates. At this point, models for two farm sites have been calibrated using data obtained between 2012 and 2015/2017. The actual operations on those farms can now be simulated many times over, each time using a new daily weather data set generated from statistics that describe the underlying climate conditions. Running many simulations allows properties such as the mean, 95%-ile, and standard deviation of the outcomes to be characterized. It is then possible to ask, for example, how variable crop yields would be under the proposed climates, or whether formerly extreme erosion events would become routine.

The next steps, then, are to (1) select a set of specific climate scenarios and derive the statistics that describe them, (2) run many repeats of the calibrated PAW1 and WIL2 models using sets of daily weather generated from those climate statistics, and (3) analyze the distributions of outcomes obtained for each climate scenario. This chapter details all of those steps, beginning with the considerations involved in selecting hypothetical climates, and how the selection of scenarios is affected by how APEX represents climate information.

4.1 The Climate Scenarios

4.1.1 Choosing Climate Scenarios

For this thesis, hypothetical future climates will be defined relative to the actual climate as observed in recent years. This raises two questions. First, which recent time period should be used as the baseline from which the new climates are developed? Second, should the proposed scenarios be extrapolations of recent climate trends, or based on model projections, or generated using other criteria such as likely relevance to the agricultural community? Regardless of which option is chosen, it would be useful to put the final scenarios into the context of recent and projected changes.

These issues are complicated by a number of factors. First, as explained in Ch. 1, the climate in the northeastern US appears to have undergone especially pronounced changes since approximately the turn of the century. To recap: Hoerling et al. (2016) note that 10 of the 12 years since 2002 in their data set had experienced unusually high rainfall from extreme events, Frei et al. (2015) find that increases in warm season heavy precipitation have been especially marked since 2000, and Huang et al. (2017) detect changepoints in extreme precipitation in 1996 and 2002. On one hand, a single baseline beginning before and ending after ~2000 could be chosen to represent a recent "average" climate. On the other hand, the change point represents an opportunity: periods prior to and since ~2000 could potentially be used as a natural laboratory to explore the effects of the climatic changes that have already occurred.

The second issue is that climate models do not appear to capture this recent behavior. For example, Guilbert et al. (2014) find that in the Lake Champlain basin region, "the 0.95 quantile of daily precipitation is projected to increase by 8.9% by midcentury", relative to a 1961 –2000 baseline. However, Huang et al. (2017) observe that "averaged over the Northeast, extreme precipitation from 1996 to 2014 was 53% higher than from 1901 to 1995". Creating a climate scenario by increasing the 95%-ile of precipitation by 9% above pre-2000 levels may result in less extreme precipitation than has actually been experienced recently. Conversely, given that $\lesssim$20 years have elapsed since the apparent step change in precipitation, and that it may be a short-term phenomenon that is not primarily due to anthropogenic climate change (Hoerling et al., 2016), attempting to extrapolate from recent data may result in extreme scenarios that are unlikely to occur.

Even a simple quantitative comparison of proposed scenarios to historical and model climates turns out not to be straightforward, because different papers use different time periods and statistical measures to quantify the climate. It is beyond the scope of this thesis to convert all the published statistics to a consistent scale and compare the trends and projections. As Moriasi et al. (2007, 2015) recommend for hydrological modeling, it would be helpful if the authors of climate assessments could report a standard set of statistics along with whichever numbers are of particular interest to their own studies.

Given these considerations, this thesis examines a small set of climates that are qualitatively consistent with observed and expected trends and/or may be of particular significance for agriculture. All of the future scenarios include warmer temperatures. Precipitation either increases year-round, shifts from summer to spring, or becomes more intense. The specifics of the scenarios, and in particular the selection of the historical baseline, are partly determined by how APEX represents climate information, so this is discussed next.

4.1.2 Selecting a Baseline Scenario

APEX can handle weather and climate inputs in two different ways. First, it can directly use files containing daily weather data supplied by the user. This was how APEX was run in Chapters 2 and 3. Second, it can use a file containing monthly statistics for temperature, precipitation, etc. to generate its own daily weather. The monthly statistics can be supplied directly by the user or they can be calculated from daily weather files using one of two supporting programs, Weather Import (for WinAPEX) or WXPM[1]. In principle, constructing climate scenarios simply means feeding Weather Import or WXPM daily weather data for a baseline climate period, obtaining monthly statistics, and editing the statistics as necessary to obtain climates with the desired characteristics.

In practice, the way in which weather data are represented by the monthly statistics ("WP1") files constrains the climates that can be meaningfully simulated. The WP1 files contain mean monthly numbers of rainy days that are converted into a set of probabilities that govern when precipitation occurs. They also contain parameters describing a skewed normal distribution of event sizes. APEX generates weather by first deciding whether rain occurs on a given day, and if so, generating an event from the skewed normal distribution (Williams et al., 2012).

[1] https://epicapex.tamu.edu/model-executables/

Specifically, if the probabilities of a wet day after a wet day ($P(W|W)$) and a wet day after a dry day ($P(W|D)$) are available, the code generates a random number, and if that number is smaller than the appropriate probability, precipitation occurs. If $P(W|W)$ and $P(W|D)$ are not available, they are estimated from the monthly number of rainy days:

$$PW = NWD/ND \tag{4.1}$$

$$P\left(W|D\right) = 0.75 * PW \tag{4.2}$$

$$P\left(W|W\right) = 1 - 0.75 + P\left(W|D\right) \tag{4.3}$$

Where ND is the number of days in the month and ND the number of wet days. Event sizes are generated as follows :

$$RF = XLV \times RST2_{Mo} + RST1_{Mo} \tag{4.4}$$

$$XLV = \left(X1^3 - 1\right) \times 2/RST3_{Mo} \tag{4.5}$$

$$X1 = (SND - R6) \times R6 + 1 \tag{4.6}$$

$$R6 = RST3_{Mo}/6 \tag{4.7}$$

Where RST1, RST2, RST3 are the monthly storm mean, standard deviation, and skew coefficient, and SND is the standard normal deviate.

The extent to which this distribution fits the input weather data governs the accuracy with which the data can be represented by the parameters in the file, and the degree to which the properties of the weather generated from those statistics match those of the original data set. Also, any climate scenario created by editing the WP1 parameters must be capable of being represented by the parameters of the skewed normal distribution.

To investigate how these issues may affect the modeling in this thesis, WP1 files were generated from Rutland and Burlington weather station data using Weather Import. Then, Python code was created to mimic the process that APEX uses to generate daily weather from the WP1 files, using Equations 4.1 – 4.7 above. For each WP1 file, 1000 years' worth of events were generated. The cumulative distributions of the real and generated events were then obtained, along with a set of simple statistics describing the distribution of events.

Figure 4.1 shows actual and generated cumulative distributions for different time periods for the Burlington weather station. Statistics for both Rutland and Burlington are given in Table 4.1. The left-hand panel in the figure shows the actual distribution of events for 1970 - 1999, and the distribution of the events generated from the WP1 file output by Weather Import. The generated distribution clearly contains fewer small events, and the median simulated rainfall is about 5.0 mm compared to the actual value of 2.5 mm. The 95%-ile and 99%-ile are also about 3 mm higher than observed, although the total precipitation is approximately correct. The Weather Import statistics include a much smaller number of rainy days per month than is actually present in the 1970 – 1999 data, which – together with the correct total rainfall – presumably is the cause of the higher median precipitation.

Table 4.1: Comparison of measured and generated precipitation statistics for various periods for the Rutland (for PAW1) and Burlington Airport (for WIL2) weather stations. Days with zero precipitation are excluded from the calculations. The climate models adopted as the basis for generating hypothetical future climates are shown in **bold**.

Period	Data[a]	50%-ile mm	95%-ile mm	99%-ile mm	Total mm	Days w/ precip.	p[b]
Rutland							
1970 – 1999	Historical	4.3	25.7	42.4	948	127	0.15
1970 – 1999	Generated, Weather Import	5.2	26.9	44.9	1001	118	–
1970 – 1999	Generated, corrected	4.5	26.0	44.3	997	128	–
2000 – 2018	Historical	3.8	27.3	43.2	1064	143	0.15
2000 – 2018	Generated, corrected	4.3	26.7	47.2	1134	143	–
1980 – 2008	Historical	4.1	26.9	44.7	1007	133	–
1980 – 2008	**Generated, corrected**	**4.1**	**26.4**	**47.1**	**1032**	**133**	**–**
Burlington							
1970 – 1999	Historical	2.5	21.6	37.4	911	159	0.02
1970 – 1999	Generated, Weather Import	5.0	24.4	40.8	896	114	–
1970 – 1999	Generated, corrected	3.3	21.5	38.6	981	160	–
2000 – 2018	Historical	2.8	24.1	40.9	976	156	0.02
2000 – 2018	Generated, corrected	3.8	23.7	40.8	1087	157	–
1980 – 2008	Historical	2.5	22.4	40.7	926	156	–
1980 – 2008	**Generated, corrected**	**3.2**	**21.9**	**40.3**	**976**	**156**	**–**

[a] Historical = calculated from weather station data. Generated, Weather Import = generated using monthly statistics for the historical data output by the Weather Import program. Generated, corrected = generated using Weather Import monthly statistics adjusted to better reflect the observed percentiles. See text for full details.
[b] K-S p value for the comparison between data for the 1970 – 1999 and 2000 – 2018 periods

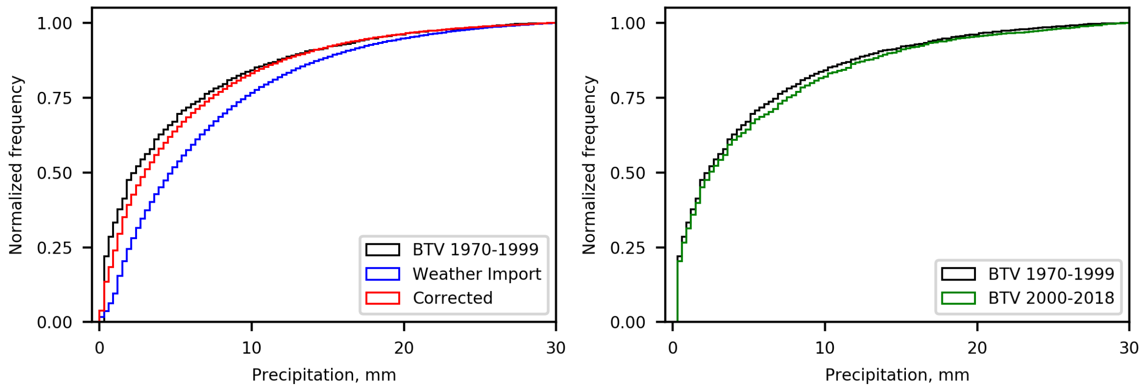

Figure 4.1: Cumulative distributions of historical and generated daily weather for two time periods at the Burlington Airport weather station. The distribution based on Weather Import monthly statistics can be improved upon (left), but the inaccuracy in the improved distribution is still comparable to the difference between the real distributions for the 1970 – 1999 and 2000 – 2018 periods (right).

An attempt was made to "correct" the WP1 file by setting the number of rainy days and monthly totals to their observed values, setting the wet-wet and wet-dry probabilities to zero (i.e., unknown), and adjusting the skew coefficient. The result of this is shown in the left-hand panel of Figure 4.1 (and in Table 4.1). The median has decreased from 5.0 to 3.3 mm, and the 95%-ile, 99%-ile are also closer to their observed values. However, the total precipitation is now too high by 8%.

Whether these inaccuracies are important depends on how the generated events will be used. As noted in the previous section, it could be interesting to ask how agricultural outcomes in the post-∼2000 climate, where a significant rise in intense precipitation appears to have occurred, may compare to those experienced prior to 2000. Statistics for the Rutland and Burlington weather stations for 1970 – 1999 and 2000 – 2018 are in broad agreement with the region-wide trends described in Ch. 1: at both sites, total precipitation is higher in the more recent period, and the 95th and 99th percentiles have also risen (Table 4.1). To be clear, the difference between the distributions of precipitation data in the two time periods is only significant at the Burlington weather station (at the $p<0.05$ level). Nonetheless, the fact that their properties are consistent with those derived from more comprehensive and sophisticated analyses suggests that it could be worth simulating agricultural outcomes from climates with these characteristics[2].

The actual distribution for the 2000 – 2018 period at Burlington is shown in the right-hand panel of Figure 4.1 (and statistics given Table 4.1). The difference between the 1970 – 1999 and 2000 – 2018 precipitation distributions at Burlington – which a K-S test finds to be significant at $p = 0.02$ – is small. In fact, it is comparable to the difference between the distributions of actual and generated events for 1970 – 1999 at that site, even after the initial Weather Import statistics have been adjusted to better represent the data. This suggests that it is not useful to separately model pre-and post-2000 climates. Instead, a single period, 1990 – 2008, will be used as the baseline climate scenario.

4.1.3 Summary of Scenarios

The properties of the climates that were ultimately selected are illustrated in Figures 4.2 and 4.3 and listed in Table 4.2. The scenarios are:

[2]We of course have records of the actual daily weather for 1970 – 1999 and 2000 – 2018, so it would be possible to simply model those periods using historical data and measure (to the model's degree of accuracy) the actual distribution of agricultural outcomes that took place. However, in practice it would be preferable to characterize the climate statistically, generate many realizations of those climates, and run APEX for each one of them. This would allow a more accurate quantification of extreme (rare) outcomes. Also, if the model does not predict absolute quantities well but can predict relative outcomes more accurately, comparing model outcomes produced in an identical manner is preferred.

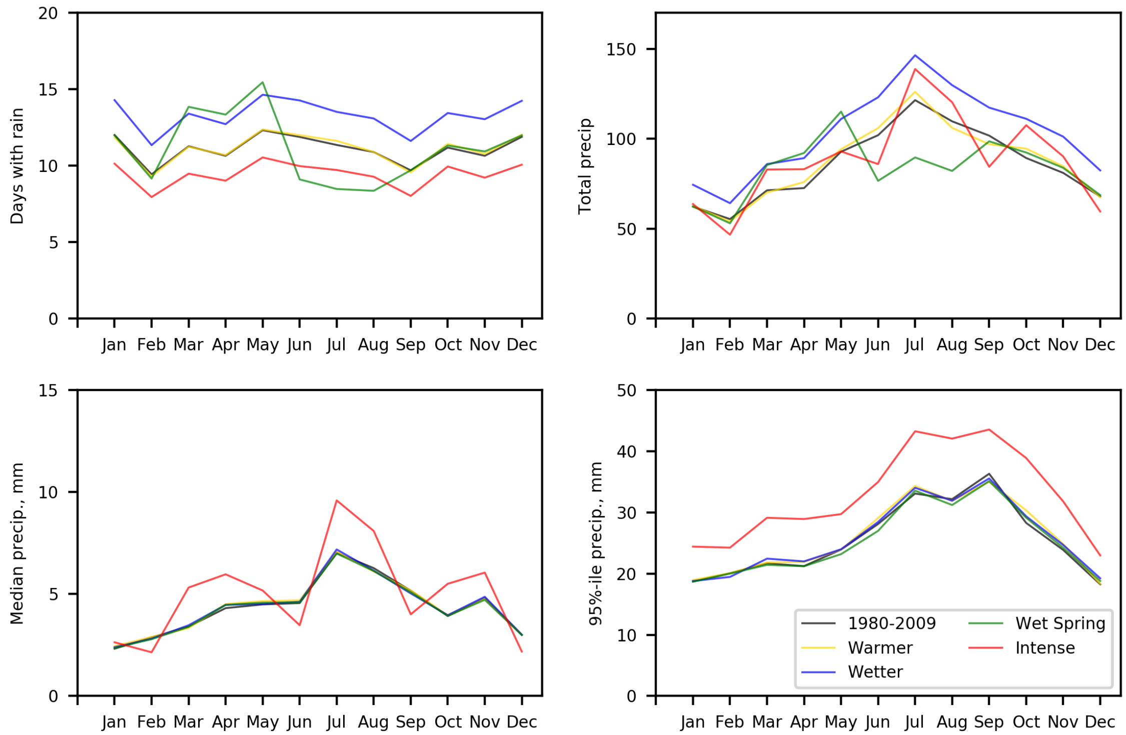

Figure 4.2: Some statistical properties of the hypothetical climate scenarios, Rutland weather station (for the PAW1 site).

1980 – 2009: All of the hypothetical future scenarios are defined relative to this baseline period. The monthly statistics that define this climate are the "corrected" Weather Import numbers described in the previous section.

Warmer: Each month's maximum and minimum temperature is increased by 2°C relative to the 1980 – 2009 climate. This is similar to the temperature rise of 2 – 3° C between ∼1970 – 1999 and the mid-21st century found in the model projections of Hayhoe et al. (2007) and Guilbert et al. (2014). (Note that the kind of analysis presented in §4.1.2 was not done for temperature, and it was simply assumed that APEX's monthly statistics can provide an adequate representation of a 2° C difference.)

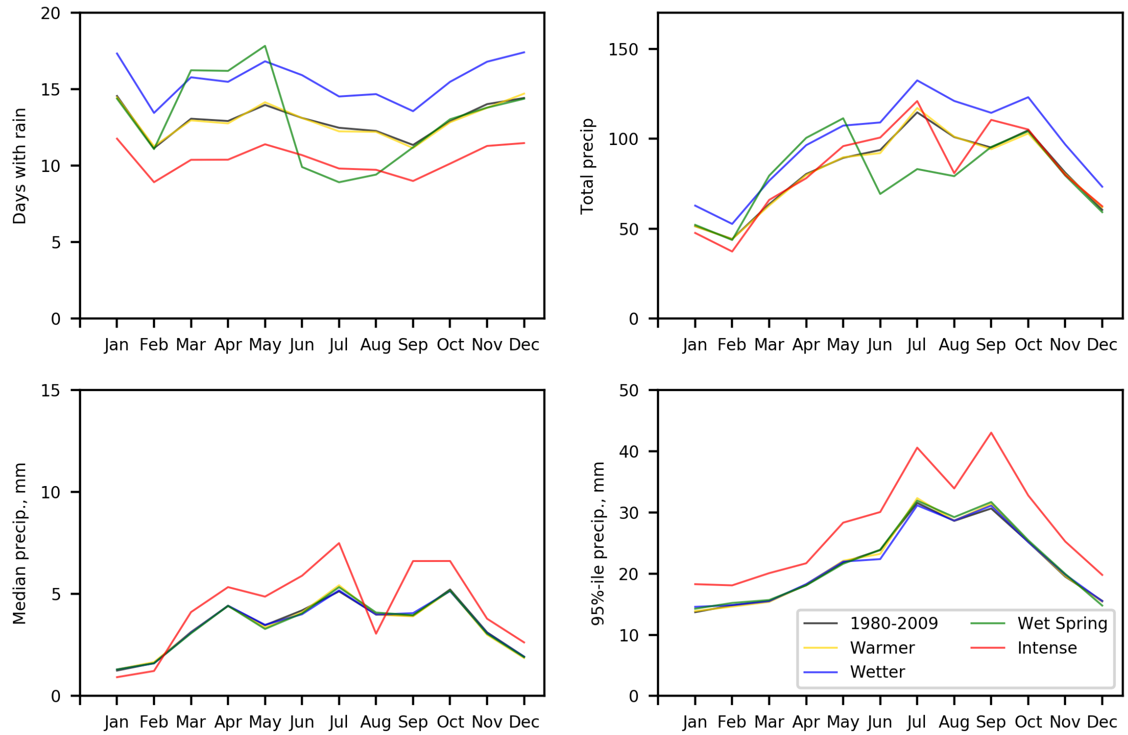

Figure 4.3: Some statistical properties of the hypothetical climate scenarios, Burlington weather station (for the WIL2 site)

All of the subsequent scenarios use the temperatures of the "Warmer" scenario.

Wetter: Total precipitation is raised by 20%. This is achieved by increasing the monthly number of rainy days by 20%. This only affects the total precipitation; the median and 95%-ile are unchanged. Annual increases in total precipitation of ∼5-8% by mid-century were modeled by Hayhoe et al. (2007), but total precipitation already appears to have risen by 13% since 2002 (Huang et al., 2017).

Wet Spring: 25% of the precipitation in June, July, and August is added to March, April and May. This is achieved by decreasing/increasing the monthly number of rainy days by 25%. This changes the total annual precipitation whilst the median and 95%-ile remain approximately the same. This scenario is of interest because farm operations are particularly sensitive to wet spring conditions (§1.2). Also, Hayhoe et al. (2007) find that climate models expect an increase in cool season precipitation (although they only discuss December – February) possibly combined with a slight decrease in warm-season rainfall.

Intense Rain: In this scenario the 95%-ile of precipitation rises by 30%. This is achieved by increasing the monthly storm standard deviation by a factor of 1.25 (WIL2) or 1.28 (PAW1). Increasing the standard deviation also results in higher total and median precipitation. To isolate the effect of intense precipitation, the total for this scenario was "reset" by decreasing the number of days on which precipitation occurs by a factor of 0.8 (WIL2) or 0.85 (PAW1). Median precipitation remains higher and has more month-to-month variability than in 1980 – 2009.

Table 4.2: Annual precipitation properties of the climate scenarios. Days with zero precipitation have been excluded from the calculations.

Scenario	Description	50%-ile mm	95%-ile mm	99%-ile mm	Total mm	Days w/ precip.
PAW1						
1980 – 2008	Local weather station[a]	4.1	26.4	47.1	1032	133
Warmer[b]	1980 – 2009 plus T+2°C, year-round	4.1	26.1	47.2	1024	133
Wetter	"Warmer" plus total precip. + 20%	4.1	26.1	46.2	1227	160
Wet Spring	"Warmer" plus 25% of Jun/Jul/Aug precip. added to Mar/Apr/May	4.0	25.5	45.5	1007	133
Intense Rain	"Warmer" plus 95%-ile precip. + 30%	4.7	33.9	59.9	1059	113
WIL2						
1980 – 2008	Local weather station[a]	3.2	22.0	40.3	976	156
Warmer[b]	1980 – 2009 plus T+2°C, year-round	3.2	21.9	39.9	975	156
Wetter	"Warmer" plus total precip. + 20%	3.2	21.7	39.8	1166	187
Wet Spring	"Warmer" plus 25% of Jun/Jul/Aug precip. added to Mar/Apr/May	3.2	21.4	38.2	959	157
Intense Rain	"Warmer" plus 95%-ile precip. + 30%	4.1	28.4	50.9	982	125

[a] Rutland for the PAW1 watershed, Burlington Airport for the WIL2 watershed; see §2.2.2

[b] "Warmer" uses the same monthly precipitation statistics as "1980 – 2009". The small differences in the descriptive statistics result from random number generator seeds not being fixed in the Python code used to recreate APEX's daily weather generation method (§4.1.2).

4.2 APEX Simulation Procedure

APEX was run 998 times for each of the scenarios listed in §4.1.3. The APEX setup was exactly that derived for the final, calibrated model at each site (§3.3.5, 3.4.1), except that NGN was set to -1 (telling APEX to generate daily weather) and for each run the random number generator seed (IGN parameter) was incremented so that a unique set of daily weather files was generated from the monthly statistics files. The WP1 files were generated as described in the previous sections. Annual runoff, sediment, and nutrient loss values were collected for each run, along with annual crop yields and stresses. These distributions are presented and interpreted in §4.3.1 and 4.3.2.

In addition, a separate set of runs was made to investigate whether the climate scenarios might affect the scheduling of farm operations. This was achieved by reducing the value of PARM78 from 10 to 1 and then repeating the runs described above. PARM78 sets the ratio of soil water:field capacity above which operations will be delayed, so a value of 10 effectively means that operations will never be postponed. Farmers do not commonly base operations decisions on a specific, measured value of soil water, but some extension publications recommend avoiding working in the fields when soil moisture is roughly equal to field capacity (Al-Khaisi and Licht, 2005), i.e., PARM78=1.

Unlike runoff, crop yield, etc., APEX does not summarize the number of days by which an operation is delayed in annual output files with a regular format. Instead, that information is printed as part of the program's main log (.OUT) file in free text format. Whenever an operation does not occur on the scheduled date, preceding lines

```
  YCW  0.2141E+01  0.1523E+02  0.1831E+02  0.9872E+01  0.1585E+02  0.4137E+02  0.1431E+03  0.0000E+00  0.1750E+02  0.3314E+02  0.2012E+02  0.0000E+00
  0.3166E+03   YCW
1
    APEX1501  2018  12  23  18:42:32
          RUN #    1
          Winapex.SIT
          9 PAW1_BASEWeathersheld
          11/2/2018 4:44:24 PM


          SA(# ID)   Y M D OPERATION
         ATMOS CO2 =  400.
        1       1 2005 512  IDON=   1  HRD#=  0  VT-PLUS   RATE=  35870.kg/ha DPTH=    0.00mm ELEM WT(kg/ha)-- MN=    54. NH3=    54. ON=    57. MP=
   14. OP=     4. MK=    75. HUSC= 0.14
        1       1 2005 512  (Added)   DPTH =      0. mm  HUSC =   0.14
        1       1 2005 512  (Added)   DPTH =    150. mm  HUSC =   0.14
        1       1 2005 529  IDON=   1  HRD#=  0  27-9-18   RATE=    224.kg/ha DPTH=    0.04mm ELEM WT(kg/ha)-- MN=    60. NH3=    60. ON=     0. MP=
    9. OP=     0. MK=    34. HUSC= 0.21
        1       1 2005 529  (Added)   DPTH =     40. mm  HUSC =   0.21
        1       1 2005 531  PCUS  RSD =   1.2t
        1       1 2005 531  (Added)   DPTH =     40. mm  HUSC =   0.22
        1       1 2005   6   7 PCUS  GERMINATION--0.2 M SW =      1. mm  HU =    55.c  HUSC =  0.28
     ^^^^^       1       1  24   9  27 PDSW =    0.76122E+01 PDAW =    0.74225E+01
        1       1 2005   9  28  (Added)   PCUS  YLD=   9.72t/ha  YSD=   0.04t/ha  AGPM=  11.37t/ha  STL=  11.37t/ha  RWT=   3.80t/ha  HI=  0.900  NCN
=  0.013G/G  HUSC= 0.75  YLN= 122.kg/ha  YLP=  26.kg/ha
        1       1 2005 928  (Added)   DPTH =    -50. mm  HUSC =   0.75
        1       1 2005 928  KILL      DPTH =      0. mm  HUSC =   0.07
1
    APEX1501  2018  12  23  18:42:32
          RUN #    1
```

Figure 4.4: Screenshot of part of an APEX .OUT file. The highlighted string "PDSW" indicates that the subsequent operation did not occur on Sep 27th because of excessive soil water. The following line indicates that the harvest operation was carried out on Sep 28th. Searching for instances of "PCUS YLD=" preceded by "PDSW" allow the days of delay to be counted for each year in the simulation.

in the file print the value of "PDSW", the plow depth soil water, for the days on which PARM78 was exceeded and the operation could not be executed (Figure 4.4). Python code was therefore written to parse this output by counting sequential occurrences of PDSW and linking them to a regular expression identifying the subsequent delayed operation. The results of this investigation are described in §4.3.3.

.
Table 4.3: Median and 95th percentile of annual runoff and erosion, and totals over the indicated time period, in the five climate scenarios described in §4.1.3

Scenario	Runoff, mm			Sediment, t ha^{-1}		
	50%-ile	95%-ile	Total	50%-ile	95%-ile	Total
PAW1 2012 – 2015						
1980 – 2008	309	519	1282	1.61	3.36	7.05
Warmer	299	505	1246	1.44	3.08	6.35
Wetter	304	515	1262	1.50	3.14	6.58
Wet Spring	305	514	1259	1.42	3.06	6.29
Intense Rain	338	595	1425	1.76	4.01	7.98
WIL2 2012 – 2017						
1980 – 2008	31.4	85.5	224	0.050	0.200	0.424
Warmer	30.3	84.8	218	0.051	0.200	0.429
Wetter	32.5	88.0	231	0.056	0.212	0.462
Wet Spring	29.2	83.8	213	0.048	0.194	0.408
Intense Rain	42.7	132	315	0.079	0.333	0.668

4.3 Results

4.3.1 Runoff, Erosion, and Nutrient Losses

Distributions of annual runoff, sediment, and nutrient loss values for PAW1 and WIL2 for the five climate scenarios are shown in Figures 4.5 – 4.9. Some statistical properties are given in Tables 4.3 and 4.4. These statistics relate to the years 2012 – 2015 for PAW1 and 2012 – 2017 at WIL2, the years for which we have actual management records and can ask “what might have happened on these farms in a different climate?”.

Table 4.4: Median and 95th percentile of annual nutrient losses[a], and total losses over the indicated time period, in the five climate scenarios described in §4.1.3

Scenario	Runoff P, kg ha^{-1}			Sediment P, kg ha^{-1}			Sediment N, kg ha^{-1}		
	50%-ile	95%-ile	Total	50%-ile	95%-ile	Total	50%-ile	95%-ile	Total
PAW1 2012 – 2015									
1980 – 2008	0.88	1.86	3.91	1.51	3.28	6.67	4.67	9.73	20.5
Warmer	0.90	1.80	3.93	1.44	3.08	6.34	4.76	9.68	20.7
Wetter	0.97	1.94	4.23	1.51	3.16	6.63	4.77	9.59	20.8
Wet Spring	0.81	1.69	3.60	1.41	2.98	6.23	4.67	9.61	20.5
Intense Rain	0.99	2.12	4.46	1.74	3.89	7.80	4.76	10.3	21.2
WIL2 2012 – 2017									
1980 – 2008	–	–	–	0.129	0.470	1.04	–	–	–
Warmer	–	–	–	0.136	0.469	1.06	–	–	–
Wetter	–	–	–	0.153	0.491	1.15	–	–	–
Wet Spring	–	–	–	0.129	0.449	1.01	–	–	–
Intense Rain	–	–	–	0.198	0.719	1.56	–	–	–

[a] Runoff N is omitted because neither the PAW1 model nor the WIL2 model was capable of satisfactorily simulating that quantity (§3.5). The WIL2 model only gave reasonable results for sediment-bound P.

For runoff and sediment, a similar pattern holds at both sites: little change relative to 1980 – 2009 except in the case of Intense Rain. In Warmer, Wetter, and Wet Spring, median runoff decreases by $\leq$3% at PAW1, and changes by $\leq$ 7% at WIL2. The lower runoff in the Warmer scenarios presumably results from increased evapotranspiration/lower soil moisture. Runoff is also marginally lower in the Wetter and Wet Spring scenarios at PAW1 and in the Wet Spring scenario at WIL2. This can probably be explained by the slightly higher crop yield in those cases (§4.3.2). The small changes in sediment loss in Warmer, Wetter, and Wet Spring are always in the same direction as runoff.

During the setup and calibration process it was found that the PAW1 model reproduced events in a high-runoff year reasonably well while overestimating runoff in a low-runoff year. It is therefore possible that the model overestimates runoff in the 1980 – 2009 scenario relative to the higher-rainfall scenarios, at least at PAW1. The "real" effect of higher temperatures and total rainfall could actually be an increase in runoff. Nonetheless, in a warmer climate, increasing total rainfall or shifting rain from summer into spring appears to have relatively minor effects on runoff.

The intensity of precipitation appears to be a more important factor than the total amount. In the Intense Rain scenario, runoff and sediment increase at both farms. Median runoff and soil loss increase by 9% at PAW1, and by 36% and 58% respectively at WIL2 (albeit from very low initial values at WIL2).

As far as nutrients are concerned, sediment-bound P behaves in a similar manner to runoff and sediment: little change except for the Intense Rain scenario where median P loss increases somewhat (15% at PAW1, 53% at WIL2). Runoff P could only be simulated reasonably at the PAW1 site. Again, changes are fairly small. The

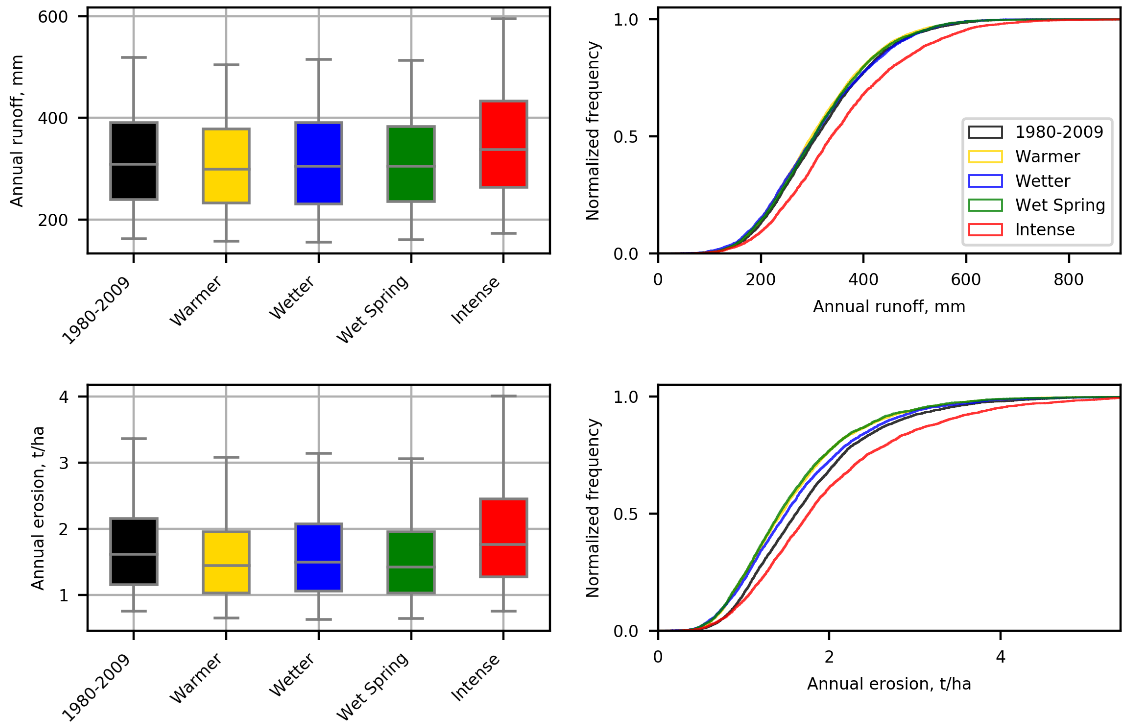

Figure 4.5: Distributions of annual runoff (upper panels) and sediment (lower panels) at the PAW1 site for the five climate scenarios described in §4.1.3. Box plots (left-hand panels) are used to indicate the median (horizontal line), first and third quartiles (box limits) and 5th and 95th percentiles (whiskers). The right-hand panels show cumulative frequency distributions for the same outcomes and scenarios. Only the years 2012 – 2015 are included in the PAW1 figures in this chapter, as the aim is to simulate actual, recorded farm operations in a range of climates.

largest shifts are increases of 10% from 1980 – 2009 to Wetter, and 13% to Intense Rain.

The behavior of N in sediment (which, again, could only be simulated at PAW1), does not follow the same clear pattern. Median sediment N changes by only 2% between scenarios. The main (but still small) difference is that the 95%-ile in Intense Rain is higher than in the other scenarios.

More generally, the 95%-ile of outcomes increases by more than the median in the

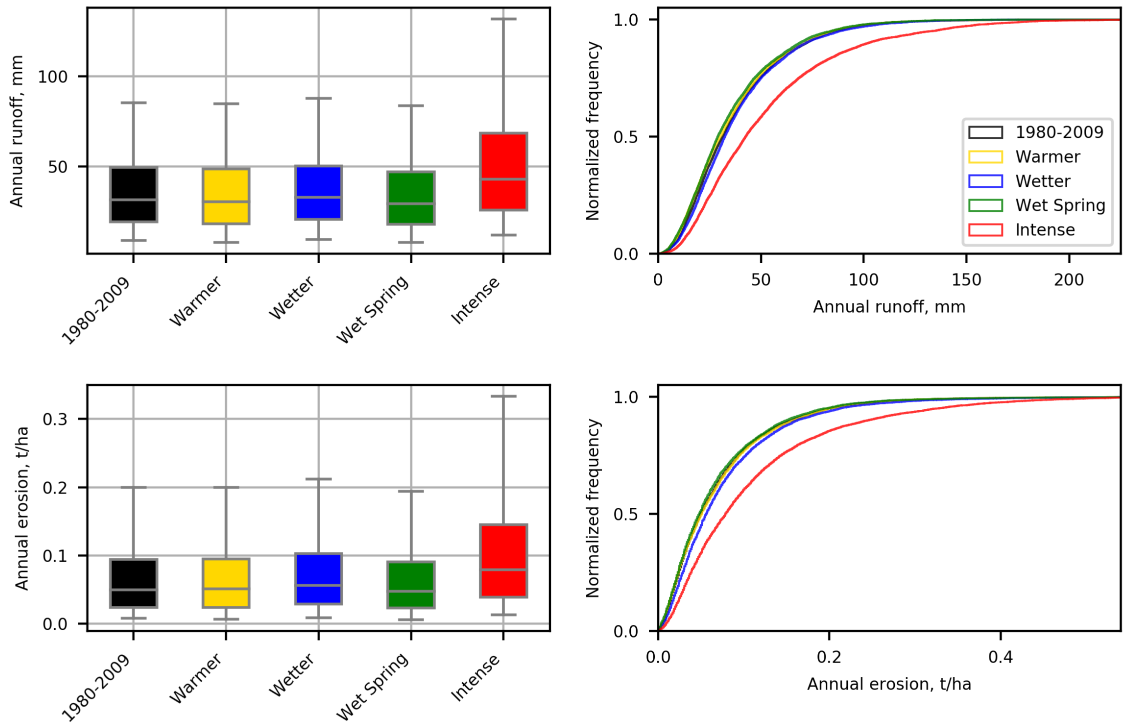

Figure 4.6: Distributions of annual runoff (upper panels) and sediment (lower panels) at the WIL2 site for the five climate scenarios described in §4.1.3. Box plots (left-hand panels) are used to indicate the median (horizontal line), first and third quartiles (box limits) and 5th and 95th percentiles (whiskers). The right-hand panels show cumulative frequency distributions for the same outcomes and scenarios. Only the years 2012 – 2017 are included in the PAW1 figures in this chapter, as the aim is to simulate actual, recorded farm operations in a range of climates.

Intense Rain scenario. For sediment at PAW1, for example, the median rises by 9% but the 95%-ile increases by 19%. Because the distribution of erosion events is skewed in this way (Figure 4.5), the total soil loss also increases by more than the median, by 13%. The PAW1 calibrated model tended to underestimate the magnitude of large sediment and nutrient loss events (§3.3.5), so it is possible that these numbers should be even higher. Higher total losses, and a higher fraction of large events, could have implications for the effectiveness of management practices designed to reduce

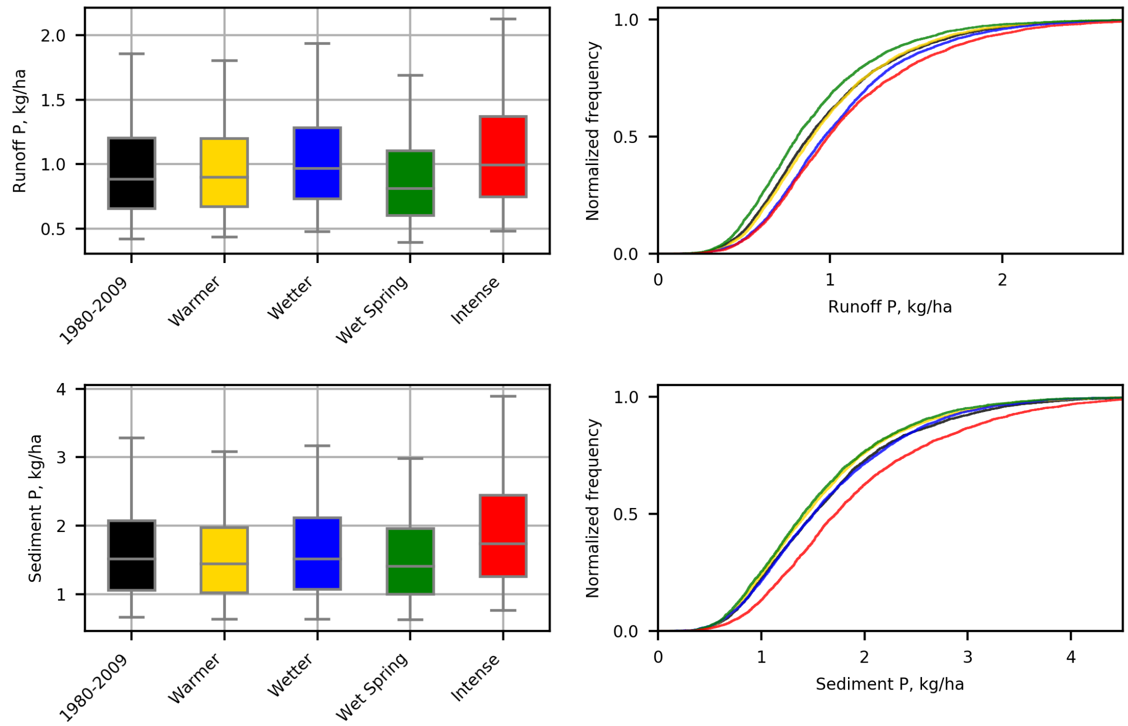

Figure 4.7: Distributions of annual runoff P and sediment P losses at the PAW1 site for the five climate scenarios described in §4.1.3.

sediment and nutrient losses. This is revisited in §4.4.

4.3.2 Crop Yields

Annual crop yields in the various climate scenarios are shown in Figures 4.10 and 4.11, along with annual days of water (i.e. drought) and temperature stress (WS, TS; all other stresses are negligible in all scenarios). The median and standard deviation of the yields are given in Table 4.5. The standard deviation was selected because it may be a useful indicator of the predictability of crop yields in future climates (ignoring other factors like pest damage), and could potentially be translated into

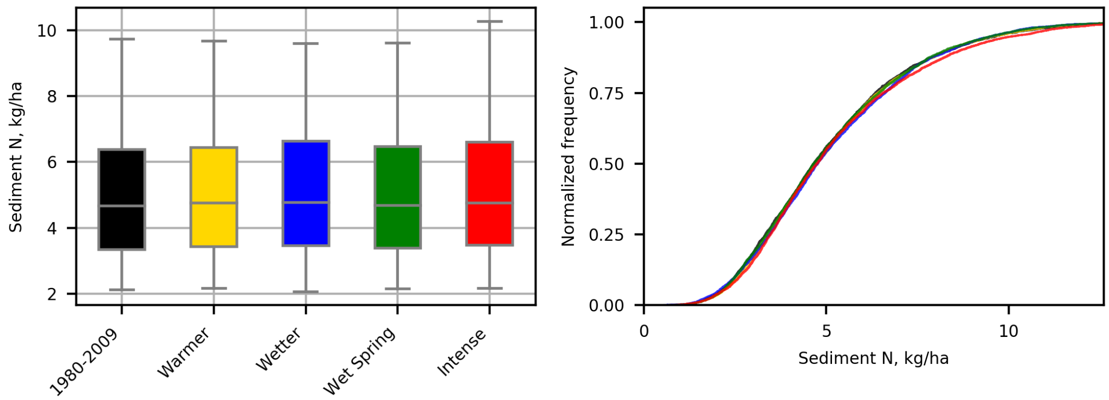

Figure 4.8: Distributions of annual sediment N loss at the PAW1 site for the five climate scenarios described in §4.1.3. (The PAW1 model was not able to produce acceptable results for runoff N.)

the likelihood of achieving an economically viable yield.

The way in which forage yields change between the climate scenarios is quite different from the changes in runoff etc. At PAW1 the median yield is lowest in the 1980 – 2009 scenario, rises in the Warmer scenario, decreases marginally as rainfall increases in Wetter, then rises slightly again in Wet Spring before dropping slightly again in Intense Rain. In contrast, at WIL2 the highest median yield occurs in 1980 – 2009. The relative yields in the other scenarios are qualitatively the same as at PAW1, but none of the hypothetical climates produces as much forage as 1980 – 2009. Overall, though, the yield changes are quite small; <10% over all scenarios at both sites.

The pattern of yields is not clearly related to patterns of crop stress. APEX reports the number of days each year when each source of stress is the one that regulated (decreased) crop growth, and these numbers change as temperature and rainfall are altered. At PAW1, increasing temperatures in the Warmer scenario decreases days of temperature stress. Days of water stress increase, but total stress days decrease.

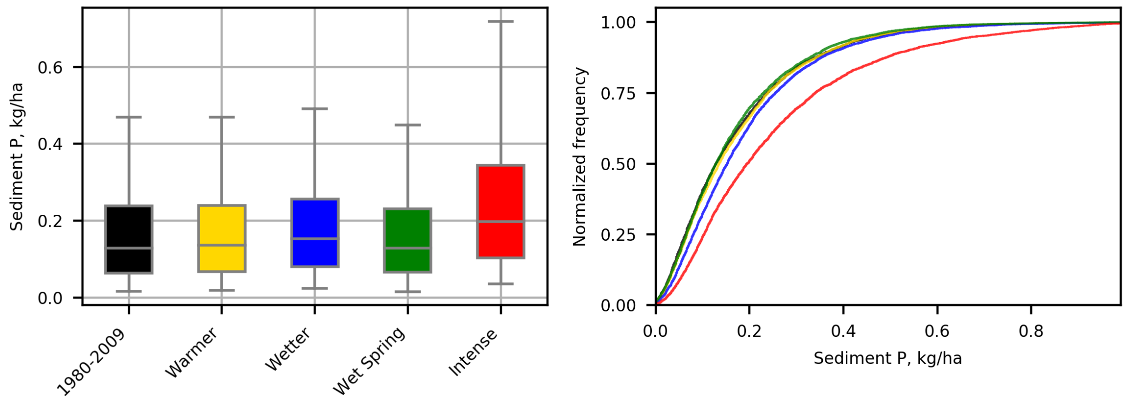

Figure 4.9: Distributions of annual sediment P loss at the WIL2 site for the five climate scenarios described in §4.1.3. (The WIL2 model was not able to produce acceptable results for runoff P or for N.)

This, combined with the higher heat unit uptake that presumably also occurs, may be the reason for the generally higher yields in the future, warmer climates.

At WIL2, however, increased temperatures contribute to a larger reduction in total stress days, but yields are also reduced. There are other, smaller oddities as well. At PAW1, WS increases slightly in Wetter even though precipitation occurs on more days, and TS decreases despite the fact that APEX simulates lower temperatures on rainy days (Williams et al., 2012). Days of water stress are fewer in Wet Spring despite 25% of the June – August precipitation having been shifted to March – May. These counterintuitive results may simply arise from the kind of complex, nonlinear interactions between weather, soil processes, and crop growth that require the use of a hydrological model in the first place.

The increase in temperature and changes in precipitation modeled in this Chapter appear to have a modest, positive effect on median silage corn yields at PAW1, but slightly negative effects at WIL2. Regardless of whether yields increase or decrease, variability increases somewhat at both sites in all scenarios relative to 1980 – 2009,

Table 4.5: Forage yield, water stress, and temperature stress for the five climate scenarios described in §4.1.3.

Scenario	Yield		WS	TS	WS+TS
	50%-ile	Std. dev	50%-ile	50%-ile	50%-ile
	t ha^{-1}	t ha^{-1}	days	days	days
PAW1					
1980 – 2008	12.2	1.81	9.22	29.6	38.9
Warmer	13.0	2.14	16.7	18.2	34.8
Wetter	12.7	2.21	18.2	16.0	34.9
Wet Spring	13.4	2.12	15.9	18.5	34.4
Intense Rain	13.2	2.31	17.7	17.6	35.3
WIL2					
1980 – 2008	13.7	2.00	9.47	36.1	45.5
Warmer	12.8	2.17	14.9	12.8	27.7
Wetter	12.5	2.19	16.4	11.2	27.6
Wet Spring	13.3	2.11	12.6	14.0	26.7
Intense Rain	12.9	2.28	15.37	13.1	28.5

and more markedly at PAW1. However, the magnitude of the variability is small compared to the ∼factor-of-3 variations in yield observed during the few-year duration of the APME project (Figure 2.24), so it is probably not very relevant for planning purposes.

4.3.3 Farm Operations

Results from the simulations in which soil water content can delay field operations are given in Table 4.6. The spring manure applications and the planting operations at PAW1 are rarely delayed; fewer than 1% of operations do not happen on time. Any delays that do occur are small, ≤2 days. This is true for all climate scenarios.

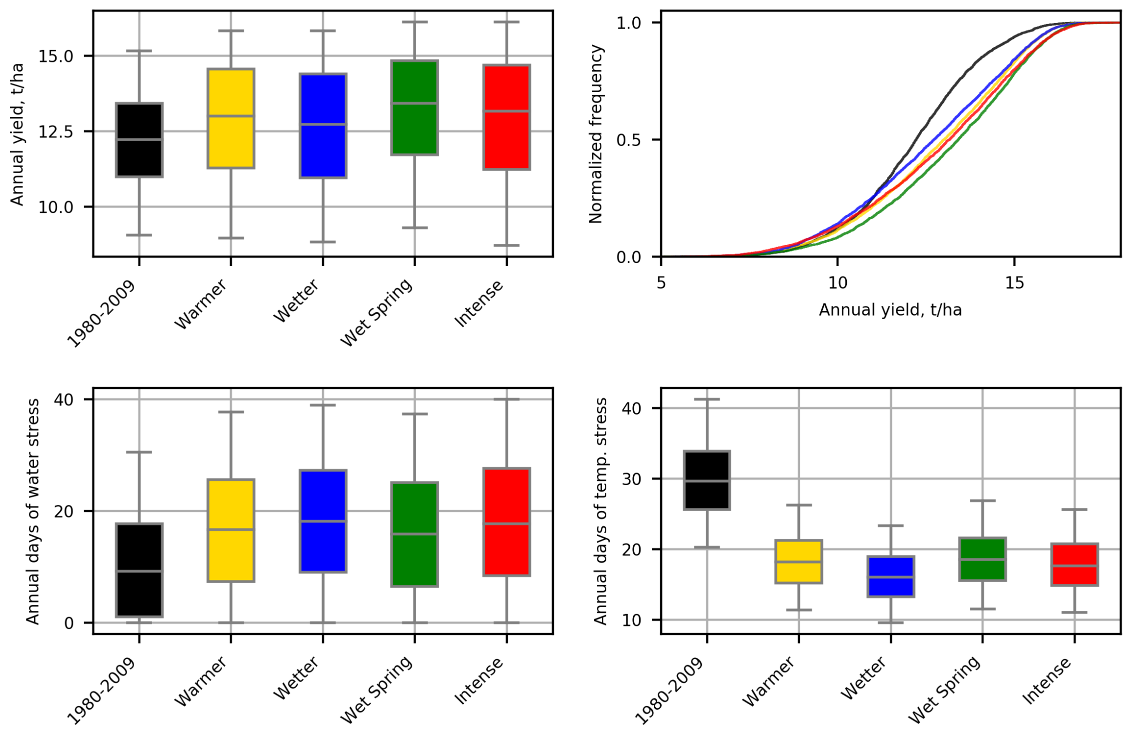

Figure 4.10: Upper panels: Percentiles and cumulative distribution functions for forage yield at the PAW1 site for the five climate scenarios. Lower panels: Water stress (left) and temperature stress (right).

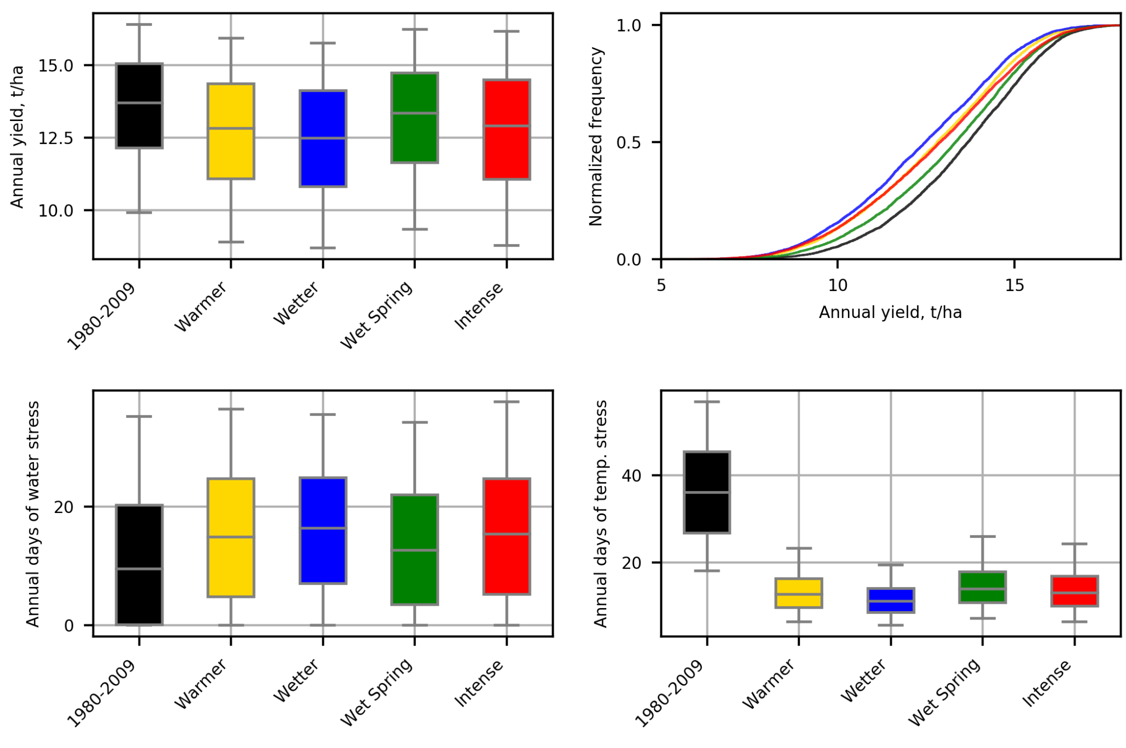

Figure 4.11: As for Figure 4.10 but for the WIL2 site.

Table 4.6: Statistics describing delays to farm operations. All years are included in these statistics, including the run-up period.

Scenario	Spring manure[a]			Planting			Harvest		
	% delayed[b]	Mean[c]	Max[d]	% delayed	Mean	Max	% delayed	Mean	Max
PAW1									
1980 – 2008	0.09	1.03	2	0.10	1.03	2	13.9	1.92	13
Warmer	0.06	1.00	1	0.10	1.05	2	14.5	1.83	14
Wetter	0.05	1.00	1	0.06	1.00	1	14.0	1.57	10
Wet Spring	0.04	1.00	1	0.05	1.00	1	14.3	1.82	14
Intense Rain	0.10	1.00	1	0.11	1.07	2	12.4	1.63	12
WIL2									
1980 – 2008	0.00	–	–	0.00	–	–	31.9	3.04	22
Warmer	0.00	–	–	0.00	–	–	27.8	2.36	22
Wetter	0.00	–	–	0.00	–	–	23.5	1.92	22
Wet Spring	0.00	–	–	0.00	–	–	29.2	2.41	22
Intense Rain	0.00	–	–	0.00	–	–	24.4	2.38	22

[a] Fall manure at WIL2 is not shown, but no delays were found in any scenario.
[b] Fraction of runs in which the operation was behind schedule.
[c] Mean of the non-zero delays, in days. The median of the non-zero delays is 1.0 for all scenarios at both sites.
[d] Maximum delay, in days.

Only the harvest operation turns out to be delayed for any significant fraction of the simulation years. This operation occurs late about 14% of the time at PAW1 and 23 – 32% at WIL2 in all scenarios, and can be delayed by as much as 14 days at PAW1 and 22 days at WIL2. The proportion and magnitude of delays at PAW1 are similar across all climates, while there is a little more variation at WIL2. These delays do not affect the statistics for runoff, sediment, nutrients, or yields.

If delays essentially occur only in the autumn, APEX must consistently estimate higher soil water values at that time of year. Figure 4.12 shows daily soil moisture for 2013 at PAW1, from the setup/calibration simulations that used that year's actual daily weather as input. In 2013 Braun et al. (2016) describe the weather as being dry in mid-March to mid-May, very wet in late May – early July, and average or drier than usual through the rest of the summer and fall. In those conditions, APEX soil moisture fell rapidly during the spring, stayed relatively low but with large excursions as the crop took up water and large rain events occurred in June and July, then began to rise again in ∼September.

If this temporal pattern of soil water is representative, harvest events in September – October will be more prone to delays than manure spreading and planting operations in ∼May. The larger fraction and extent of delays at WIL2 compared to PAW1 is probably due to the later dates at which harvest took place at that site; as late as November 9th in 2012.

This behavior seems contrary to experience in the field. Quantitative data regarding delayed farm operations are scarce, but anecdotally, delays are usually associated with wet spring weather and planting operations. Vermont extension personnel have reported that "most farmers planted late [in 2017]... [2013 had] one of the wettest

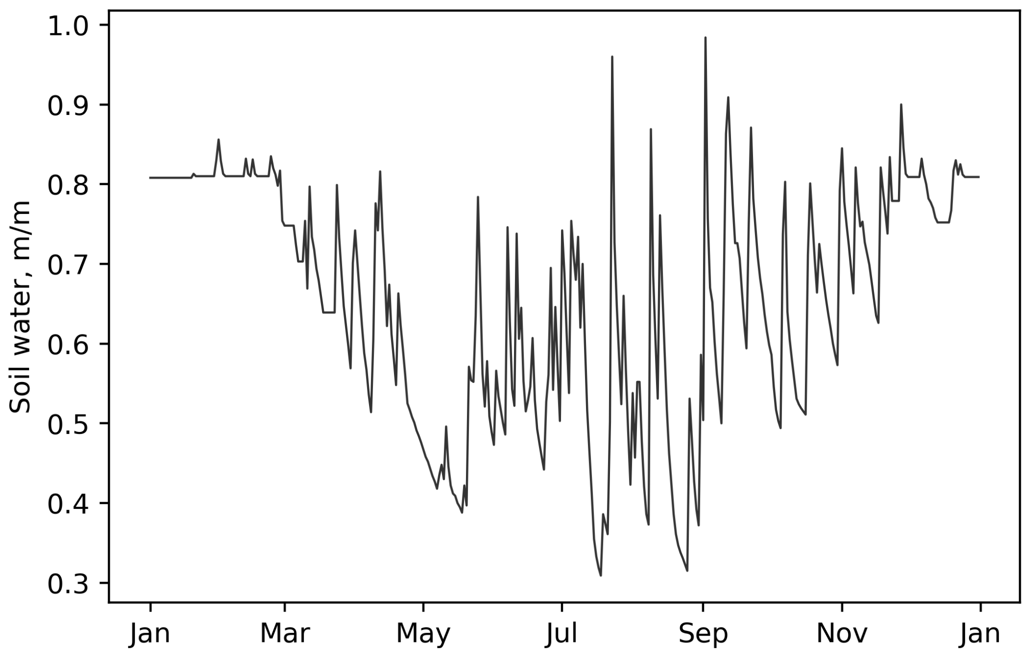

Figure 4.12: Model soil water for the PAW1 watershed in 2013, using actual daily weather.

Junes on record, so no doubt that a lot of farmers lost their initial plantings and had to replant or had to plant late" (J. Faulkner, personal communication, 2018). There is only one comment on weather-related delays at Pawlet and Williston during the APME project (Braun et al., 2016), and that is that "wetness delayed planting of the PAW2 field" in 2014.

The reasons why these APEX results appear to differ greatly from farmers' experience are not understood by this author at this time. One contributing factor could be that APEX delays planting operations until the soil temperature reaches the crop's base temperature (§2.2.4; the user has no control over this), so if low temperatures and wet weather tend to occur together, this could mask delays caused by wet soils.

Overall, though, it is not clear whether high soil moisture in the fall is a fundamental characteristic of APEX models in Vermont's precipitation regime or whether there are parameters that influence soil water that should be changed during the setup and calibration process. This may be a useful avenue for further investigation.

4.4 Discussion

It sounds straightforward to design realistic climate scenarios based on historical data and model projections, obtain their monthly temperature and precipitation statistics, and then use APEX to simulate agricultural outcomes in those climates. In practice, several complications are encountered. First, the climate is not well-behaved. Gradual trends turn into abrupt changes and the models are uncertain and inconsistent. Second, the mass of statistical measures used to describe historical, recent, and projected regimes makes it difficult to compare the published climate studies in a consistent way. Also, the climates that can be simulated by APEX are constrained to be those that can be adequately described by a skewed normal distribution[3].

Once these issues have been taken care of, the mechanics of simulating agricultural outcomes in a range of climates are fairly simple. In most cases, differences turn out to be small. In the Warmer, Wetter, and Wet Spring scenarios, median runoff, sediment, and nutrient loss changed by $< 10\%$. These changes could be in either direction, and differed slightly between sites, but the general picture in these scenarios is that

[3]Actually this isn't quite true. APEX also gives the option of using a modified exponential distribution. Brief experiments with this option did not produce reasonable rainfall, but it may be worth further investigation. Also, the user could produce their own set of daily weather files generated from an arbitrary distribution and feed these to APEX in lieu of the program generating its own daily weather.

outcomes do not change substantially. It should be noted that when the PAW1 model was compared against edge-of-field data in Chapters 2 and 3, it tended to overestimate runoff in low-runoff years relative to high-runoff years. This could imply somewhat worse relative outcomes in the wetter climates than the model predicts, but this possible bias cannot be quantified on the basis of the work done for this thesis.

Larger changes were found for the Intense Rain scenario. At the PAW1 site, where runoff etc. are already at relatively high levels, median runoff, sediment, and nutrient losses increase by $\approx 9 - 15\%$. At WIL2, the changes are much more pronounced: median runoff increases by 36%, erosion by 56%, and sediment-bound P by 53% (other nutrient loss pathways could not be simulated at WIL2). Fortunately, these increases are relative to low baseline values.

In the Intense Rain scenario, the 95%-ile of outcomes increases by a larger factor than the median. This means that total runoff and sediment/nutrient losses also rise by more than the median. At PAW1 the increases in median, 95%-ile, and total sediment losses are 9%, 19% and 13%, respectively. Higher total losses, and a larger fraction occurring in large events, could have several implications for farms and their surrounding environments.

First, severe erosion events can pose a threat to crops. In the APME study, Braun et al. (2016) comment that "Several large runoff events in the spring of 2013 caused substantial soil erosion of the PAW1 field... A thick layer of sediment was deposited... at the lower end of the field, smothering young corn plants". The farmer reported that erosion and crop yield that year were the worst he had ever experienced.

Second, management practices that aim to reduce sediment and nutrient losses could become less effective. BMPs generally work by some combination of encouraging soil aggregation and water infiltration, reducing the impact energy of raindrops and the velocity at which rainwater moves across the land, and filtering out sediment and nutrients. It is possible that there exist thresholds, either in terms of long-term total loads or short-term pressures, above which these practices "saturate". Indeed, some modeling work finds that BMPs become "more necessary but less effective" under climate change (Bosch et al., 2014). The Intense Rain scenario probably bears the most relation to recent, observed changes in the climate in the US Northeast, but it is not clear whether these trends will continue, strengthen, or weaken in the future.

Crop yields change in different ways at PAW1 and WIL2. At PAW1, all of the climate scenarios have a mildly positive effect on median yields, which increase by up to 10%. At least some of this increase is likely due to lower temperature stress and higher heat unit uptake. At WIL2, yields decrease relative to 1980 − 2009 levels by similar amounts, and the reason for this is unclear. Model yields become more variable in all scenarios, but at a level that is dwarfed by the normal year-to-year variations seen on real farms.

This behavior seems contrary to many published reports, which have warned that temperature stress, drought stress, and pressures from weeds, insects and diseases will cause field crop yields to decrease in this region (Frumhoff et al., 2007; Galford, 2014; Horton et al., 2014; Markowitz, 2017; Tobin et al., 2015). In one of the studies underlying these reports, for example, Wolfe et al. (2008) derive temperature and precipitation data from downscaled global circulation models for two emissions scenarios, and use that as input to a model that calculates infiltration, evapotranspiration, etc. They point out that more frequent summer droughts will be occurring at the same time as crop water requirements are increasing because of warmer temperatures, and find that all crops will be negatively affected by water deficits.

There are a few possible reason for this apparent discrepancy. One is that the APEX models in this thesis and/or the published models could be generating inaccurate results. To some extent, this is very likely true: model results should generally be regarded as exploratory and indicative, rather than as firm predictions (Oreskes et al., 1994).

Also, this thesis has modeled two particular sites in Vermont using future climates that are based on fairly conservative changes to recent conditions at those locations. These hypothetical, location-specific climates may be quite different from those based on region-wide projections from downscaled GCMs. In particular, the most pronounced temperature and drought effects occur under higher emissions scenarios towards the end of the century (Guilbert et al., 2014; Hayhoe et al., 2007; Wolfe et al., 2008). The climate scenarios modeled here do not include the much higher temperatures of those scenarios. The crop model in APEX is also fairly simple, and does not include effects such as the shortened grain-filling period expected in a warmer climate (Wolfe et al., 2008).

The APEX models used in this section were calibrated to reproduce observed runoff, erosion, nutrient losses, and crop yields. They were also used to investigate the possibility that a wetter future climate will cause losses to farmers by delaying spring farm operations. At least in the configuration used in this study, the models calculate that soil moisture will delay autumn harvest operations but not spring planting and manure spreading. This seems contrary to real-world experiences. Further work is needed to understand the underlying cause of this behavior, and whether parameters can be found that lead to more realistic outcomes.

5 Conclusions

Dairy farming is in many ways at the heart of Vermont. It sustains rural communities, creates delicious food, and plays a big role in shaping the much-loved landscape of the state. At the same time, the industry has also contributed to environmental problems, and – along with dairying in much of the rest of the country – finds itself in a deeply troubled economic state.

Climate change threatens to make the current situation worse. Increasingly intense storms could sweep more soil, phosphorus, and nitrogen into the state's waterways. A wetter start to the year could keep farmers from planting their corn on time, and crops may suffer from hot days and summer droughts. These issues have been investigated in general, region-wide modeling studies, but many gaps in our knowledge remain. Better predictions of how climate change will affect agricultural outcomes, including delays to farm operations, should help farms adapt to a warmer, wetter world.

This thesis has started to fill in those gaps. The first task was to investigate whether the APEX model could adequately simulate runoff, sediment, nutrient losses, and crop yields on two local farms, comparing model outputs with data obtained on the farms themselves. The second was to carry out a pilot study of outcomes, including weather delays, in a handful of hypothetical climates.

The scope of this study has been narrow but deep. Partly, this is because of the need to critically evaluate the models. Silberstein (2006) asked, "If hydrological models are so good, do we still need data?". According to several recent papers, the answer is "yes". Baffaut et al. (2017) found that APEX could not be relied upon to give satisfactory results without being calibrated using dedicated water quality data. In other cases, even calibrated models could not adequately reproduce sediment or water quality measurements (Ramirez-Avila et al., 2017), or performed poorly when used to model management that was not used in the calibration data set (Bhandari et al., 2016).

Calibration is important, but to the best of this author's knowledge no standard procedure or worked examples exist. APEX is extensively documented in terms of its underlying equations and principles (Williams et al., 2012), and its input and output files and variables (Steglich et al., 2016). General calibration guidelines exist in the literature (Daggupati et al., 2015; Wang et al., 2012), some supporting tools are also available (e.g. Wang et al., 2014; Wang and Jeong, 2016), and modeling workshops are offered from time to time. However, papers based on APEX often do not fully document their procedures and choice of model parameters. Some have argued that the hydrology/water quality modeling community is only beginning to grasp issues of transparency and reproducibility in research (Saraswat et al., 2015).

For those reasons, the APEX models in this thesis were created and calibrated using a rich set of data collected by a project that aimed to quantify the effects of best management practices on local dairy farms (the APME project; Braun et al., 2016). A calibration procedure was developed that paid close attention to some of the many processes occurring in the model, attempting to understand the causes of problems and find opportunities to improve the model. Extensive use was made of the graphical and statistical performance indicators recommended by Moriasi et al. (2007, 2015).

The setup and calibration steps, described in Chapters 2 and 3, are most likely imperfect, incomplete, and will not apply to all modeling applications. However, they might be a useful starting point for others who are new to modeling, either as a case study that can illuminate some of the considerations involved in model calibration, or as a living document that can be improved upon by the community. The APEX input and output files for the calibrated models, and some supporting materials, are made available as described in Appendix 1.

The performance of the uncalibrated, "baseline" models was not good. For most of the outputs the Nash-Sutcliffe Efficiency (NSE) was <0, meaning that the mean of the data was a better predictor than the model. Runoff is generally the easiest quantity for a model to get right, and the total runoff summed over all events was within 25% of the observed value at PAW1. However, it was overestimated by a factor of three in the WIL2 model.

Calibration improved the performance measures significantly. NSE for runoff and sediment increased from 0.41 and -0.05 to 0.47 and 0.22, respectively, at PAW1; and from -11.7 and 0.68 to 0.39 and 0.78 at WIL2. The PBIAS measure for all outputs was within ±10%, indicating that the models reproduced the observed total crop yield, runoff, sediment, and nutrient losses to within a small margin.

Still, the calibrated models fell short of what is considered a satisfactory model in the scheme of Moriasi et al. (2007, 2015). The PAW1 model tended to predict too much runoff etc. in low-runoff years compared to high-runoff years, and also systematically underestimated large events. These tendencies were not apparent in the WIL2 model, although the smaller number of comparison events made that model more difficult to evaluate. In addition, NSE values for some of the nutrient outputs (runoff N at PAW1 and all but sediment P at WIL2) remained below zero. Model predictions for those nutrient loss pathways were excluded from the remainder of the study.

The calibrated models were used in Chapter 4 to simulate runoff, erosion, and some nutrient losses on the two farms in a small set of hypothetical future climates. Selecting these climate scenarios was surprisingly difficult. Ideally, the way in which hypothetical climates relate to recent observations and climate model predictions would be clear, but climate assessment papers do not always report the same statistical measures, baseline time periods, and other quantities of interest. Nevertheless, four scenarios were constructed in which monthly temperatures were increased by 2° C ("Warmer"), total precipitation increased by 20% ("Wetter"), 25% of precipitation shifted from June, July, and August to March, April, and May ("Wet Spring"), and the 95%-ile of precipitation increased by 30% ("Intense Rain").

In the Warmer, Wetter, and Wet Spring scenarios, there was little change to runoff, sediment, and nutrient losses. Only the Intense Rain scenario was different. In that scenario, median runoff, sediment, and nutrients changed by 2 – 15% at PAW1 and 36 – 53% at WIL2 (albeit from very low baseline values at WIL2). More importantly, the 95%-ile of all environmental outcomes increased by more than the median, as did total runoff, sediment, and nutrient losses. Having more total sediment and nutrient loss, and more of it happening in large events, could pose problems for both farm operations and field practices that aim to reduce environmental problems.

Chapter 4 also found that changes in crop yields are small, ≤10%. These findings are rather muted compared to predictions in recent assessments of agriculture and climate change (Frumhoff et al., 2007; Galford, 2014; Horton et al., 2014; Markowitz, 2017; Tobin et al., 2015). Such reports warn that hot, dry summers will cause field crop yields to fall, but any signature of that in the APEX simulations was minor. In fact, yields at PAW1 actually increased slightly relative to the recent climate baseline.

The reasons for this are probably related to the specifics of the climates that were modeled. The scenarios presented here do not include the 4–5° C temperature increase expected by the end of the century in some emissions pathways, and those are the conditions in which the largest yield declines are seen in other simulations.

As well as runoff, water quality, and yields, APEX can in principle also estimate how often and for how long farm operations will be delayed due to wet soils. This was also investigated in Chapter 4. Contrary to real-world experiences, only the fall silage harvesting operation ever experienced significant delays. The reasons for this are presumably to do with how APEX calculates soil moisture, but it is not clear at this point whether and how this could be improved.

5.1 Limitations

For clarity, the limitations of this study are highlighted below:

- The model was calibrated using a limited set of field data with unquantified, and possibly large, errors and uncertainties. In addition, some model parameters and outputs could be only approximately mapped to the field measurements.
- Because of limitations in the available automatic calibration software, the cal-

ibration and some of the sensitivity analysis were done manually. Automating these activities would have allowed a wider search of the model parameter space, possibly leading to the identification of better-performing models.

- Partly due to the limited number of runoff events in the calibration data set, the models were calibrated using all the available data for each site. No data were held out for model validation.
- The final model performance statistics were usually lower than those that have been suggested in the literature as indicating satisfactory model performance (although those criteria should depend on the purpose of the modeling, are not well-established for APEX, and are probably affected by publication biases). Some outputs (e.g. some nutrient losses) could not be reliably simulated at all, and delays to farm operations were not well reproduced.
- The model results could not be associated with error bars or confidence intervals.
- The range of farming systems and climate scenarios simulated in this pilot study was small.
- Further work is needed to elucidate how the results in this thesis compare to those of other studies that may have examined different climates and included different processes (e.g. pest damage) in their simulations.

5.2 Future Work

Although the present work is limited in scope, it suggests numerous possibilities for future research. Obtaining reliable results from hydrological models is important for a number of reasons, from climate change adaptation to the possibility of reducing pollution through nutrient credit trading between farms. Evaluating the accuracy of model outputs, or the range of outcomes that could be obtained by reasonable modelers, would therefore be of interest.

Even though automatic calibration tools are now available to aid with obtaining reproducible results in a systematic manner, there are still many subjective decisions to be made when setting up a model. Many manually-calibrated models also exist in the literature. It would be an interesting exercise to assign the same modeling task to two or more people, requiring them to work independently, and see whether they obtained consistent results. A similar test could be done using software to find all parameter combinations that produced acceptable model results, then use the distribution of model outcomes to assign confidence intervals to the results (see Moriasi et al., 2016, for related work).

The calibrated PAW1 and WIL2 models could be used to simulate outcomes in a wider range of climates. If the full range of historical, recent, and projected climates were quantified in a consistent manner, the full parameter space could be explored. A grid like this could identify sudden changes in outcomes, or climates with especially negative or beneficial effects. Calibrating models for the hay farms in the APME study would allow climate effects on those systems to be evaluated, in addition to the corn-growing farms modeled in this thesis. The work could also perhaps be extended to rotational grazing systems using data from Gilker (2005).

The APME data could also be used to help construct models of BMPs in future climates. As noted above, BMP efficacy may be impaired when more runoff etc. comes from large events, and there are indications that models should be calibrated using data obtained under the same management that they will be used to model (Bhandari et al., 2016). This would probably not be a straightforward task, though. The period in which BMP data were gathered in the APME project was short compared to the control period, so calibration data will be more limited, and the project itself concluded that the effects of the BMPs were often the opposite of what was expected (Braun et al., 2016).

The present work was not successful in using APEX to obtain predictions for how farm operations may be delayed by wet soils in future, wetter climates. Resolving this issue would require digging into the mechanics of how APEX calculates soil moisture, and probably writing dedicated software to identify and optimize influential parameters. Current APEX auto-calibration tools are not equipped to handle operations delays as a calibration variable.

5.3 Final Thoughts

Times are hard for Vermont's dairy farms. The price paid for milk cannot cover the cost of production, and another round of sell-offs and farm consolidation appears to be underway (D'Ambrosio, 2018). Even the historically less vulnerable organic sector, and the newer, niche "grass-milk" producers are struggling.

Many observers connect the current problems to structural issues within the dairy industry and in agriculture in general, including the perennial problem of ever-higher production leading to ever-lower prices. To this, climate change adds many challenges, and perhaps some opportunities for those who have the capacity to adapt. It is hoped that the work in this thesis provides a foundation for more extensive work that can warn of dangers and highlight possibilities in our changing world. In the meantime, to the farms that are struggling to get through this difficult period and (hopefully) on to the next high point in the cycle, I wish the very best of luck.

Bibliography

Al-Khaisi, M. and Licht, M. (2005). Soil moisture conditions - consideration for soil compaction. https://crops.extension.iastate.edu/soil-moisture-conditions-consideration-soil-compaction.

Alley, M. M., Martz, M. E., Davis, P. H., and Hammons, J. L. (2009). Nitrogen and Phosphorous Fertilization of Corn. https://pubs.ext.vt.edu/424/424-027/424-027.html.

Anomaa Senaviratne, G. M. M. M., Udawatta, R. P., Baffaut, C., and Anderson, S. H. (2013). Agricultural Policy Environmental eXtender Simulation of Three Adjacent Row-Crop Watersheds in the Claypan Region. *J. Environ. Qual.*, 42(3):726–36. http://www.ncbi.nlm.nih.gov/pubmed/23673939.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998). Large area hydrologic modeling and assessment - Part 1: Model development. *J. Am. Water Resour. Assoc.*, 34(1):73–89.

Baffaut, C., Nelson, N. O., Lory, J. A., Senaviratne, G. A., Bhandari, A. B., Udawatta, R. P., Sweeney, D. W., Helmers, M. J., Van Liew, M. W., Mallarino, A. P., and Wortmann, C. S. (2017). Multisite Evaluation of APEX for Water Quality: I. Best Professional Judgment Parameterization. *J. Environ. Qual.*, 46(6):1323. https://dl.sciencesocieties.org/publications/jeq/abstracts/46/6/1323.

Baker, J. L., David, M. B., Lemke, D. W., and Stewardship, L. (2006). IMPORTANCE OF HYDROLOGY IN DETERMINING LOSSES, AND POTENTIAL IMPLICATIONS ON MANAGEMENT SYSTEMS TO REDUCE THOSE LOSSES. Technical report.

Beckage, B., Osborne, B., Gavin, D. G., Pucko, C., Siccama, T., and Perkins, T. (2008). A rapid upward shift of a forest ecotone during 40 years of warming in the Green Mountains of Vermont. *Proc. Natl. Acad. Sci.*, 105(11):4197–4202.

Beegle, D. and Peters, J. (2011). Common Manure Test Results Conversions. http://articles.extension.org/pages/18998/common-manure-test-results-conversions.

Betts, A. K. (2011). Vermont Climate Change Indicators. *Weather. Clim. Soc.*, 3(2):106–115.

Betts, A. K. (2017). Climate Change in Vermont. Technical report. papers2://publication/uuid/046F0732-055C-4137-982D-F5A18A8E941C.

Bhandari, A. B., Nelson, N. O., Sweeney, D. W., Baffaut, C., Lory, J. A., Senaviratne, A., Pierzynski, G. M., Janssen, K. A., and Barnes, P. L. (2016). Calibration of the APEX Model to Simulate Management Practice Effects on Runoff, Sediment, and Phosphorus Loss. *J. Environ. Qual.*

Bishop, P. L., Hively, W. D., Stedinger, J. R., Rafferty, M. R., Lojpersberger, J. L., and Bloomfield, J. A. (2005). Multivariate Analysis of Paired Watershed Data to Evaluate Agricultural Best Management Practice Effects on Stream Water Phosphorus. *J. Environ. Qual.*, 34:1087–1101.

Bosch, N. S., Anne, M., Scavia, D., and Allan, J. D. (2014). Interacting effects of climate change and agricultural BMPs on nutrient runoff entering Lake Erie. *J. Great Lakes Res.*, 40(3):581–589.

Braun, D., Meals, D., Moore, J., and Macrellis, A. (2016). Agricultural Practice Monitoring and Evaluation: Year 4 Report. Technical report.

Braun, D. and Moore, J. (2014). EFFECTS OF A GRASSED WATERWAY ON NUTRIENT AND SEDIMENT RUNOFF: CHARLOTTE SITE CHARACTERIZATION. Technical report.

Burkholder, J., Libra, B., Weyer, P., Heathcote, S., Kolpin, D., Thorne, P. S., and Wichman, M. (2007). Impacts of waste from concentrated animal feeding operations on water quality. *Environ. Health Perspect.*, 115(2):308–312.

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environ. Model. Softw.*, 22(10):1509–1518.

Carter, P. R. (1992). Selecting corn hybrids. *UWEX Bull.*, A3265:8 pp.

CERN (2000). Integrated Assessment of Hypoxia in the Northern Gulf of Mexico. *Natl. Sci. Technol. Counc.*, (May):58.

Chapman, M. and Duggan, J. (2016). The Transition Towards the 2016 Lake Champlain TMDL: A Survey of Select Water Quality Litigation in Vermont from 2003-2015. *Vermont J. Environ. Law*, 17(4):629–650.

Charles, D. (2016). Why Do Milk Prices Spike And Crash? Because It's Like Oil. https://www.npr.org/sections/thesalt/2016/08/05/488708017/why-do-milk-prices-spike-and-crash-because-its-like-oil.

Chiang, L.-c., Chaubey, I., Hong, N.-m., Lin, Y.-p., and Huang, T. (2012). Implementation of BMP Strategies for Adaptation to Climate Change and Land Use Change in a Pasture-Dominated Watershed. *Int. J. Environ. Res. Public Health*, pages 3654–3684.

Collins, M. J. (2009). Evidence for changing flood risk in New England since the late 20th century. *J. Am. Water Resour. Assoc.*, 45(2):279–290.

Corwin, E. (2018). When Nitrate Threatens Well Water, The State's Response Is Inconsistent, Often Undocumented. http://digital.vpr.net/post/when-nitrate-threatens-well-water-states-response-inconsistent-often-undocumented#stream/0.

Daggupati, P., Pai, N., Ale, S., Douglas-Mankin, K. R., Zeckoski, R. W., Jeong, J., Parajuli, P. B., Saraswat, D., and Youssef, M. A. (2015). A Recommended Calibration and Validation Strategy for Hydrologic and Water Quality Models. *Trans. ASABE*, 58(6):1705–1719.

D'Ambrosio, D. (2018). Vermont dairy is in crisis: 4 years of bad prices take a toll even on industry leaders. https://www.burlingtonfreepress.com/story/money/2018/05/29/vermont-dairy-crisis-nordic-farms-folds-can-any-farms-survive/486456002/.

Darby, H. and Bosworth, S. (2004). Early Corn Growth and Development. http://pss.uvm.edu/vtcrops/?Page=articles/CornEmergence.html.

Darby, H., Monahan, S., Cummings, E., Harwood, H., and Madden, R. (2012). 2012 Cover Crop Termination & Reduced Tillage Study. *Univ. Vermont Ext. Publ.*, (March):1–10. http://www.uvm.edu/extension/cropsoil/wp-content/uploads/2012-CC-Term-Reportfinal.pdf.

Dolan, K. (2016). The importance of interagency collaboration and public engagement in the development of the implementation plan for the nonpoint source-focused Vermont Lake Champlain phosphorus TMDL. *Vermont J. Environ. Law*, 17(4):663–687.

Duncan, E. W., Dell, C. J., Kleinman, P. J. A., and Beegle, D. B. (2017). Nitrous Oxide and Ammonia Emissions from Injected and Broadcast-Applied Dairy Slurry. *J. Environ. Qual.*, 46(1):36.

Duriancik, L. F., Bucks, D., Dobrowolski, J. P., Drewes, T., Eckles, S. D., Jolley, L., Kellogg, R. L., Lund, D., Makuch, J. R., O'Neill, M. P., Rewa, C. A., Walbridge, M. R., Parry, R., and Weltz, M. A. (2008). The first five years of the Conservation Effects Assessment Project. *J. Soil Water Conserv.*, 63(6):185A–197A.

Fan, F., Bradley, R. S., and Rawlins, M. A. (2015). Climate change in the Northeast United States: An analysis of the NARCCAP multimodel simulations. *Jounral Geophys. Res. Atmos.*, 120:10569–10592.

Francesconi, W., Williams, C. O., Smith, D. R., Williams, J. R., and Jeong, J. (2016). Phosphorus Modeling in Tile Drained Systems Using APEX. *J. Fertil. Pestic.*, 7(1):1–7.

Frei, A., Kunkel, K. E., and Matonse, A. (2015). The Seasonal Nature of Extreme Hydrological Events in the Northeastern United States. *J. Hydrometeorol.*, 16(5):2065–2085.

Frumhoff, P., McCarthy, J., Melillo, J., Moser, S., and Wuebbles, D. (2007). Confronting Climate Change in the US Northeast: Science, Impacts and Solutions. Technical report. http://www.cabdirect.org/abstracts/20083177506.html.

Galford, G. (2014). Considering Vermont's Future in a Changing Climate. Technical report.

Gassman, P. W., Osei, E., Saleh, A., Rodecap, J., Norvell, S., and Williams, J. (2006). Alternative practices for sediment and nutrient loss control on livestock farms in northeast Iowa. *Agric. Ecosyst. Environ.*, 117(2-3):135–144.

Gassman, P. W., Williams, J. R., Wang, X., Saleh, A., Osei, E., Hauck, L. M., Izaurralde, R. C., and Lowers, J. D. (2010). The Agricultural Policy Environmental Extender (APEX) model: An emerging tool for landscape and watershed environmental analyses. *Trans. ASABE*, 53(3):711–740.

General Assembly of the State of Vermont (2015). No. 64. An act relating to improving the quality of State waters.

Gilker, R. (2005). *Water Quality in Management Intensive Grazing and Confined Feeding Dairy Farm Watersheds.* PhD thesis.

Green, W. H. and Ampt, G. A. (1911). Studies on Soil Physics 1: Flow of Air and Water Through Soils. *J. Agric. Sci.*, 4:1–24.

Guilbert, J., Beckage, B., Winter, J. M., Horton, R. M., Perkins, T., and Bomblies, A. (2014). Impacts of projected climate change over the Lake Champlain basin in Vermont. *J. Appl. Meteorol. Climatol.*, 53(8):1861–1875.

Guilbert, J., Betts, A. K., Rizzo, D. M., Beckage, B., and Bomblies, A. (2015). Characterization of increased persistence and intensity of precipitation in the northeastern United States. *Geophys. Res. Lett.*, 42(6):1888–1893.

Hansen, N. C., Daniel, T. C., Sharpley, A. N., and Lemunyon, J. L. (2002). The fate and transport of phosphorus in agricultural systems. *J. Soil Water Conserv.*, 57(6):408–417.

Hanson, J., Johnson, D., Lichtenberg, E., and Minegishi, K. (2013). Competitiveness of management-intensive grazing dairies in the mid-Atlantic region from 1995 to 2009. *J. Dairy Sci.*, 96(3):1894–1904.

Harmel, R. D., Cooper, R. J., Slade, R. M., Haney, R. L., and Arnold, J. G. (2006). Cumulative Uncertainty in Measured Streamflow and Water Quality Data for Small Watersheds. *Trans. ASABE*, 49(3):689–701.

Hayhoe, K., Wake, C. P., Huntington, T. G., Luo, L., Schwartz, M. D., Sheffield, J., Wood, E., Anderson, B., Bradbury, J., DeGaetano, A., Troy, T. J., and Wolfe, D. (2007). Past and future changes in climate and hydrological indicators in the US Northeast. *Clim. Dyn.*, 28(4):381–407.

Hoerling, M., Eischeid, J., Perlwitz, J., Quan, X. W., Wolter, K., and Cheng, L. (2016). Characterizing recent trends in U.S. heavy precipitation. *J. Clim.*, 29(7):2313–2332.

Horton, R., Yohe, G., Easterling, W., Kates, R., Ruth, M., Sussman, E., Whelchel, A., Wolfe, D., and Lipschultz, F. (2014). NCA 2014: Chapter 16 Northeast. Technical Report October. http://nca2014.globalchange.gov/report/regions/northeast.

Houska, T., Kraft, P., Chamorro-Chavez, A., and Breuer, L. (2015). SPOTting model parameters using a ready-made python package. *PLoS One*, 10(12):1–22.

Howarth, R. W., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., Downing, J. A., Elmgren, R., Caraco, N., Jordan, T., Berendse, F., Freney, J., Kudeyarov, V., P, M., and Z, Z.-L. (1996). Regional Nitrogen Budgets and Riverine N & P Fluxes for the Drainages to the North Atlantic Ocean: Natural and Human Influences. *Biogeochemistry*, 35(1):75–139.

Howden, S. M., Soussana, J.-F., Tubiello, F. N., Chhetri, N., Dunlop, M., and Meinke, H. (2007). Adapting agriculture to climate change. *Proc. Natl. Acad. Sci.*, 104(50):19691–19696.

Huang, H., Winter, J. M., Osterberg, E. C., Horton, R. M., and Beckage, B. (2017). Total and Extreme Precipitation Changes over the Northeastern United States. *J. Hydrometeorol.*, 18(6):1783–1798.

Hudzik, S. (2018). Routine Mailing To Dairy Farmers Included A Rare Note: Suicide Hotline Info. http://digital.vpr.net/post/amid-low-prices-farmers-agri-mark-sends-suicide-hotline-info-milk-checks#stream/0.

IPCC (2014). Climate Change 2014 Synthesis Report Summary Chapter for Policymakers. *Ipcc*, page 31.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2000). *An Introduction to Statistical Learning*.

Jayakody, P., Parajuli, P. B., and Cathcart, T. P. (2014). Impacts of climate variability on water quality with best management practices in sub-tropical climate of USA. *Hydrol. Process.*, 28(23):5776–5790.

Jesse, E. and Cropp, B. (2008). Basic Milk Pricing Concepts for Dairy Farmers. Technical report.

Johnson, M.-V. V., Norfleet, M. L., Atwood, J. D., Behrman, K. D., Kiniry, J. R., Arnold, J. G., White, M. J., and Williams, J. (2015). The Conservation Effects Assessment Project (CEAP): a national scale natural resources and conservation needs assessment and decision support tool. *IOP Conf. Ser. Earth Environ. Sci.*, 25(1):012012.

Jokela, B., Magdoff, F., Bartlett, R., Bosworth, S., and Ross, D. (2004). Nutrient Recommendations for Field Crops in Vermont.

Jones, C. A., Koenig, R. T., Ellsworth, J. W., Brown, B. D., and Jackson, G. D. (2007). Management of Urea Fertilizer to Minimize Volatilization. http://www.uwyo.edu/soilfert/pubs/urea volatilization.pdf.

Kardashian, K. (2012). *Milk Money: Cash, Cows and the Death of the American Dairy Farm*. University of New Hampshire Press, Durham, New Hampshire.

Kaye, J. P. and Quemada, M. (2017). Using cover crops to mitigate and adapt to climate change. A review. *Agron. Sustain. Dev.*, 37(1).

Keim, B. D., Fischer, M. R., and Wilson, A. M. (2005). Are there spurious precipitation trends in the United States Climate Division database? *Geophys. Res. Lett.*, 32(4):1–3.

Kellogg, R. L., Lander, C. H., Moffitt, D. C., and Gollehon, N. (2000). Manure Nutrients Relative to The Capacity Of Cropland And Pastureland To Assimilate Nutrients: Spatial and Temporal Trends for the United States. *Proc. Water Environ. Fed.*, 2000(16):18–157.

Kirkby, C. A., Kirkegaard, J. A., Richardson, A. E., Wade, L. J., Blanchard, C., and Batten, G. (2011). Stable soil organic matter: A comparison of C:N:P:S ratios in Australian and other world soils. *Geoderma*, 163(3-4):197–208.

Kleinman, P. J., Sharpley, A. N., Saporito, L. S., Buda, A. R., and Bryant, R. B. (2009). Application of manure to no-till soils: Phosphorus losses by sub-surface and surface pathways. *Nutr. Cycl. Agroecosystems*, 84(3):215–227.

Kunkel, K. E., Karl, T. R., Brooks, H., Kossin, J., Lawrimore, J. H., Arndt, D., Bosart, L., Changnon, D., Cutter, S. L., Doesken, N., Emanuel, K., Groisman, P. Y., Katz, R. W., Knutson, T., O'brien, J., Paciorek, C. J., Peterson, T. C., Redmond, K., Robinson, D., Trapp, J., Vose, R., Weaver, S., Wehner, M., Wolter, K., and Wuebbles, D. (2013a). Monitoring and understanding trends in extreme storms: State of knowledge. *Bull. Am. Meteorol. Soc.*, 94(4):499–514.

Kunkel, K. E., Stevens, L. E., Stevens, S. E., Sun, L., Janssen, E., Wuebbles, D., Rennells, J., DeGaetano, A., and Dobson, J. G. (2013b). NOAA Technical Report NESDIS 142-1 Regional Climate Trends and Scenarios for the U . S . National Climate Assessment. Part 1. Climate of the Northeast U.S. Technical Report January. http://scenarios.globalchange.gov/report/regional-climate-trends-and-scenarios-us-national-climate-assessment-part-1-climate-northeast.

Lake Champlain Basin Program (2015). 2015 State of the Lake and Ecosystems Indicators Report. Technical report. http://sol.lcbp.org/images/State-of-the-Lake_2015.pdf.

Lal, R. (2008). Promise and limitations of soils to minimize climate change. *J. Soil Water Conserv.*, 63(4):113A–118A.

Lazor, J. (2016). 100% Grass-Fed Dairying- A Farmer's Thoughts. http://butterworksfarm.com/100-grass-fed-dairying-a-farmers-thoughts/.

Macdonald, J. M., Cessna, J., and Mosheim, R. (2016). Changing Structure, Financial Risks , and Government Policy for the U. S. Dairy Industry. Technical Report March.

Magdoff, F., Lanyon, L., and Liebhardt, B. (1997). Nutrient cycling, transformations, and flows: implications for a more sustainable agriculture. In *Adv. Agron.*, pages 1–73.

Markowitz, D. (2017). Climate Adaptation in the Agricultural Sector: Lessons from Vermont.

Marshall, E. and Randhir, T. (2008). Effect of climate change on watershed system: A regional analysis. *Clim. Change*, 89(3-4):263–280.

McKenzie, R. H. (2010). Agricultural Soil Compaction: Causes and Management. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/agdex13331/$file/510-1.pdf?OpenElement.

Mkhabela, M. S., Madani, A., Gordon, R., Burton, D., Cudmore, D., Elmi, A., and Hart, W. (2008). Gaseous and leaching nitrogen losses from no-tillage and conventional tillage systems following surface application of cattle manure. *Soil Tillage Res.*, 98(2):187–199.

Moral, R., Moreno-Caselles, J., Perez-Murcia, M. D., Perez-Espinosa, A., Rufete, B., and Paredes, C. (2005). Characterisation of the organic matter pool in manures. *Bioresour. Technol.*, 96(2):153–158.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE*, 50(3):885–900.

Moriasi, D. N., Gitau, M. W., Pai, N., and Daggupati, P. (2015). Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Trans. ASABE*, 58(6):1763–1785.

Moriasi, D. N., King, K. W., Bosch, D. D., Bjorneberg, D. L., Teet, S., Guzman, J. A., and Williams, M. R. (2016). Framework to parameterize and validate APEX to support deployment of the nutrient tracking tool. *Agric. Water Manag.*, 177:146–164.

Morris, M. D. (1991). Factorial plans for preliminary computational experiments. *Technometrics*, 33(2):161–174.

National Agricultural Statistics Service (2010). Overview of the United States Dairy Industry. Technical report.

Nearing, M., Prusk, F., and O'Neal, M. (2004). Expected climate change impacts on soil erosion rates: a review. *J. Soil Water Conserv.*, 59(1):43–50.

Nearing, M. A. (1998). Why soil erosion models over-predict small soil losses and under-predict large soil losses. *Catena*, 32:15–22.

Nielsen, R. L. (2012). Interpreting Corn Hybrid Maturity Ratings. https://www.agry.purdue.edu/ext/corn/news/timeless/hybridmaturity.html.

Nolan, B. T. and Ruddy, B. C. (2016). Nitrate in the Ground Waters of the United States–Assessing the Risk. https://pubs.usgs.gov/fs/1996/fs-092-96/fs092-96.html.

NRCS (2014). Single Species Cover Crop: Vermont Conservation Practice Job Sheet. https://efotg.sc.egov.usda.gov/references/public/VT/JS340VT_SingleSpecies.pdf.

Oreskes, N., Shrader-frechette, K., and Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263:641–646.

Parsons, B. (2010). Vermont's Dairy Sector: Is There a Sustainable Future for the 800 lb. Gorilla? *Univ. Vermont Cent. Rural Stud.*, 1(4):1–12.

Pettygrove, G. S., Heinrich, A. L., and Eagle, A. J. (2009). Dairy Manure Nutrient Content and Forms. http://manuremanagement.ucdavis.edu/files/134369.pdf.

Pribyl, D. W. (2010). A critical review of the conventional SOC to SOM conversion factor. *Geoderma*, 156(3-4):75–83.

Rabchevsky, G. (1997). Phosphate rock. In US Geological Survey, editor, *Miner. Commod. Summ. 1997*. https://minerals.usgs.gov/minerals/pubs/commodity/phosphate_rock/540495.pdf.

Radel, C., Schmook, B., and McCandless, S. (2010). Environment, transnational labor migration, and gender: Case studies from southern Yucatan, Mexico and Vermont, USA. *Popul. Environ.*, 32(2):177–197.

Ramirez-Avila, J. J., Radcliffe, D. E., Osmond, D., Bolster, C., Sharpley, A., Ortega-Achury, S. L., Forsberg, A., and Larry Oldham, J. (2017). Evaluation of the APEX model to simulate runoff quality from agricultural fields in the southern region of the United States. *J. Environ. Qual.*, 46(6):1357–1364.

Rawlins, M. A., Bradley, R. S., and Diaz, H. F. (2012). Assessment of regional climate model simulation estimates over the northeast United States. *J. Geophys. Res. Atmos.*, 117(23):1–15.

RMA (2012). Why Vermont Crops Fail.

Sakovitz-Dale, J. (2006). Vermont Farmstead Cheese Marketing Study. Technical report.

Saraswat, D., Frankenberg, J. R., Pai, N., Ale, S., Daggupati, P., and Youssef, M. A. (2015). Hydrologic and Water Quality Models: Documentation and Reporting Procedures for Calibration, Validation, and Use. *Trans. ASABE*, 58(6):1787–1797.

Sharpley, A., Kleinman, P., and Weld, J. (2004). Assessment of best management practices to minimise the runoff of manure-borne phosphorus in the United States. *New Zeal. J. Agric. Res.*, 47(4):461–477.

Sharpley, A. and Moyer, B. (2000). Phosphorus forms in manure and compost and their release during simulated rainfall. *J. Environ. Qual.*, 29:1462–1469.

Sharpley, A. N., Chapra, S. C., Wedepohl, R., Sims, J. T., Daniel, T. C., and Reddy, K. R. (1994). Managing Agricultural Phosphorus for Protection of Surface Waters: Issues and Options. *J. Environ. Qual.*, 23(3):437–451.

Sharpley, A. N., Tillman, R. W., and Syers, J. K. (1977). Use of Laboratory Extraction Data to Predict Losses of Dissolved Inorganic Phosphate in Surface Water and Tile Drainage. *J. Environ. Qual.*, 6(1):33–36.

Silberstein, R. P. (2006). Hydrological models are so good, do we still need data? *Environ. Model. Softw.*, 21(9):1340–1352.

Smit, B. and Skinner, M. W. (2002). Adaptation Options in Agriculture To Climate Change : a. *Mitig. Adapt. Strateg. Glob. Chang.*, 7(UNFCCC 1992):85–114. http://www.springerlink.com/index/10.1007/s11027-008-9156-3.

Steglich, E. M., Jeong, J., and Williams, J. R. (2016). Agricultural Policy / Environmental eXtender Model User's Manual Version 1501.

Stokstad, E. (2018). The Truth Squad. *Science*, 361(6408):1189–1191.

Stone (2015). Development of a Web-Based APEX Tool, VT STAR, for Optimizing Best Management Practices and Conservation Planning on Vermont Farms: Final Report. Technical report.

Thompson, R. B. and Meisinger, J. J. (2002). Management Factors Affecting Ammonia Volatilization from Land-Applied Cattle Slurry in the Mid-Atlantic USA. *J. Environ. Qual.*, 31(4):1329.

Tobin, D., Janowiak, M., Hollinger, D., R.H.Skinner, Swanston, C., Steele, R., Radhakrishna, R., Chatrchyan, A., Hickman, D., Bochicchio, J., Hall, W., Cole, M., Hestvik, S., Gibson, D., Kleinman, P., Knight, L., Kochian, L., Rustad, L., Lane, E., Niedzielski, J., and Hlubik, P. (2015). Northeast Regional Climate Hub Assessment of Climate Change Vulnerability and Adaptation and Mitigation Strategies.

Technical report. http://climatehubs.oce.usda.gov/sites/default/files/Northeast Regional Hub Vulnerability Assessment Final.pdf.

University of Arizona (2000). Manure Use and Management. https://cals.arizona.edu/animalwaste/farmasyst/awfact8.html#salt.

US EPA (2015). Phosphorus TMDLs for Vermont segments of Lake Champlain.

USDA-SCS (1972). Hydrology Section 4, Chapters 4-10. In *Natl. Eng. Handb.*

USDA/NASS (2016). 2016 STATE AGRICULTURE OVERVIEW: VERMONT.

UVM Extension (2017). Champlain Valley Crop, Soil and Pasture Team Newsletter.

VAAFM (2016). A Summary of the Required Agricultural Practices.

VAAFM and VANR (2017). Vermont Subsurface Agricultural Tile Drainage Report. Technical report.

Vaze, J., Post, D. A., Chiew, F. H., Perraud, J. M., Viney, N. R., and Teng, J. (2010). Climate non-stationarity - Validity of calibrated rainfall-runoff models for use in climate change studies. *J. Hydrol.*, 394(3-4):447–457.

Vermont Dairy Promotion Council (2014). The Role of Dairy in Vermont: An Economic Assessment. Technical report. http://www.vermontdairy.com/download/VTDairy_EconomicReport.pdf.

Vermont Department of Environmental Conservation (2006). 303 (d) LIST OF WATERS PART A - IMPAIRED SURFACE WATERS IN NEED OF TMDL. Technical report. http://www.champlainparkway.org/_resources/documents/2009FSEIS/Appendix 4 State of Vermont 303d List.pdf.

Walsh, J., Wuebbles, D., Hayhoe, K., Kossin, J., Kunkel, K., Stephens, G., Thorne, P., Vose, R., Wehner, M., Willis, J., Anderson, D., Doney, S., Feely, R., Hennon, P., Kharin, V., Knutson, T., Landerer, F., Lenton, T., Kennedy, J., and Somerville, R. (2014). Ch 2: Our Changing Climate. In *Clim. Chang. Impacts United States Third Natl. Clim. Assess.*, number October, pages 19–67.

Wang, X. and Jeong, J. (2016). APEX-CUTE 4 User Manual.

Wang, X., Potter, S. R., Williams, J. R., Atwood, J. D., and Pitts, T. (2006). Sensitivity Analysis of APEX for National Assessment. *Trans. ASABE*, 49(3):679–688.

Wang, X., Williams, J. R., Gassman, P. W., Baffaut, C., Izaurralde, R. C., Jeong, J., and Kiniry, J. R. (2012). EPIC and APEX: Model Use, Calibration, and Validation. *Trans. ASABE*, 55:1447–1462.

Wang, X., Yen, H., Liu, Q., and Liu, J. (2014). An Auto-Calibration Tool for the Agricultural Policy Environmental eXtender (APEX) Model. *Trans. ASABE*, 57(2012):1087–1098.

Weider, K. and Boutt, D. F. (2010). Heterogeneous water table response to climate revealed by 60 years of ground water data. *Geophys. Res. Lett.*, 37(24):10–15.

Williams, J. R., Izaurralde, R. C., and Steglich, E. M. (2012). Agricultural Policy / Environmental eXtender Model Theoretical Documentation Version 0806.

Williams, J. R., Jones, C. A., Kiniry, J. R., and Spanel, D. A. (1989). THE EPIC CROP GROWTH-MODEL. *Trans. ASAE*, 32(2):497–511.

Williams, J. R., Potter, S. R., Wang, X. S., Atwood, J. D., Norfleet, M. L., Gerik, T., Lemunyon, J., King, A. D., Steglich, E., Wang, C., Pitts, T., and Meinardus, A. (2010). APEX model validation for CEAP. Technical report. ftp://ftp-nhq.sc.egov.usda.gov/NHQ/nri/ceap/APEX_Model_Validation_for_CEAP.pdf.

Winchell, M., Meals, D., Folle, S., Moore, J., Braun, D., Deleo, C., Budreski, K., and Environmental, S. (2011). Identification of Critical Source Areas of Phosphorus Within the Vermont Sector of the Missisquoi Bay Basin. Technical report.

Winslow, M. (2016). A Natural and Human History of Lake Champlain. *Vermont J. Environ. Law*, 17(4):482–500.

Winsten, J., Kerchner, C., Richardson, A., Lichau, A., and Hyman, J. (2010). Trends in the Northeast dairy industry: Large-scale modern confinement feeding and management-intensive grazing. *J. Dairy Sci.*, 93(4):1759–1769.

Winsten, J., Parsons, R. L., and Hanson, G. D. (2000). A profitability analysis of dairy feeding systems in the northeast. *Agric. Resour. Econ. Rev.*, 29(2):220–228. http://ageconsearch.umn.edu/bitstream/31299/1/29020220.pdf.

Winterkorn, H. F. (1959). Theory and Practice of Soil Densification. *Electr. Eng.*, 78(4):302–302.

Wironen, M. B., Bennett, E. M., and Erickson, J. D. (2018). Phosphorus flows and legacy accumulation in an animal-dominated agricultural region from 1925 to 2012. *Glob. Environ. Chang.*, 50(September 2017):88–99.

Wolfe, D. W., Ziska, L., Petzoldt, C., Seaman, A., Chase, L., and Hayhoe, K. (2008). Projected change in climate thresholds in the Northeastern U.S.: Implications for crops, pests, livestock, and farmers. *Mitig. Adapt. Strateg. Glob. Chang.*, 13(5-6):555–575.

Wolkowski, R. and Lowery, B. (2008). Soil compaction : Causes , concerns , and cures. http://www.soils.wisc.edu/extension/pubs/A3367.pdf.

Woznicki, S. A. and Nejadhashemi, A. P. (2012). Sensitivity analysis of best management practices under climate change scenarios. *J. Am. Water Resour. Assoc.*, 48(1):90–112.

Yellen, B., Woodruff, J. D., Cook, T. L., Newton, R. M., and Sites, F. (2016). Historically unprecedented erosion from Tropical Storm Irene due to high antecedent precipitation. *Earth Surf. Process. Landforms*, 684(February):677–684.

A Appendix 1

The APEX files needed to run the PAW1 and WIL2 calibrated models can be found here. This folder contains:

- PAW1_NP19: Input and output files for the PAW1 calibrated model, using historical daily weather data.
- PAW1_climate: Input files for running multiple PAW1 simulations for the climate scenarios in Ch. 4, and some sample output files.
  - For 1980-2009, use RUT_80_09.WP1
  - For Warmer, use RUT_warme.WP1
  - For Wetter, use RUT_plus.WP1
  - For Wet Spring, use RUT_wetSp.WP1
  - For Intense Rain, use RUT_inten.WP1
  - batchJob.ipynb can be used to run multiple simulations
- WIL2_NP8: Input and output files for the WIL2 calibrated model, using historical daily weather data.

- WIL2_climate: Input files for running multiple WIL2 simulations for the climate scenarios in Ch. 4, and some sample output files.
    - For 1980-2009, use BTV_80_09.WP1
    - For Warmer, use BTV_warme.WP1
    - For Wetter, use RUT_plus.WP1
    - For Wet Spring, use BTV_wetSp.WP1
    - For Intense Rain, use BTV_inten.WP1
    - batchJob.ipynb can be used to run multiple simulations

All other material used in the production of this thesis – baseline models, intermediate calibration steps, code used for data manipulation and figure creation, etc. – may be obtained from the author on request.