Chapter 6:

Collaborative groundwater modeling: Open source, cloud-based, applied science at a small-island water utility scale

Abstract

The process traditionally used to execute most groundwater modeling projects has several drawbacks. The typical client-consultant relationship is expensive, produces products with limited longevity, and is technologically dated. Recent advancements in cloud-computing and social-networking are influencing how we communicate professionally, work collaboratively, and approach data-science tasks. Here we show how the groundwater modeling process is especially well positioned to benefit from these technological advancements. This work presents a case study detailing a vertically-integrated, collaborative modeling framework jointly developed by participants at the American Samoa Power Authority and at the University of Hawaii Water Resources Research Center. The framework includes a chain of modular components extending from the direct collection and analysis of climatic and streamflow data through the development of a water budget model and a dynamic regional groundwater model. The process we present is entirely open-source and employs newly available data-science infrastructure such as Python-based tools compiled with Jupyter Notebooks, GitHub, Binder and Microsoft Azure. These resources facilitate the collaborative model development process and deliver seamless integration of multiple computational components into a dynamic cloud-based workflow that is immediately accessible by stakeholders, resource managers, or anyone with an internet connection.

6.1 Introduction

For the last half-century, computational modeling has become one of the primary tools in the water resource manager's toolbox. Groundwater models have become indispensable, industry standard methods for estimating the availability and sustainability of groundwater resources (e.g. Young & Bredehoeft 1972; Cummings & McFarland, 1974; Willis & Yeh 1987). However, because of the inherent complexity of numerical models and the significant time, effort, and expertise needed for their development, it is often challenging for stakeholders and water managers to access models that are appropriate for their needs (Essawy et al. 2018). Oftentimes smaller water utilities or resource management agencies simply do not have the resources to develop their own models or the in-house expertise to assess the validity, conceptualization, calibration, or usefulness of existing models. Within the traditional model development paradigm, water management agencies usually take one of two approaches, (1) dedicate significant resources to building internal modeling capacity or (2) contract with outside 'experts' to deliver models that typically cannot be interacted with once completed. The latter approach generally produces static models that may lose relevance quickly, and are often produced in a format that does not allow end-users to easily modify parameters or address new questions. This approach also suffers from the inherent temporariness of typical funding mechanisms, whereas the modeling process lasts only as long as the project account is solvent, after which a final report is delivered and model files are archived for long-term storage on a server in the back of an office somewhere. Compounding the issue, is the fact that these sorts of models typically require proprietary software or specialized computational environments to run, making it prohibitively challenging for end-users to open and interact with the model. On the other hand, a major drawback to the approach of handling all modeling tasks in-house, is the high cost of training, software, and salary required for agencies to retain personnel with sufficient skills to create and maintain effective models. This level of resource dedication is often only possible for larger utility companies or management agencies, thereby leaving small, remotely located agencies with few to no options for accessing quality modeling tools.

Here we document the ongoing development of a collaborative modeling framework conceived as a joint effort between the University of Hawaii Water Resources Research Center (UHWRRC) and the American Samoa Power Authority (ASPA). Our intended outcomes for the models developed within this framework are commensurate with the motivations behind the participatory and collaborative modeling movement that has, in recent decades, become a highly utilized approach in environmental management. These outcomes are centered around addressing the need for enhanced researcher – stakeholder engagement (e.g. Argent and Grayson, 2003; Liu et al., 2008) and producing, practical, defendable models that sufficiently address stakeholder needs and promote model use in guiding important water management decisions. Such models are intended to incorporate the views, needs, and knowledge of as many stakeholders as possible, including scientists, policy makers, and resource managers.

The terms collaborative and participatory modeling are often used interchangeably in the literature to describe modeling endeavors that involve scientists, modelers, stakeholders, or community members in decision making processes. Basco-Carrera et al. (2017) distinguishes these terms, defining collaborative modeling as a more intensive sub-discipline of participatory modeling. With collaborative modeling, the actual model development responsibilities are more equally shared between participants and the decision making process has a higher level of participation across all levels. Moran (2016) describes the collaborative process as one where, "*…model developers, decision-makers, stakeholders and others work together to develop a shared understanding of* [the region's] *management objectives and the model's role in supporting those objectives*. Langsdale et al. (2013) further refines collaborative modeling as a process where, "*Both the model and the process remain accessible and transparent to all participants, Collaborative modeling builds trust and respect among parties*." Participatory and collaborative modeling methods have been applied across many disciplines ranging from computer science (e.g. Bidarra et al., 2001), to economics (e.g. Mendoza and Prabhu, 2006), and social sciences (e.g. Flint et al., 2017). This approach is especially pertinent in water resource management, as water's indispensable and ubiquitous nature inherently makes any issue a multi-stakeholder concern. In the water resources field, collaborative or participatory approaches have been applied in numerous case studies across a range of technical foci including watershed modeling (e.g. Liu et al., 2008), groundwater modeling (e.g Barfield, 2009; Beall et al., 2011), and water policy and planning (e.g Tidwell et al., 2004).

It is incontrovertible that the adoption of research results by decision makers relies on the usability of the information produced. Such information must be both scientifically credible as well as tailored to the priorities of managers and policy-makers. This means models should be easy for managers and end-users to understand and to interact with. In this context, the success of a model is measured by more than just its calibration, a successful model is able to provide stakeholders with the opportunity to dynamically change model inputs, assess future scenarios, and evaluate model uncertainties. This requires a model design that is physically accessible, i.e. based on open-source codes, and conceptually accessible, meaning the modeling process facilitates sharing of the required core skills for data management, workflow efficiency, and visualization (Pease et al, 2018). Accessible models should also be portable, flexible, use small file sizes, and have short run times to enhance their ease of adoption (Argent and Grayson, 2003). This paradigm views model development as a process, not necessarily an end-goal. Such a process-based approach does require some redefinition of the typical client-consultant relationship, requiring a longer period of interaction and collaboration. However, without this time commitment, it is difficult to support the type of continuous model development that allows updated information to be incorporated as new data or changing conditions are encountered.

The cooperative paradigm is beneficial to both model developers and end users, as it not only encourages the adoption and trust of models but also encourages their responsible use, as users are more aware of limitations, assumptions, and appropriate uses of their tools (Argent and Grayson, 2003). Another benefit is simply the more effective use of limited resources and avoidance of duplicative efforts. When a model is built cooperatively, there is less need for technology transfer, since participating stakeholders are already familiar with the model and its development. This helps to reduce user error leading to poor decision making, which can result from miscommunication of model application or results. Perhaps most importantly, this paradigm allows decision making to be an exploratory process by accommodating new data or information and simulating alternative ideas quickly (Barfield, 2009; Langsdale et al., 2013). The collaborative framework facilitates interaction and dynamic management, by allowing models to be modified based on stakeholder discussion, concerns, and ideas (Pease et al, 2018).

6.1.1 Objectives

The main objective of this chapter is to present a framework for, and a case study of, a small scale, vertically integrated, collaborative groundwater modeling process that takes advantage of recent advancements in cloud-computing and open-source modeling tools. We refer to this framework as vertically integrated because it includes a diverse chain of modular components extending from the direct collection and processing of basic hydrologic parameters, through to the development of a dynamic regional groundwater model. A primary objective of this process is to make every part of it publically available online, so that the process is transparent, reproducible, and easy to understand for interested stakeholders. Anyone with the skill and interest is able to modify inputs, test scenarios, and continue model development for their own ends, and where stakeholders do not wish to interact directly with the model, the open-source code and in-line documentation facilitates discussion and planning between participants. Another objective of the chapter is to demonstrate the ease of use and the applicability of modern code sharing and cloud-computing tools in a scientific modeling setting involving participants at different and remotely located institutions. The tools presented are able to connect researchers and stakeholders through ready-built data science infrastructure that allows advanced modeling techniques to be easily shared, even when either party has limited software development experience.

6.2 Case Study Setting

The island of Tutuila is the main population center of the U.S. territory of American Samoa. It is located near 14° S and 170° W, and at 142 km^2 is the third largest island in the Samoan hot-spot island chain. Geologically, Tutuila contains two distinct provinces. The bulk of the island is composed of an older, highly eroded basaltic shield edifice (1.5 to 1.0 Ma), henceforth referred to as the older-volcanics. Recent (Holocene age) rejuvenation-stage volcanism erupting through the eroded southwestern flank of the older shields created the Tafuna-Leone Plain, a lava delta composed mostly of thin bedded pahoehoe lava flows (Stearns, 1944; McDougall, 1985). The pahoehoe flows of the plain impart a significantly higher overall hydraulic conductivity than is found in the Older-Volcanic Unit, which is composed of a heterogeneous mixture of a'a lava flows, pyroclastic materials, and trachyte domes (Stearns, 1944; Eyre and Walker, 1991). Geological subdivisions within each of these units exist, and may be used as the basis for further refinement into zones with different hydrogeologic properties (Izuka et al., 2007). Tutuila's climate is warm and humid with abundant, year-round rainfall due to its position within the South Pacific Convergence Zone. The island experiences a wetter season with increased precipitation amounts from October to May, and a drier season with less, though still significant, precipitation from June to September. Rainfall varies considerably with location and elevation, and ranges between 1,800 mm/yr near the Tafuna Airport up to more than 5,000 mm/yr along the crest of the highest mountains (Daly et al., 2006). The region is also influenced by tropical storms and hurricanes, and an average of 25 to 30 significant thunderstorms affect the island annually (Kennedy et al., 1987).

In American Samoa, groundwater resources supply over 90% of domestic, and nearly 100% of industrial water use. However, these resources are afflicted by multiple threats to their long-term sustainability. Since 2009, portions of the public water supply system have been unsafe to drink, necessitating one of the longest standing boil-water-advisories in U.S. history. This is partly caused by the vulnerability of Tutuila's young and highly-permeable aquifers to anthropogenic and surface water contamination (Shuler et al., 2017; Shuler et al., 2018). Other aquifers on Tutuila produce high salinity water, presumably caused by salt-water intrusion (Izuka, 1999). In some cases, the island's wells produce water with Cl^- concentrations exceeding the U.S. Environmental Protection Agency drinking water standards by four to five times. Multiple local stakeholders see groundwater models as a tool that will greatly facilitate management of these issues (ASPA, 2013, Anderson-Taggarino personal communication Oct, 2018). While development of groundwater models has been a long standing priority for ASPA, the island's only water utility, limitation of financial and personnel resources has so far, precluded realization of that goal. As of this writing, there have been four known groundwater models developed for portions of Tutuila (Izuka et al., 2007; ASPA, 2013, Shuler et al., 2014, Shuler et al., 2017). While each of these models addressed a specific question, ranging from defining well-capture zones to modeling nutrient transport, none have satisfied the requirements to fully address ASPA's water management needs. The static nature of these models also restricts their ability to be modified, and by nature, most small-scale water utilities do not have the time or resources needed to support building and maintaining the technical capacity necessary for maintaining active modeling projects.

6.2.1 Collaborative Groundwork

The foundation for this collaborative modeling project was based on a formalized working relationship between ASPA and UHWRRC. The American Samoa Power Authority is the sole water utility in American Samoa, and is also responsible for all municipal power, wastewater, and solid waste services. American Samoa is a unique environment as it is small (population of approximately 60,000), geographically isolated (4,000 km to the nearest continent), and a sovereign society still retaining much of its indigenous culture and tradition. Therefore, ASPA is particularly invested in not only meeting customer needs, but also in conservation and responsible stewardship of the island's limited natural resources. The Water Resources Research Center is a technical research unit at the University of Hawaii, and its stated mission is, "*To promote understanding of critical state and regional* [including the U.S. Affiliated Pacific Islands] *water resource management and policy issues through research, community outreach, and public education*." To fulfil this mission in American Samoa, UHWRRC has been working with ASPA and other agencies since 2013 to develop an integrated water resources research program in the territory that strives to incorporate on-island stakeholder concerns into research priorities. The relationship between UHWRRC and ASPA was officially ratified in 2015 through a memorandum of understanding (MOU) focused on collaborative water resources related work. The MOU established long-term goals specifically to (1) develop infrastructure for collection of hydrologic and climatic data, and (2) generate and apply hydrologic data in support of ASPA's water resources management priorities. Additionally, a formal mechanism for identifying stakeholder needs and research priorities was conceived by UHWRRC through establishment of the American Samoa Water Resources Stakeholders Committee. Water-related agency heads were invited to serve as committee members. Since its formation, the committee has participated in annual stakeholder input workshops and meetings on Tutuila, and also provides annual review of Water Resources Research Institute Program grant proposals. A consistently identified priority by the committee is a need for additional groundwater modeling.

6.3 Methods

6.3.1 Modeling Framework

The modeling framework presented in this chapter was developed using a collaborative modeling paradigm, where input, feedback, and participation between ASPA and UHWRRC participants were integrated into the development process from its beginning. Because ASPA's and UH-WRRC's interests overlap in terms of groundwater modeling, development of a management focused groundwater model was prioritized as an end-goal of this effort. The first step identified for achieving this goal was to begin collecting hydrologic data after a 7-year long data gap resulting from the cessation of U.S. Geological Survey (USGS) monitoring operations in 2008. Prior to this time, the USGS maintained numerous weather stations, rain gauges, and stream gauges, many with decades long periods of record. These datasets are imperative for estimating groundwater recharge, which is perhaps the most important input variable for development of groundwater models, especially in island settings with very steep rainfall gradients. To fill this need, UHWRRC and ASPA collaboratively created and now currently maintain a monitoring network consisting of eight stream gauging stations and six weather monitoring stations. ASPA took responsibility for regular station maintenance and downloading data, and UHWRRC covered the initial equipment costs and took responsibility for data processing, quality assurance / quality control (QA/QC) procedures, as well as archiving and distribution of data. To accommodate the extra field-based workload of this project, ASPA agreed to develop a full-time position for a hydrologic technician. This step was a cornerstone in the success of the participatory process, as it significantly increased ASPA's ability to commit to the program success.

These weather stations and stream gauges continually produce raw data in need of QA/QC processing and integration into each complete station record with previous data. This is accomplished with cloud-based processing routines that create output datasets, which can subsequently be supplied as input to a water budget model. We selected the Soil-Water Balance 2 (SWB2) water budget mode developed by the USGS (Westenbroek et al., 2018) to calculate spatially distributed groundwater recharge rates. All water budget inputs and outputs can be processed in the same Python development environment as the aforementioned monitoring network data, resulting in a seamless integration of processing steps, including running the SWB2 code itself. While linking these steps creates a complex network of interdependencies, consistency and organization are maintained through the cloud-based version management application GitHub.

Groundwater model development is the final step in this collaborative modeling process, and this step like those previously described, is performed in the same Python-based development environment, again resulting in seamless integration of all modeling steps to achieve a cohesive workflow. Specifically, the Python module FloPy was selected for most of the necessary pre- and post-processing tasks. The primary input from the ASPA-UHWRRC processing/modeling workflow to the groundwater modeling component is the SWB2 generated groundwater recharge coverage, which is formatted so that it can be used as direct input in a FloPy script. As of this writing, groundwater model calibration is ongoing and dynamic Python based inverse-modeling optimization procedures are currently being explored. The adaptability of this framework greatly facilitates updating the existing model, for example, as ASPA drills more wells and additional water level observations and new pump test results become available. Figure 6.1 shows the entire data and modeling workflow for the ASPA-UHWRRC cooperative modeling case study and these components are individually described in the sections below.

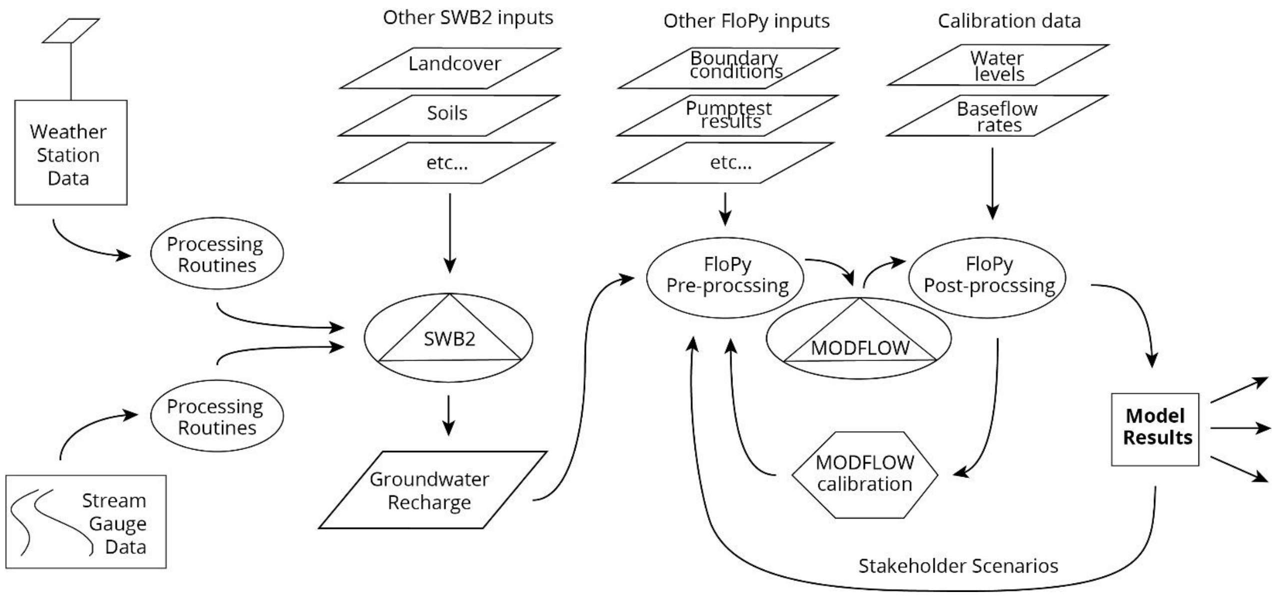

Figure 6.1: Schematic of data and modeling workflow for the ASPA-UHWRRC case study. Datasets or geospatial layer components are shown in quadrilaterals, code-based processes are contained in ovals, and external model executables are contained in triangles, which are themselves within ovals since they are run as Python sub-processes.

6.3.2 Cyber-infrastructure Framework

The cloud based cyber-infrastructure framework used in the UHWRRC-ASPA modeling process was imperative for facilitating communication, providing the ability to collaboratively code, and for taking care of the basic cloud-computing needs that would have been overwhelmingly resource intensive for either participant to develop independently. The selection of processing and computing tools was originally based on the need to make the weather station processing routines and data accessible to both parties, and also to store data in a way that allowed automatic updates to results each time new data became available. We selected Jupyter Notebooks (https://jupyter.org) for coding and project development, GitHub (https://github.com) for project storage and version control, Skype and Google Hangouts (www.skype.com and https://hangouts.google.com) for communication, and (Binder or Microsoft Azure) (https://mybinder.org or https://azure.microsoft.com) for live, cloud-based code execution. All of these services are intuitive, simple, open-source, and integrable with each other, as well as with other widely used cloud-computing services.

The computing languages Python, R, and MATLAB are ubiquitous and commonly used tools in the sciences, especially for model development and application. (e.g. Borah and Bhattacharjya, 2013; Bakker et al., 2016; Yin et al., 2017). However, these tools have historically been difficult for non-experts to access, with steep learning curves and sometimes requiring costly licenses. While mastery of any of these languages can take a lifetime, recent developments in code compilation, support, and shareability are opening up these tools to non-traditional users in ways that have not been previously possible (Perez and Granger, 2015). Jupyter Notebooks (previously called ipython notebooks) are designed to bridge the gap between "coders" and the uninitiated by integrating live code, equations, visualizations, web links, and explanatory documentation, thereby drastically increasing their accessibility to more novice users (Kluyver et al., 2016). A key aspect of Jupyter Notebooks is a code-cell based modularity where blocks of code can be run independently, making the learning or exploration process more manageable and allowing explanations or visualizations to augment each step (Fig. 6.2). While these advances may seem unimportant to those more familiar with coding, in a collaborative framework where team members with variable degrees of expertise wish to be involved in the modeling process, simplicity and ease of access is paramount for everyone's engagement. Owing to its current popularity, this format is very well supported in the development community and a large number of tools are being made available to increase the shareability and accessibility of notebooks. Almost every step in the Tutuila modeling framework project is compiled in Jupyter Notebook format (.ipynb). These notebooks can then directly integrated into the GitHub and Binder or Azure platforms to allow remote access for sharing, developing, and collaboration.

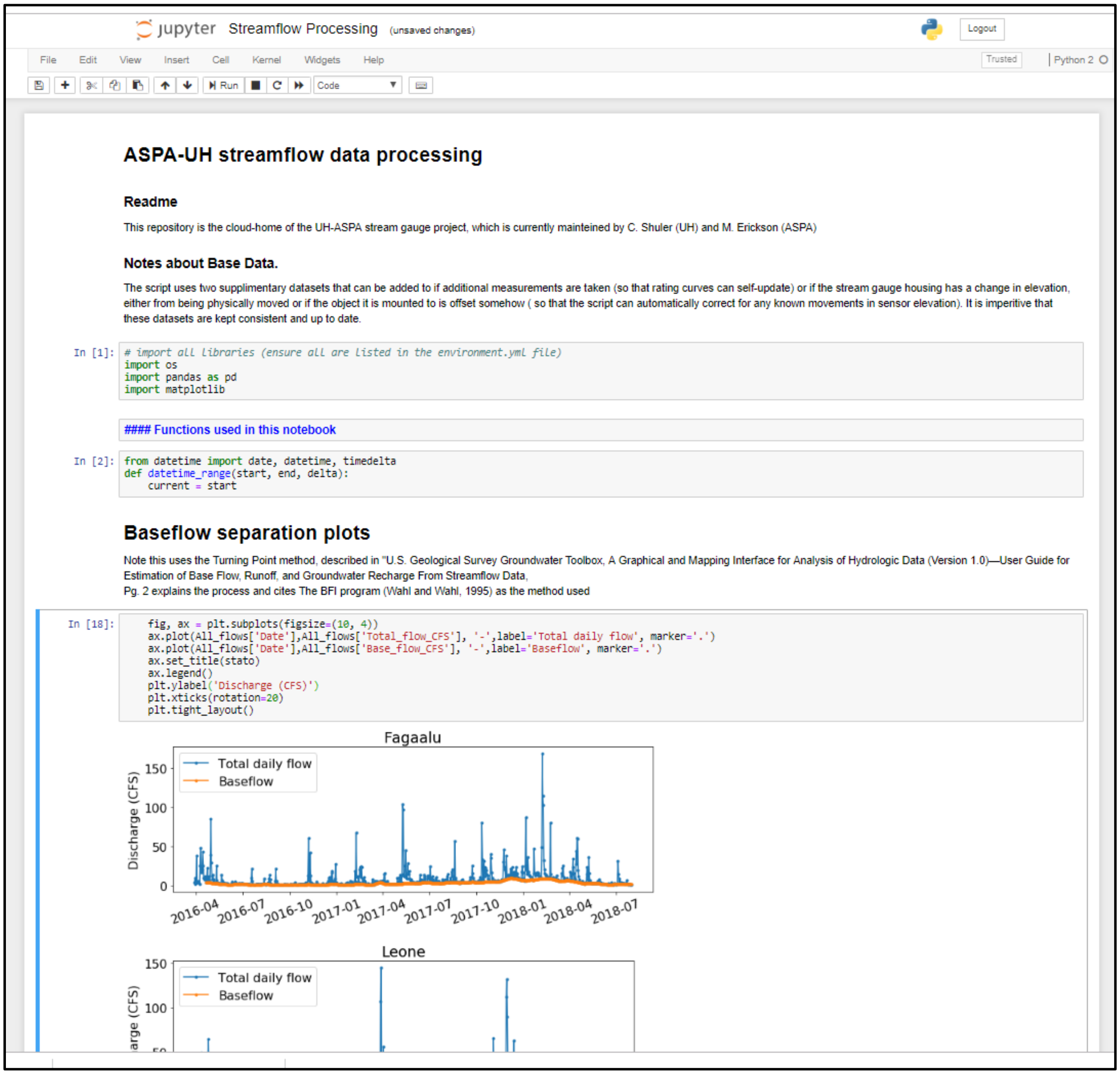

Figure 6.2: Example of a Jupyter Notebook showing code execution cells, live code with notes/comments, explanatory markdown cells, and in-line visualizations. These attributes make notebooks significantly more user friendly to team members with limited data science backgrounds.

GitHub is a free cloud-based file hosting service for version control and management of multi-participant coding projects, and it is becoming increasingly popular as a collaborative coding and data-driven project management tool (Dabbish et al., 2012). It is presently the computer-science industry-standard application for storing, managing, and tracking changes to code. Over 80% of professional developers indicate they use GitHub on a regular basis (Stack Overflow, 2018). GitHub is free (for open-access code) and provides unlimited data storage (as long as each file is under 100 mb). More importantly, the platform provides version control and workflow organization, keeping track of changes made by different users and providing a browser friendly graphical user interface (GUI) to view and explore files, datasets, and results without needing specialized software or computing resources. For this project, GitHub was used to share processing routines and models and also to store and organize all data. Projects are managed in individual units termed “repositories”, which contain all of the data, the code, and the file structure for a process to run. The project repository can be directly downloaded from the web by anyone, providing the ability to share, reproduce, and work on models from start to finish. The repository can be pushed (uploaded) back to GitHub (by authorized team members), and any changes can be viewed and managed before incorporation into the master branch of the model. This not only allows for direct collaborative model development by multiple parties, but also acts as a system for archiving and distributing an entire modeling process including input, historical development, and output files. It also allows the modeling process to be completely transparent, as all the input data, and everything done with it is visible and reproducible by anyone.

There are numerous online video telecommunications services currently available, and we primarily used Skype or Google Hangouts as the main channels of communication between participants on different islands. While this is a fairly simple of part of the framework, it nonetheless was one of the most important, as it facilitated simple, affordable communication and helped to avoid expensive international calling charges to American Samoa. The ability to share computer screens, so that users can see visual output directly, was especially helpful for the modeling process.

While GitHub provides online file storage and organization, it does not provide facilities for running models. Although it is not difficult to install the required open-source software modules needed to run these models on a user’s own computer, it does require the user to install Python and ensure version compatibility between the developer’s and the user’s installed software. This equates to a couple of extra steps and more importantly, introduces the potential for creating problems from incompatible software versions or operating systems. However, there are a number of small scale, cloud-computing resources available online that are dedicated to simplifying this particular issue. These services operate by opening a cloud-based Python environment on a remote server and installing all needed software at the time of use. Current options include Microsoft Azure, Google Colaboratory, and Binder, all of which allow for seamless integration with Jupiter Notebooks and GitHub. Binder is perhaps the most intuitive and straightforward of these options, simply requiring the web address of a GitHub project repository to start a live Python instance. Microsoft Azure, operates in essentially the same way, but does require an account login. At present, we are experimenting with both of these services to determine which will be most useful for this project. These recent advancements in cloud-based, remote-collaboration computing technologies have so far proved to be incredibly useful for facilitating participation between participants in the ASPA-UHWRRC modeling project, despite that fact that we are separated by 3000 km of ocean.

6.3.3 Weather Station Infrastructure

Because of the data gap created in 2008 when the USGS left American Samoa, historical, spatially-distributed rainfall and weather parameter data was becoming more and more outdated. To address this need and to supply data for our modeling purposes, UHWRRC and ASPA worked collaboratively to develop and maintain a network of seven weather stations throughout the island (Fig. 6.3). The stations all have the capability to record precipitation, temperature, relative humidity (RH), wind speed and direction, and solar radiation (SR). The network was initially developed using Spectrum Technologies Inc. WatchDog 2900ET weather stations (Spectrum item number 3350WD2) and Davis Vantage Pro2 Plus weather stations (Davis Item #61612) (Fig. 6.4). These less-expensive and less robust instruments are currently being replaced by solar powered Campbell Scientific stations consisting of an RM Young Wind Sentry Set (03002-L12-PT), a CSL Temperature/RH Probe (CS215-L7-PT), an Apogee SP-110 Pyranometer (CS300-L12-PT), and a Texas Electronics Rain Gauge (TE525-L10-PT) (part numbers refer to Campbell Scientific catalog numbers). All stations are mounted on 2-3 m poles and placed at sites with the best balance of station-siting characteristics considering the available terrain (WMO, 1983; USEPA, 1987). Specifically, the ASPA-WRRC weather station sites were selected based on the following criteria:

1) **Land ownership**—sites already leased or owned by ASPA were highly prioritized.
2) **Minimizing obstructions**— open fields or locations with fewer trees were prioritized.
3) **Spatial distribution**—Representativeness of the variability in Tutuila's climate based on elevation, aspect, location was a priority in site distribution.

All weather stations were set to log data every 15 minutes, and data is downloaded in csv format every 1 to 3 months by the ASPA hydrologic technician. Once downloaded, the ASPA technician simply uploads the raw data files to the project repository on GitHub for processing and for long-term storage. Additional metadata and information about the ASPA-UHWRRC network is presented in Shuler and El-Kadi (2017).

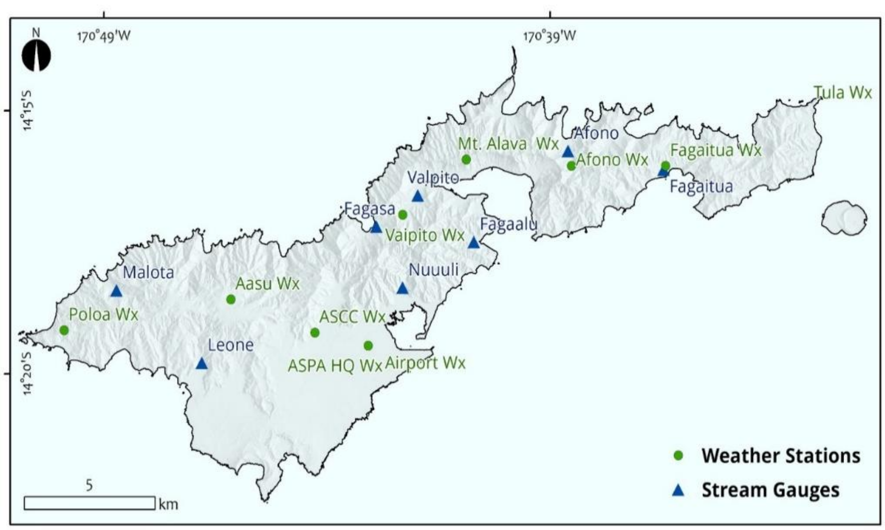

Figure 6.3: Monitoring network instrument map for Tutuila Island with weather stations shown as green circles and stream gauges shown as blue triangles. blue triangles.

6.3.4 Stream Gauge Infrastructure

The ASPA-UHWRRC collaborative framework also included developing on-island capacity for continuous streamflow measurement. Stream gauges were installed cooperatively by staff from both UHWRRC and ASPA over the course of two years. At present, the stream network consists of eight separate continuous record gauges located on different streams throughout Tutuila, and two open-air barometers, all recording at 15 minute intervals. The gauges consist of stainless steel, water level logging, pressure transducers (PT) (HOBO model # U20-001-01) installed in durable steel housings, which are permanently mounted to immobile structures such as bridges or bedrock outcroppings. Housings are made from perforated square galvanized steel pipe with a locking mechanism at the top (Fig. 6.4). The site selection process involved field scoping and soliciting input from multiple departments at ASPA, other on-island stakeholders such as the American Samoa EPA, and also with hydrologists at UH to ensure maximization of data utility. Site selection criteria included considering site access, proximity to historical gauges, bank and channel-control stability, and representativeness of the variability in Tutuila's different climatic and geological regions. Additional metadata and information about the stream gauge network is available in Shuler and El-Kadi (2017).

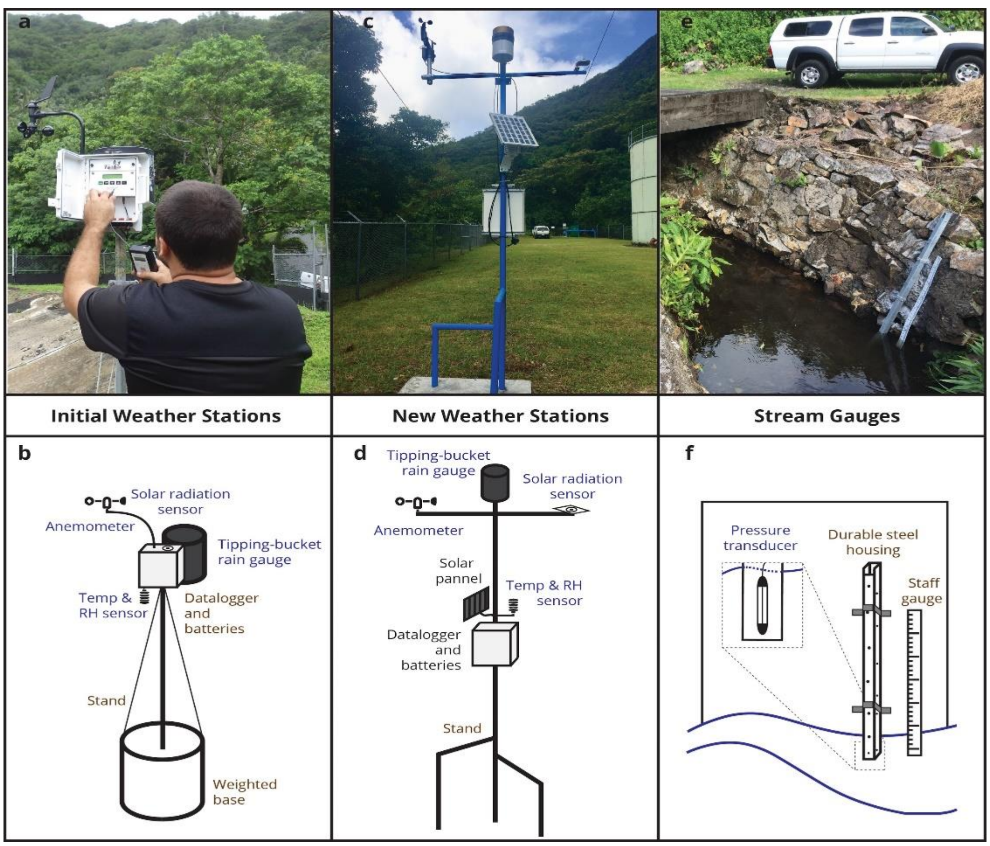

Figure 6.4: Examples of ASPA-UHWRRC weather station and stream gauge network instruments with schematics. Lower grade weather stations (a and b) initially deployed in 2015, are being replaced with new higher-quality stations (c and d) as time and funding permits. Stream gauges are shown in panels e and f. Blue text on schematics indicates sensors and brown text indicates infrastructure used.

6.3.5 SWB2 Model Development

The USGS developed the SWB2 Model (Westenbroek et al., 2018) to allow users to easily calculate water budget components, and specifically groundwater recharge. For this study SWB2 was used to develop a groundwater recharge coverage, which could then be used as an input to the FloPy groundwater model. The SWB2 model is based on a modified Thornthwaite-Mather (1955) soil-water balance approach, which in a simplified form is represented by the following:

$$\textit{Recharge} = \textit{Rainfall} - \textit{Runoff} - \textit{Actual Evapotranspiration}$$

The SWB2 model also includes recent updates based on the Hawaii Water Balance Code (Engott et al., 2017) that make the model better suited to modeling tropical basaltic islands such as the Hawaiian or Samoan Islands. All water balance components are calculated at a daily resolution and output files are produced in NetCDF format. This format can be post-processed with most geospatial software or python scripts can be used to process data into different temporal or spatial resolutions. Runoff-to-rainfall ratios and temporal rainfall distributions derived from our monitoring network data were used as key input variables to the SWB2 model. All other input datasets used for the SWB2 model were obtained from existing publications or databases, with each being described in the respective documentation as cited in Chapter 5 of this dissertation. All SWB2 inputs are either in the form of tabular lookup tables or if spatially-distributed data are used, they are required to be in the ESRI ascii grid format. Input files and sources included:

- Gridded monthly precipitation data (Daly et al., 2006)
- Precipitation gauge data used to represent temporal rainfall distributions (this study)
- Land use data (Meyer et al., 2016)
- Impervious surface ratios (Meyer et al., 2016)
- Canopy coverage ratios (Meyer et al., 2016)
- Soil type data consistent with the NRCS SSURGO database (Nakamura, 1984)
- Direct infiltration data from municipal water line leaks (ASPA, personal communication)
- Direct infiltration data from OSDS effluent discharge (AS-DOC, 2009)
- Runoff-to-rainfall ratios (this study; Perrault (2010); Wong 1996)
- Potential evapotranspiration data in monthly gridded format (Izuka et al., (2005).
- Canopy evaporation data (Engott et al., 2015; AWS Truepower, 2014)
- Gridded monthly maximum and minimum temperature data (Daly et al., 2006)
- Mountain front recharge information (Izuka et al., 2007)

The water budget model, just like the other routines and models used in this project, was designed to be used in a collaborative processed-based manner. Newly downloaded streamflow or rainfall datasets can be used to update the model on an ongoing basis, and if desired, participants can modify input files, change model parameters, run the model, and post-process model results. The version control capabilities of GitHub help to manage different participant's contributions or new scenarios, and also provides the ability to track the model's evolution. Although UHWRRC performed the majority of SWB2 model development, ASPA provided advice in designing portions of the model, specifically relating to the magnitude and distribution of non-revenue water / leaking water lines and direct net-infiltration from OSDS units. Additionally, as of this writing, we are developing additional future land-use scenarios with local stakeholders to assess the effects of possible land-use change on groundwater recharge. Full documentation of the SWB2 model development is provided in Chapter 5 of this dissertation.

6.3.6 FloPy Model Development

Although calibration and validation of the groundwater modeling component of the framework remains ongoing as of this writing, a fully functioning MODFLOW modeling process has been established. One of the primary motivations for development of this component is to ensure that all steps are directly integrable into the existing cyber-infrastructure framework and that the model only depends on tools that can be seamlessly implemented in the Python environment. Additional stipulations for the groundwater modeling process included:

1. Input data need to be simple to modify
2. Recharge from SWB2 has to be seamlessly integrated with future SWB2 updates
3. Updated observation data must be easily incorporated into calibration routines
4. Cell size resolution needs to be simple to modify
5. All model files, need to be small enough to be hosted on GitHub

Meeting these requirements was simplified by using FloPy, an open-source Python package developed by the USGS (Bakker et al., 2016). The primary functionality of FloPy is to support pre-and post-processing methods for the MODFLOW family of models, such as MODFLOW (Harbaugh et al., 2000), MT3DMS (Zheng and Wang, 1999), or SEAWAT (Guo and Langevin, 2002). Presently, the FloPy package is relatively new, but it is rapidly gaining in popularity due to its modularity, open-source availability, and use and support by USGS modelers (e.g. Rotzoll et al., 2016; Feo et al., 2018, Foglia et al., 2018). The bulk of the work in model development typically lies in converting and standardizing input datasets from variable formats into the grid-based formats accepted as input by the MODFLOW executable. FloPy contains many functions to perform pre-processing tasks, and since it is open-source and continually in-development, new functionality is constantly being added. Once input files are converted to appropriate formats, FloPy organizes the input data into a model object, which is input to MODFLOW and run by FloPy as a sub-process. FloPy also contains functionality to post-process, visualize, or reformat output data as desired by the user.

Key benefits of the FloPy method for model construction include: (1) model building and pre-processing steps are quick to execute, (2) the entire process is trivial to reproduce, (3) specific inputs are easy to modify, for example, changing cell size, and (4) the whole modeling process is transparent and easy to share with modeling team participants, as well as with end-users, other researchers, or reviewers. The ease of directly changing the model contributes to its utility in a process-based paradigm, allowing continuous model evolution based on new stakeholder needs, development of new procedures, and incorporation of updated data.

6.4 Process Implementation and Discussion

6.4.1 Weather Data Process Implementation

Weather stations are maintained, and data is downloaded at least quarterly by the ASPA technician. Protocols for data collection and maintenance were developed collaboratively between UHWRRC and ASPA and these are documented in Shuler and El-Kadi (2017). Once downloaded, the ASPA technician uploads raw weather station datasets to the cloud-based repository on GitHub (https://github.com/cshuler/ASPA-UH_Integrated_Modeling_Framework) and new datasets are automatically incorporated with previous datasets once the repository is accessed and the routine is run. Processing the available set of weather station data is simple and can be done by anybody with an internet connection. To process the raw data into summarized files, the user can download the repository from GitHub and run the routine on a personal computer, provided Jupyter Notebook is installed, or online using Binder or Azure. A user with contribution permissions to the GitHub repository, can then upload processed data back to the repository with the push of a button.

Once the weather data processing routine is open in a Python environment, the user can run through the existing steps to update the output with all raw weather data contained in the repository. If desired, the processing routine itself can be modified to change processing steps, generate different visualizations or reformat output data. The processing routine currently includes steps to consolidate and organize raw data files, perform QA/QC checks, remove previously identified bad data, visualize datasets, and summarize data at different time resolutions (Fig. 6.5). The primary output from the weather station processing routine used as input to the SWB2 model, is daily rainfall data and this dataset is directly taken in by the SWB2 pre-processing to ultimately inform calculation of the groundwater recharge coverage.

6.4.2 Streamflow Data Process Implementation

Stream gauge instruments are inspected and downloaded on a monthly basis and streamflow measurements are taken as frequently as possible, especially during high flow events. Once data is collected, the ASPA technician can upload raw water level files, and append new streamflow measurement data to the project repository on GitHub. The repository can be downloaded and processed on a personal computer or on a remote server as described for the weather station data. The streamflow processing steps include barometric compensation, removal of false readings, corrections for physical changes at gauge sites, automated development of rating curves, and summarization of data into daily series, monthly averages, and annual averages (Fig. 6.6). Monthly streamflow summaries produced by this routine are directly used as input in the SWB2 water budget model pre-processing routine to determine runoff to rainfall ratios, which are a key component in determining groundwater recharge rates.

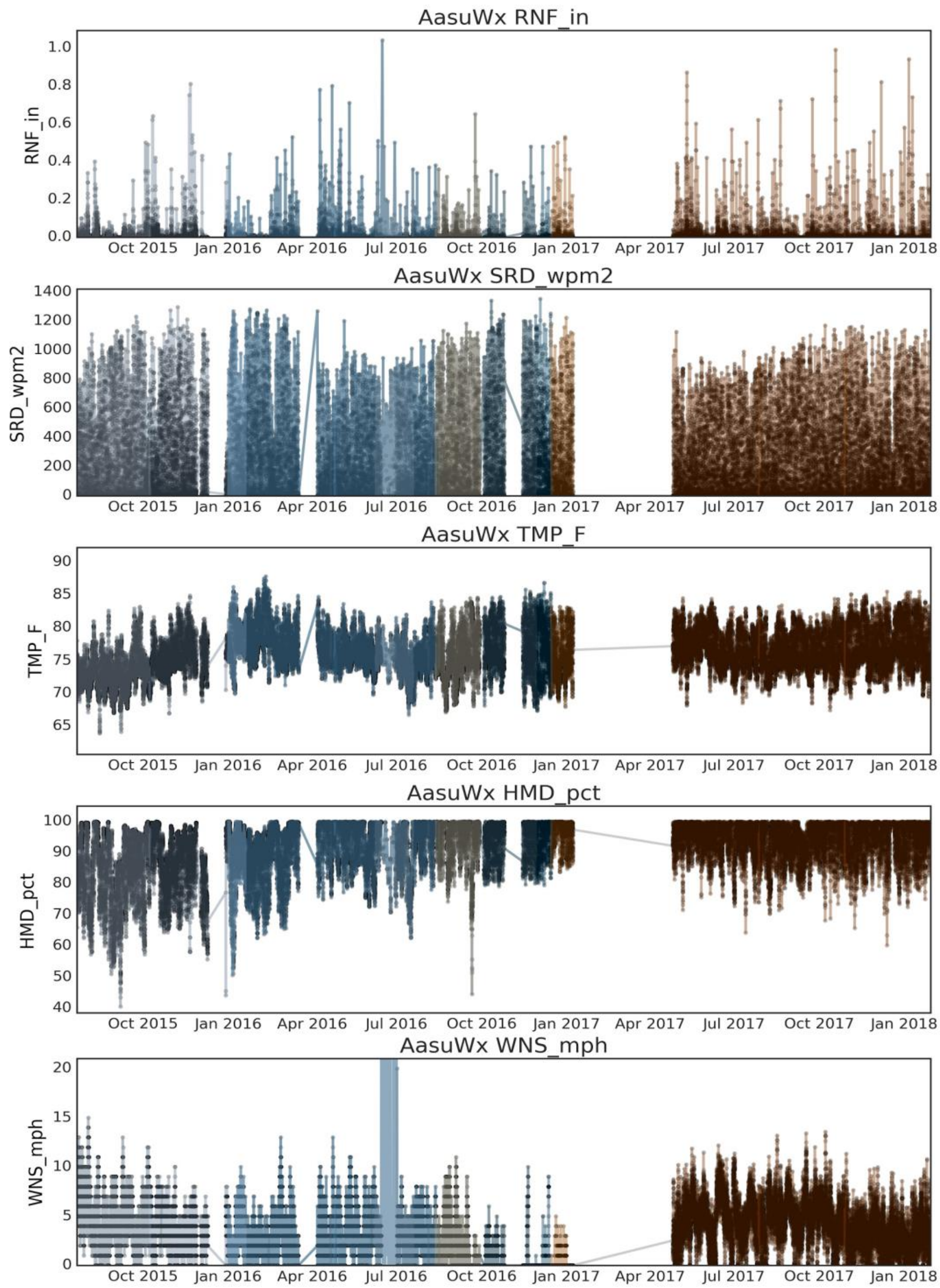

Figure 6.5: Example of consolidated weather station data for one of the ASPA-UHWRRC weather stations (Aasu site). Black line and dots represent consolidated data and colored shaded lines indicate extent of each individual raw weather station data file downloaded by ASPA technician. Note the gap in data where the original station went down, and was later replaced with a new high-quality weather station.

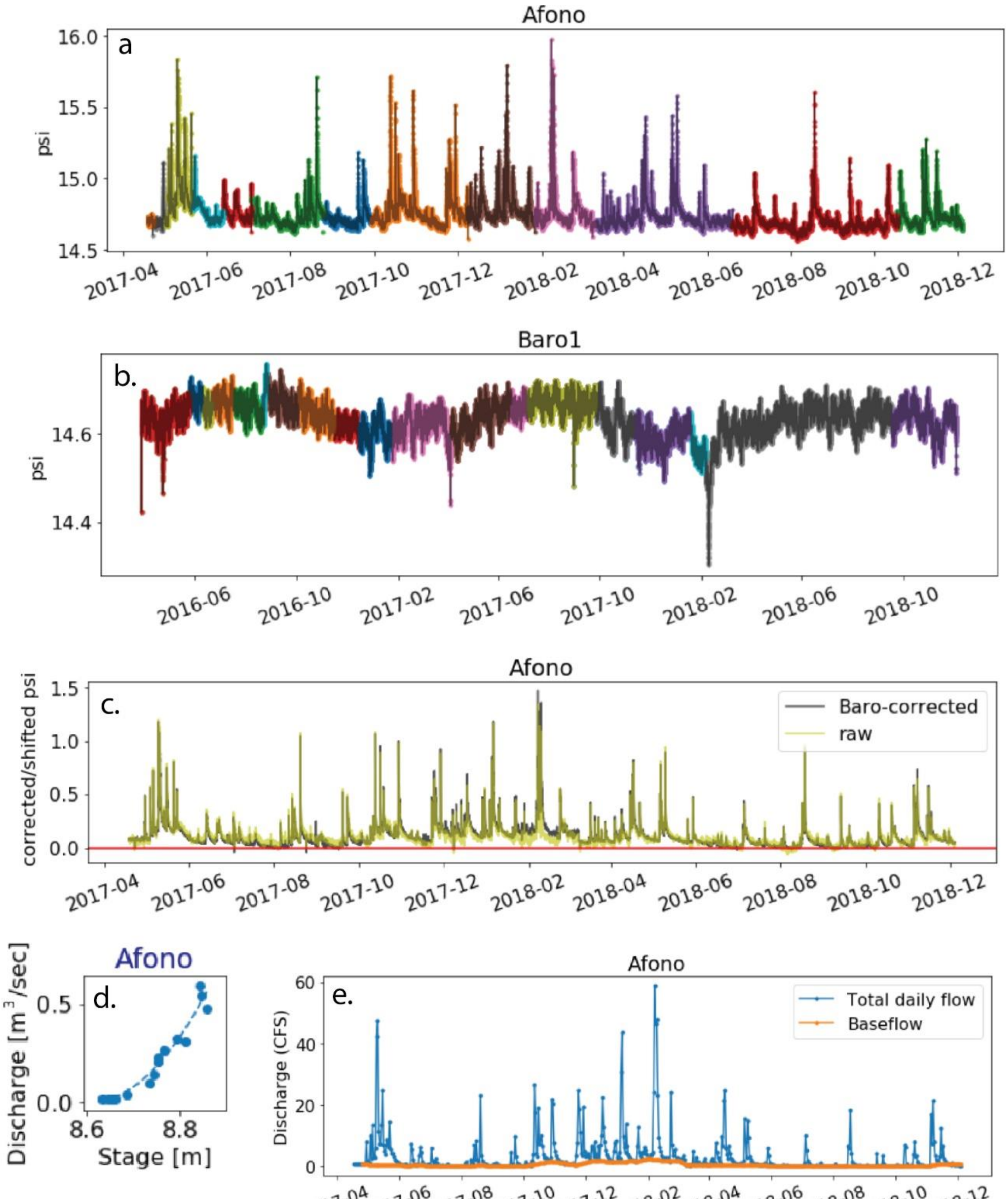

Figure 6.6: Example visualizations for a selection of the procedures implemented in streamflow processing for the Afono Stream Gauge. Steps include but are not limited to, a) consolidation of individual downloaded files (each color represents an individual data file) b) consolidation and incorporation of barometer data (each color represents an individual data file) c) barometric correction of stream stage, d) rating curve calculation, and e) baseflow separation.

6.4.3 SWB Implementation

The SWB2 model component was designed to be run as a series of modular cells, the first of which contain pre-processing routines that format shapefile or raster datasets into ascii grids for use in the SWB2 model. It is possible to modify or substitute different input datasets during any of these steps if desired, to update the model or to assess different scenarios. The SWB executable, which can be downloaded at no cost from the developer's website (https://github.com/smwesten-usgs/swb2), is run as a sub-process from the Jupyter Notebook. Once the SWB2 code is executed, the next set of cells, which contain code for post-processing the model output, are run to produce output data at any resolution or spatial aggregation the user desires. The SWB2 model produces spatially and temporally-distributed datasets for each output parameter in NetCDF file format. These datasets are computed on a daily time step, though it is typically common to summarize water budget results by month or year. At present, the SWB2 notebook is setup to produce volumetric annual totals for each water budget component. With the present array of input datasets, the SWB2 output indicates Tutuila receives a total of 402 Mgal/d as precipitation inputs, and of these inputs, 33 Mgal/d or 8% are lost to canopy interception, 61 Mgal/d or 15% are lost to evapotranspiration, 84 Mgal/d or 21% are lost to island wide runoff, leaving the remaining 54% of inputs, totaling 221 Mgal/d, as the total island wide groundwater recharge estimate (Fig. 6.7). The annual resolution groundwater recharge layer produced by SWB2 is directly integrable in the FloPy pre-processing routine to supply the MODFLOW model with recharge rates at the desired spatial resolution.

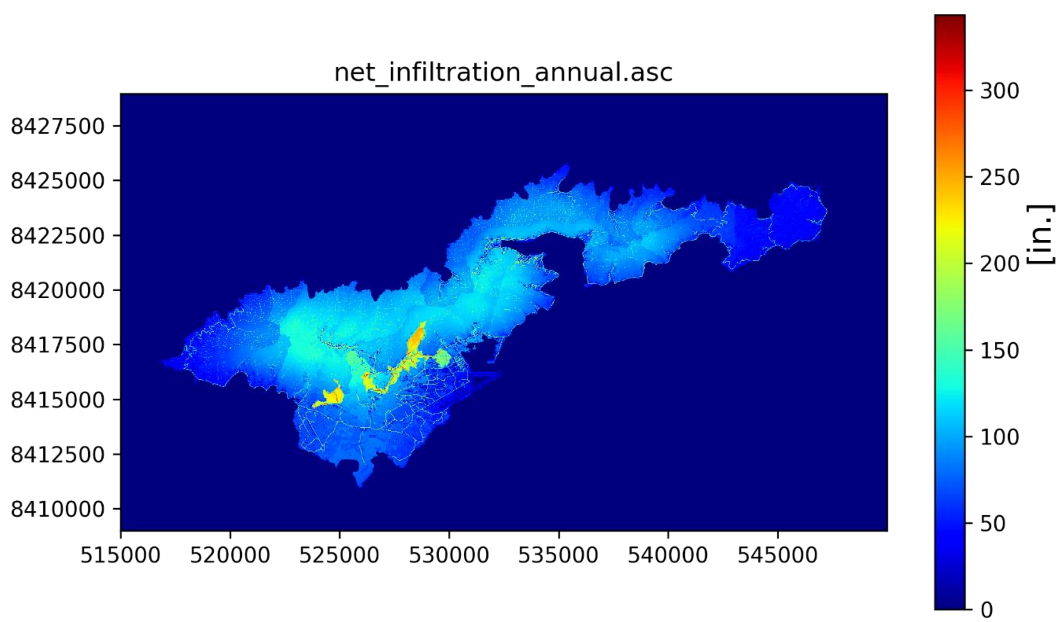

Figure 6.7: Annual average groundwater recharge (net-infiltration) as calculated by most recent SWB2 run, and in appropriate format to be supplied as input to FloPy model. Axes are labeled with UTM coordinates.

It should be noted that because this modeling process is considered to be dynamic, with additions of updated streamflow and weather station data continually being produced by the monitoring network, that this particular model output only represents the latest model iteration at the time of this writing, and may be subject to change as new rainfall, runoff or other input variable data are gathered, or as new information becomes available warranting change of the original conceptual model or decision making process. This also makes it simple to incorporate different scenarios into the model to assess possible future changes. For example, we incorporated future climate scenarios based on gridded dynamically-downscaled climate projections for American Samoa (Wang and Zhang 2016) into the SWB2 model to assess potential climate change effects on the island's water resources, and future land-use scenarios are currently under development through a participatory approach with local stakeholders.

Wang and Zhang (2016) produced 800 m x 800 m gridded-hourly precipitation and temperature predations within three specific scenarios: (1) present-day climate for the years 1990 to 2009, (2) future climate during the years 2080-2099 reflecting a lower-carbon emissions scenario (RCP4.5), and (3) 2080-2099 climate reflecting on a higher emissions scenario (RCP8.5). This set of climate predictions suggests that under both emissions scenarios future climate may see significant increases in both precipitation and temperature. When integrated into the water budget model, this translated into overall increases in all water budget components as calculated by the modified SWB2 runs. Most notably, the 11 to 18% increase in precipitation predicted by the RCP8.5 and RCP4.5 scenarios, respectively, drove increases in net-infiltration rates of 17 to 27%, respectively. However, because these scenarios are only one of the many predictions that are possible, the model framework has the ability to incorporate scenarios from other predictions (such as statistically-downscaled climate predictions) once these are made available. Future land-use scenarios will be incorporated into the SW2 model in a similar manner once the participatory scenario development process with stakeholders in American Samoa is completed.

6.4.4 FloPy Implementation

The MODFLOW model input is structured into a discrete set of required and optional packages, and each are represented by separate input files. Required packages include the basic package (.bas), the discretization package (.dis), and the output control package (.oc). While these files are generally human-readable text or ascii grid files that can be modified by hand, FloPy contains highly useful data formatting functionality to easily take shapefile- or raster-based input data and generate the MODFLOW package files in required formats. Implementation of the Tutuila groundwater modeling framework accomplished all pre-processing steps using FloPy or other Python modules within a dedicated Jupyter Notebook. FloPy also provides functionality to run the MODFLOW code, which needs to be downloaded as a separate executable file (https://water.usgs.gov/water-resources/software/MODFLOW-2005/MF2005.1_12.zip). FloPy runs MODFLOW as a sub-process and prints output directly to the notebook cell. Once the MODFLOW code is run and output files are saved, these can be accessed within the notebook for post-processing routines that display model output or conduct statistical analysis. The Tutuila groundwater model workflow currently includes definition of the required MODFLOW packages as well as those representing boundary conditions (.ghb), hydraulic conductivity zones (.lpf), head observations (.obs), spatially-resampled recharge rates (.rch) derived directly from SWB2, and salt-water interface predictions as described in sections 6.4.4.1 to 6.4.4.6 below.

It should be noted that this iteration of the Tutuila model workflow is presented as an example / proof-of-concept only. Any results presented here are not necessarily accurate or representative of reality, due to the need for additional constraint on many of the model parameters. However, as the cooperative modeling process evolves and as more specific modeling objectives are identified, better constrained parameters can be easily integrated into the existing modeling framework to generate results with lower levels of uncertainty.

6.4.4.1 MODFLOW Model Package Development: Model Geometry

The discretization package (.dis) defines the model geometry and establishes cell size and dimensions, georeferenced coordinates, cell elevations, and model stress periods. The Tutuila case study model cell size can be modified to any desired rectangular dimension and presently, the model has been developed with only one layer, although more can be easily added. The Tutuila model basic package (.bas) establishes the spatial model boundaries, which encompass the shallow marine areas surrounding the island (above 50 m below sea level) and the starting groundwater head levels. Model top elevations are shown in Figure 6.8. A general head boundary (.ghb package), with a head of 0.001 m elevation, is also defined within the region between the model boundary and the island's coastline (Fig. 6.9). This boundary condition is intended to represent the effect of ocean water overlying the submarine island slope, which can discharge water flux, thereby representing the process of submarine groundwater discharge.

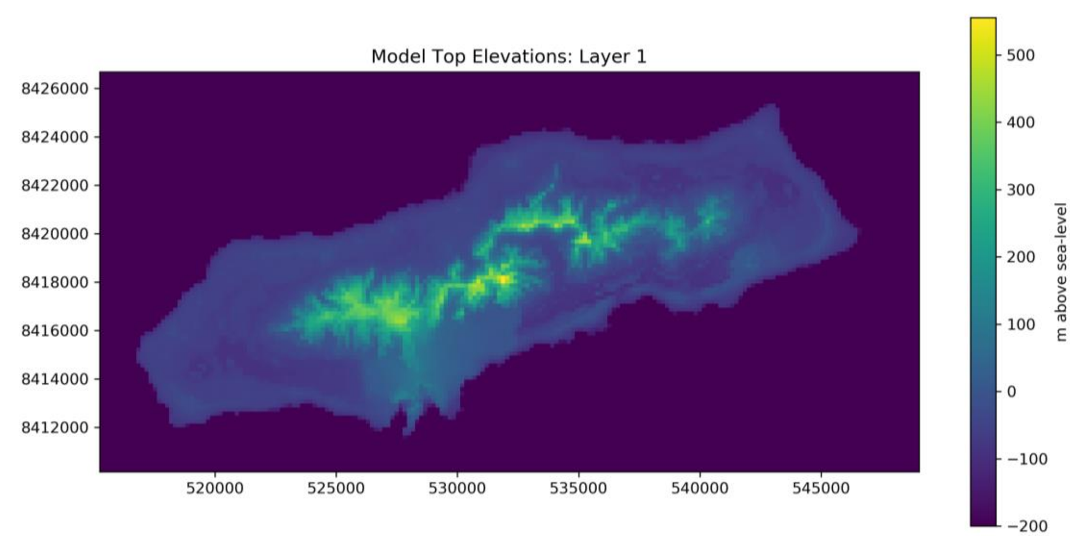

Figure 6.8: Plot of Tutuila case study model grid top elevations derived from a publically available 1/3 arc second digital elevation model.

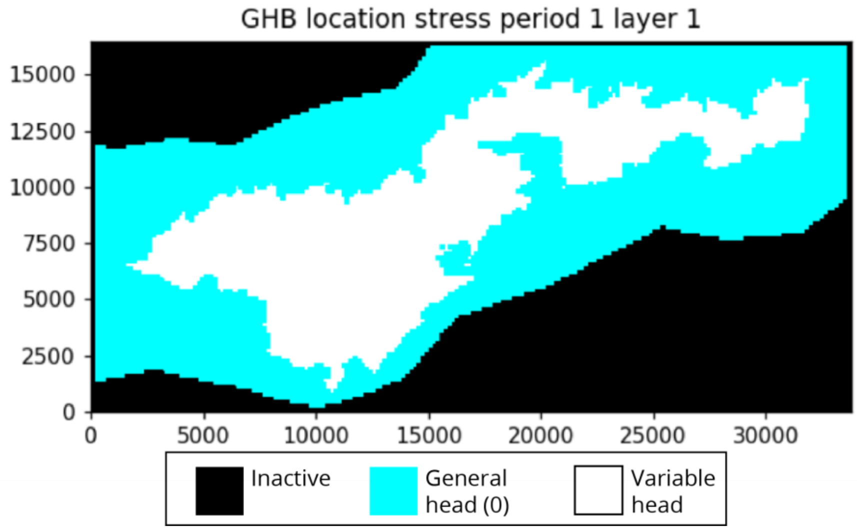

Figure 6.9: Plot of Tutuila case study model grid (at 169 x 165 m cell size) showing model active area (blue and white), and area of general head (.ghb) boundary condition (blue).

6.4.4.2 MODFLOW Model Package Development: Recharge

Spatially distributed groundwater recharge is a key input to the MODFLOW model, and for the Tutuila case study, is provided as an output file from the SWB2 water budget component. The SWB2 model produces a gridded groundwater recharge file in ascii format, and this file is spatially resampled in the groundwater modeling component to a desired grid size (Fig 6.10). Generally, a larger grid size is desirable to speed FloPy processing time. Once resampled, FloPy reformats the groundwater recharge into the MODFLOW input .rch file.

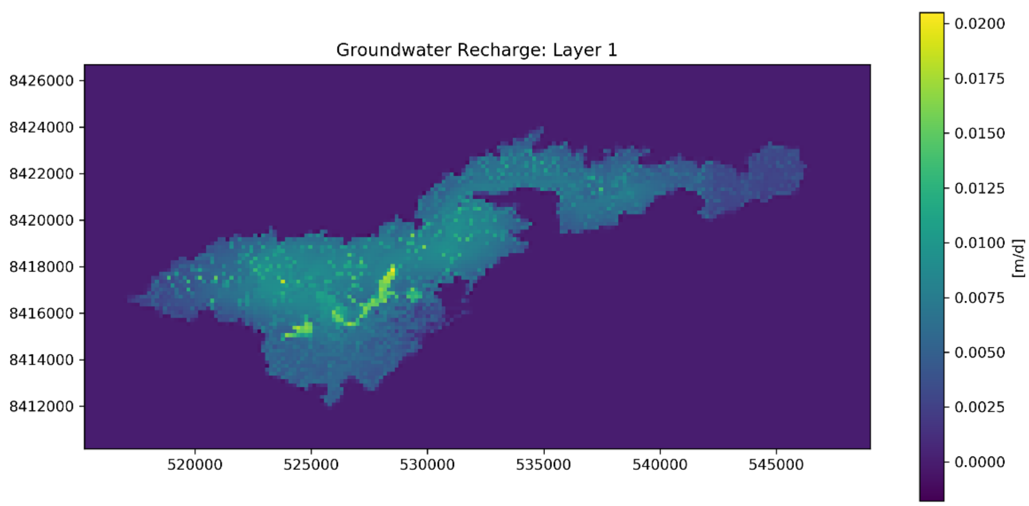

Figure 6.10: Plot of Tutuila case study MODFLOW input groundwater recharge values, resampled from SWB2 water budget model results.

6.4.4.3 MODFLOW Model Package Development: Hydraulic Conductivity

The .lpf package and input file can be used to provide subsurface-flow parameters to the MODFLOW model. These parameters include hydraulic conductivity in both the vertical and horizontal directions, and other parameters affecting subsurface flow such as inter-block transmissivity and cell wetting parameters. Groundwater models are highly sensitive to horizontal hydraulic conductivity values, and thus this property is an important, though difficult to measure parameter due to typically high spatial heterogeneity. Therefore, spatially distributed hydraulic conductivity values used in models are typically obtained through calibration. Numerous optimization approaches exist for this process and many are straightforward to execute in Python. At present, development of an appropriate hydraulic conductivity calibration method for the Tutuila case study remains ongoing. Zone based calibration using a simple Python-based optimizer was applied to develop a first-order estimate of the spatial conductivity distribution and an example of this result is shown in Figure 6.11. As the modeling process and objectives evolve, we intend to apply other techniques for conductivity calibration as well.

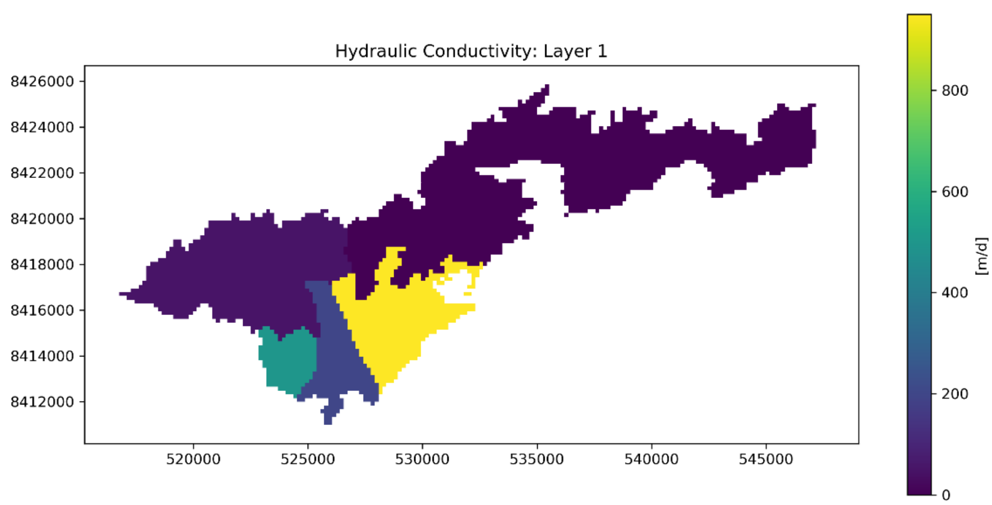

Figure 6.11: Example of hydraulic conductivity zone definition based on simplified geologic units from Stearns (1944). In the current iteration of the Tutuila case-study model each zone is assigned a single value of hydraulic conductivity. Other more distributed parameterization approaches are currently being explored as well.

6.4.4.4 MODFLOW Model Package Development: Head Observations

While essentially any variable calculated by MODFLOW can be used as a calibration parameter, water table elevations, i.e. head observations, are typically used as the primary calibration variable in groundwater models. FloPy is able to quickly format observation data into a MODFLOW-readable .obs file. For Tutuila, pre-development water level observations from wells were compiled from drillers logs and pump test records provided by ASPA or as found in documented literature. Locations of these observation points are shown in Fig. 6.12, left panel, and metadata regarding their sources and uncertainties is provided in Appendix B. To provide a basic example of conductivity calibration here, a simplified zonal calibration approach was applied to the Tutuila model to develop a first-order estimate of a possible hydraulic conductivity distribution. The resulting comparison between observed and simulated water table elevations is shown in Figure 6.12, right panel, and error analysis can either be accomplished through manual calibration, or the model can be wrapped into an object that almost any Python based optimization routine can be applied to further minimize the error. For the purpose of demonstration, the scipy.optimize.minimize solver was implemented here to develop a first order estimate of hydraulic conductivities in the zones delineated in the section above (https://docs.scipy.org/doc/scipy.14.0/reference/generated/scipy.optimize.minimize.html).

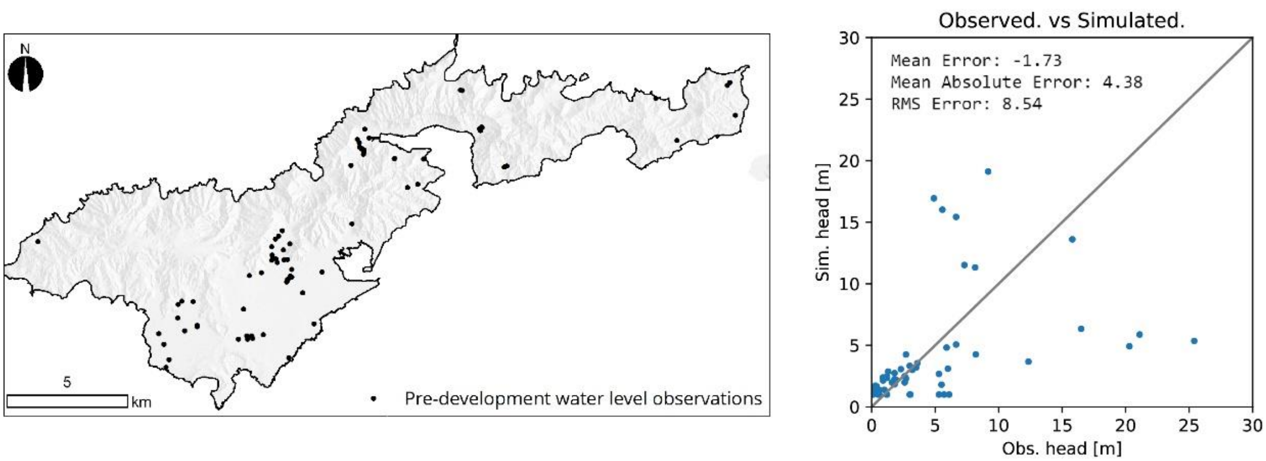

Figure 6.12 (left): Map of water level observation locations and (right): example plot of observed vs. simulated water levels and error statistics as calculated by the manually calibrated FloPy MODFLOW model, demonstrating post-processing visualization. Note grey line in scatter plot is the 1 to 1 line.

6.4.4.5 MODFLOW Model Package Development: Seawater Intrusion

The interaction between salt and freshwater is often modeled with dispersive solute transport models such as SUTRA (Voss and Provost, 2010) or SEAWAT or (Langevin et al., 2008), which are often computationally expensive and require fine vertical discretization of model cells. However, the Seawater Intrusion (SWI2) package for MODFLOW uses a vertically integrated variable-density formulation to simulate vertically integrated variable-density groundwater flow and seawater intrusion. This allows for the position of the 50% freshwater-seawater interface within the saltwater transition zone to be estimated as a discrete sharp-interface or density isosurface. This methodology is simpler and less computationally intensive than that used in dispersive solute transport models. However, because it does not account for diffusion and dispersion, SWI2 cannot definitively predict salt concentrations; it can only approximate the location of the middle of the transition zone.

The primary management utility of this package is to assess the effects of pumping on underlying seawater, which has implications for the salinity or chloride content of produced water. Because the hydraulic conductivity field is the primary control on the position of the seawater interface, at this point in the Tutuila model development, results of the SWI2 analysis should not be considered reliable and are presented for demonstration purposes only. Additionally, due to the lack of direct salinity monitoring information from depths that represent the transition zone, salinity or interface position predictions cannot be calibrated at this time. This underscores the need to enhance Tutuila's groundwater monitoring capacity and encourages development of deeper monitoring wells. As an example visualization, the position of the interface as created by the first-order calibrated model is shown in Figure 6.13.

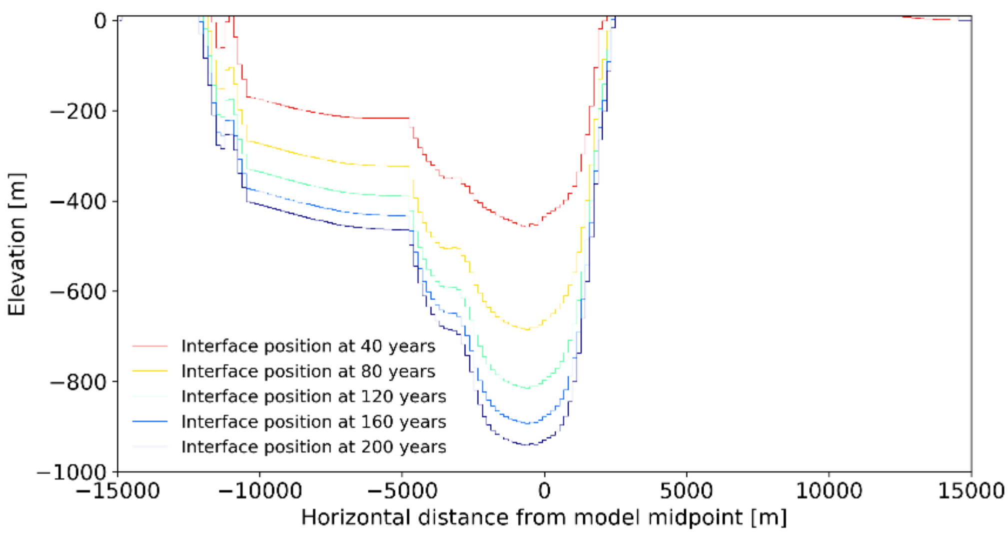

Figure 6.13 Demonstration visualization of the position of the salt-freshwater interface midpoint generated by the SWI2 package as applied to the first-order calibrated Tutuila model. Different colors represent the evolution of the interface position at different time steps as the model runs and moves towards equilibrium. Note that results are not reliable are presented for demonstration purposes only.

6.4.4.6 MODFLOW Model Output: Post-Process Visualization

Since the MODFLOW code is not able to generate plots or any type of graphical or statistical output, the FloPy module provides basic functionality for plotting contour maps and creating other output visualizations. Other Python-based methods for data visualization can be applied as well, simply by reading output files and converting data into the desired format. As an example visualization, a contour plot from the output of the Tutuila case-study model, as parameterized by the first-order zonal calibration method is shown in Figure 6.14. Note again that results are not reliable are presented for demonstration purposes only.

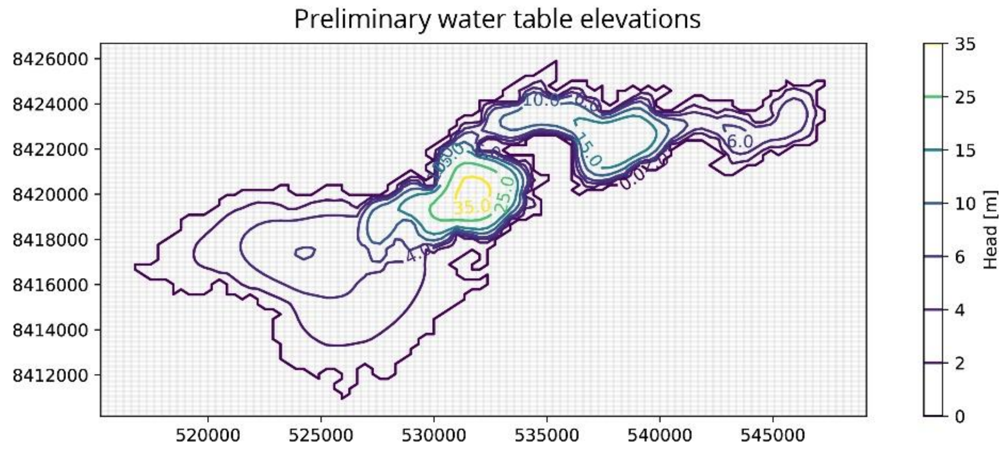

Figure 6.14: Preliminary water table elevations as produced by the Tutuila case-study MODFLOW model. Note that results are shown only for demonstration purposes as the model calibration remains, as of this writing, in an oversimplified state.

6.5 Conclusions

The modeling framework developed for this project integrates weather station data, streamflow data, water budget modeling, and groundwater modeling into a seamless data-to-model workflow. The workflow is made entirely open-source, reproducible, and dynamic by using innovative cloud-computing tools such as Jupyter Notebooks, GitHub, and Binder/Azure. These tools manage the data-science infrastructure, so the project team can focus on communicating with each other and developing models that are scientifically relevant and useful for water resources management. While this framework was deployed in American Samoa it could be easily scaled to other islands or localities.

6.5.1 Continuing Work

This case-study demonstrates a long-term, process-based groundwater modeling approach, that as of this writing is evolving and under active development. While the goals, methods, and preliminary results have been established, there remain numerous planned, and likely many unplanned additions and modifications to be made as stakeholders weigh in, and as our experience with these tools grows. Planned future objectives include continued development with the groundwater modeling component, as well as some additions to the monitoring network. Continued development of the weather station network includes installation of at least one additional station, to be located on Aunuu Island off the south-eastern coast of Tutuila, as well as replacement of older low-quality stations with higher-quality stations. One additional streamflow station is planned for installation at an old USGS gauging site on Aasu Stream located on the northern coast of Tutuila. Collection of streamflow measurements for rating curve updates remains ongoing. The SWB2 model results continues to incorporate updated rainfall and streamflow information future land use predictions are currently being developed with stakeholders and these will be incorporated once completed. The groundwater modeling process is ongoing with continued model calibration, validation using baseflow data, and running groundwater pumping scenarios to assess potential rise in transition zone. The groundwater model has already, and will likely continue to expose data gaps, which can be prioritized in the future. So far these include, (1) developing additional monitoring well capacity, especially in Western Tutuila, (2) development of deep monitoring wells or nearshore wells to monitor the saltwater-freshwater transition (3) Management plan for existing deep geothermal wells, and possible repurposing as salinity monitoring points, and (4) additional constraint on mountain front recharge behavior in the Tafuna-Leone Plain area.

6.5.2 Final Thoughts

The traditional approach to groundwater modeling has several significant drawbacks. It is expensive, it produces products with limited longevity, and it is technologically dated. Recent advancements in social-networking are spilling over into how we communicate professionally, how we work collaboratively, and how we approach data-science. Scientific endeavors, and especially computational tasks such as groundwater modeling, are well poised to take advantage of these new developments. Improvement in the sharability of information is revolutionizing how we work with each other, and this allows for a new process-based paradigm that promotes the maintenance of long-standing project partnerships. While collaborative process-based approaches do require more commitment then static, product-based ones, they allow smaller size agencies or utilities to obtain the benefits of having in-house expertise while being able to defer some of the costs of maintaining dedicated personnel. This also benefits institutions or universities dedicated to providing scientific expertise to large regions with many independent stakeholders. By using cloud-based collaborative tools, institutional staff and researchers can maintain relationships with more agencies in spread out locations while deferring the travel cost and time that would otherwise be required to retain this level of involvement.

The collaborative, process-based approach is especially well suited to development of groundwater models on small islands such as Tutuila, where there is a critical management need for environmental models, but limited resources to develop and maintain the scientific capacity to use them. Groundwater modeling is a complex process and within the traditional paradigm, often takes multiple years of project development to obtain results. During this period, the original research questions may become outdated, and newer more relevant questions may not be appropriate to answer with a model designed for an older objective. For example, in the Tutuila case study, initial modeling efforts were targeted at finding new sites for well development (ASPA, 2013). However, several years later, ASPA's well drilling program has gotten well underway and now addressing sustainable yield and salt-water intrusion is a more pertinent need. When stakeholders are involved in the modeling process at a collaborative level, products can easily evolve to remain relevant to management needs as priorities shift. Additionally, stakeholders will also be more aware of the uncertainties involved with using preliminary results and may be able to apply them to pertinent needs without having to wait for delivery of a polished product. This is evidenced by the vertical integration of multiple different components used in this study, which enhanced all of the participant's understanding of not only the data collection process leading up to the model, but also to the uncertainties implicit in the approaches used and the weight that can be placed upon results.

References: Chapter 6

Argent, R. M., & Grayson, R. B. (2003). A modelling shell for participatory assessment and management of natural resources. *Environmental Modelling & Software*, *18*(6), 541-551.

AS-DOC - American Samoa Department of Commerce. (2009). Buildings layer for Tutuila Island. [Data file]. http://doc.as.gov/

ASPA - American Samoa Power Authority. (2013). Documentation of Wellhead Analysis, Tutuila Hydrogeological Analysis for the U.S. EPA Clean Water Act—State Revolving Fund ASPA Consolidated Grant for the benefit of the Territory of American Samoa. Report prepared for American Samoa Power Authority, Pago Pago, American Samoa. By Walters, M.O.

AWS Truepower (2014). Wind Resource Maps and Data of American Samoa: Final Report. Report prepared for the American Samoa Power Authority. https://www.asrec.net/wp-content/uploads/2016/03/Wind-Resource-Study_2014_DRAFT_2014-10-27-4378826.pdf

Bakker, M., Post, V. +., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW model development using Python and FloPy. *Groundwater,* 54(5), 733-739.

Barfield, D. W. (2009). Collaborative groundwater model development. In World Environmental and Water Resources Congress 2009: Great Rivers (pp. 1-3).

Basco-Carrera, L., Warren, A., van Beek, E., Jonoski, A., & Giardino, A. (2017). Collaborative modelling or participatory modelling? A framework for water resources management. *Environmental modelling & software*, 91, 95-110.

Beall, A., Fiedler, F., Boll, J., & Cosens, B. (2011). Sustainable water resource management and participatory system dynamics. Case study: Developing the Palouse basin participatory model. *Sustainability*, 3(5), 720-742.

Bidarra, R., Van Den Berg, E., & Bronsvoort, W. F. (2001, September). Collaborative modeling with features. In *Proceedings of DET* (Vol. 1, p. 2001).

Borah, T., & Bhattacharjya, R. K. (2013). Solution of source identification problem by using GMS and MATLAB. *ISH Journal of Hydraulic Engineering*, *19*(3), 297-304

Cummings, R. G., & McFarland, J. W. (1974). Groundwater management and salinity control. *Water Resources Research*, *10*(5), 909-915.

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012, February). Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 1277-1286). ACM.

Daly, C., J. Smith, M. Doggett, M. Halbleib, and W. Gibson. (2006). High-resolution climate maps for the Pacific basin islands, 1971–2000. Final Report. National Park Service, Pacific West Regional Office.

Engott, J. A., Johnson, A. G., Bassiouni, M., & Izuka, S. K. (2015). Spatially distributed groundwater recharge for 2010 land cover estimated using a water-budget model for the island of O'ahu, Hawaii (No. 2015-5010). US Geological Survey.

Essawy, B. T., Goodall, J. L., Zell, W., Voce, D., Morsy, M. M., Sadler, J., ... & Malik, T. (2018). Integrating scientific cyberinfrastructures to improve reproducibility in computational hydrology: Example for HydroShare and GeoTrust. *Environmental Modelling & Software,* 105, 217-229.

Eyre, P., and G. Walker. (1991). Geology and ground-water resources of Tutuila American Samoa. Unpublished report in American Samoa Power Authority files.

Feo, A., Zanini, A., Petrella, E., & Celico, F. (2018). A Python script to compute isochrones for MODFLOW. Groundwater, 56(2), 343-349.

Flint, C. G., Jones, A. S., & Horsburgh, J. S. (2017). Data management dimensions of social water science: The iUTAH experience. JAWRA Journal of the American Water Resources Association, 53(5), 988-996.

Foglia, L., Borsi, I., Mehl, S., De Filippis, G., Cannata, M., Vasquez-Suñe, E., ... & Rossetto, R. (2018). FREEWAT, a Free and Open Source, GIS-Integrated, Hydrological Modeling Platform. *Groundwater*, *56*(4), 521-523.

Guo, W., & Langevin, C. D. (2002). *User's guide to SEAWAT; a computer program for simulation of three-dimensional variable-density ground-water flow* (No. 06-A7).

Harbaugh, A. W., Banta, E. R., Hill, M. C., & McDonald, M. G. (2000). MODFLOW-2000, The U. S. Geological Survey Modular Ground-Water Model-User Guide to Modularization Concepts and the Ground-Water Flow Process. *Open-file Report. U. S. Geological Survey*, (92), 134.

Izuka, S.K. (1999). Summary of ground-water eata for Tutuila and Aunuu, American Samoa, for October 1987 Through September 1997. U.S. Geological Survey, Pacific Islands Water Science Center, Open-File Report No. 99-252.

Izuka, S. K., Giambelluca, T. W., & Nullet, M. A. (2005). Potential Evapotranspiration on Tutuila, American Samoa. U.S. Geological Survey Scientific Investigations Report 2005-5200. U.S.Geological Survey

Izuka, S.K., J.A. Perreault, and T.K. Presley. (2007). Areas contributing recharge to wells in the Tafuna-Leone Plain, Tutuila, American Samoa. U.S. Geological Survey Scientific Investigations Report 2007-5167. [http://pubs.usgs.gov/sir/2007/5167/].

Engott, J. A., Johnson, A. G., Bassiouni, M., Izuka, S. K., & Rotzoll, K. (2017). Spatially distributed groundwater recharge for 2010 land cover estimated using a water-budget model for the Island of O ʻahu, Hawai ʻi (No. 2015-5010). US Geological Survey.

Kennedy, Jenks, and Chilton Consulting Engineers. 1987. *Groundwater contamination study Tafuna-Leone Plain Tutuila Island.* Final report for the Environmental Quality Commission, Office of the Governor, Tutuila, American Samoa. 168 p.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... & Ivanov, P. (2016, May). Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB* (pp. 87-90).

Langevin, C.D., Thorne, D.T., Jr., Dausman, A.M., Sukop, M.C., and Guo, Weixing, (2008). SEAWAT Version 4—A computer program for simulation of multi-species solute and heat transport: U.S. Geological Survey Techniques and Methods, book 6, chap. A22, 39 p

Langsdale, S., Beall, A., Bourget, E., Hagen, E., Kudlas, S., Palmer, R., ... & Werick, W. (2013). Collaborative modeling for decision support in water resources: Principles and best practices. JAWRA Journal of the American Water Resources Association, 49(3), 629-638.

Liu, Y., Gupta, H., Springer, E., & Wagener, T. (2008). Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. *Environmental Modelling & Software,* 23(7), 846-858.

McDougall, I. 1985. Age and evolution of the volcanoes of Tutuila, American Samoa. *Pac. Sci.* 39(4): 311–320.

Mendoza, G. A., & Prabhu, R. (2006). Participatory modeling and analysis for sustainable forest management: Overview of soft system dynamics models and applications. *Forest Policy and Economics*, *9*(2), 179-196.

Meyer, R.A., Seamon, J.O., Fa'aumu, S., and Lalogafuafua, L. (2016). Classification and Mapping of Wildlife Habitats in American Samoa: An object-based approach using high resolution orthoimagery and LIDAR remote sensing data. Report prepared for American Samoa Department of Marine and Wildlife Resources.

Moran, T. (2016). Projecting Forward: A framework for groundwater model development under the Sustainable Groundwater Management Act. Stanford Water in the West, Stanford Law School.

Nakamura, S., 1984. Soil survey of American Samoa. U.S. Department of Agriculture Soil Conservation Service, Pago Pago. 95 p.

Pease, E. C., Pfeil, A., Ibarra, V., Reid, M., Siddique, S., Apango, F., ... & Pierce, S. A. (2018, December). Groundwater Modeling with Informatics and Automated Workflows for Water Resource Management: A Case Study from the Northern Trinity Aquifer. In *AGU Fall Meeting Abstracts.*

Perez, F., & Granger, B. E. (2015). Project Jupyter: Computational narratives as the engine of collaborative data science. *Retrieved September*, *11*(207), 108.

Perreault, J.A. (2010). Development of a water budget in a tropical setting accounting for mountain front recharge. Masters Thesis, University of Hawaii at Manoa. Honolulu, HI.

Rotzoll, K., Izuka, S. K., Nishikawa, T., Fienen, M. N., & El-Kadi, A. I. (2016, December). Quantifying effects of humans and climate on groundwater resources of Hawaii through sharp-interface modeling. In *AGU Fall Meeting Abstracts*.

Shuler, C. K., El-Kadi, A.I., & Dulaiova, H. (2014, October). Hunting for High-Level Groundwater on Tutuila, American Samoa. In 2014 GSA Annual Meeting In Vancouver, British Columbia.

Shuler and El-Kadi, (2017). WRRC-ASPA Hydrologic Monitoring Network Handbook. WRRC Special Report SR-2018-02, Water Resources Research Center University of Hawaiʻi at Manoa, Honolulu, Hawaiʻi 96822

Shuler, C.K.; El-Kadi, A.I.; Dulai, H.; Glenn, C.R.; Fackrell, J. (2017) Source partitioning of anthropogenic groundwater nitrogen in a mixed-use landscape, Tutuila, American Samoa. *Hydrogeol. J.* (25), 2419–2434.

Shuler, C. K., Dulai, H., DeWees, R., Kirs, M., Glenn, C. R., & El-Kadi, A. I. (2018) Isotopes, Microbes, and Turbidity: A Multi-Tracer Approach to Understanding Recharge Dynamics and Groundwater Contamination in a Basaltic Island Aquifer. Groundwater Monitoring & Remediation.

Stack Overflow. (2018). *Stack Overflow Developer Survey*. [online] Available at: https://insights.stackoverflow.com/survey/2018/#work-version-control [Accessed 24 Mar. 2019].

Stearns, H.T. (1944). Geology of the Samoan islands. *Geological Society of America Bulletin* 55(11): 1279–1332.

Thornthwaite, C.W., and J.R. Mather. (1955). The water balance. *Publications in Climatology* (Laboratory of Climatology) 8(1): 1–86.

Tidwell, V. C., Passell, H. D., Conrad, S. H., & Thomas, R. P. (2004). System dynamics modeling for community-based water planning: Application to the Middle Rio Grande. *Aquatic sciences*, 66(4), 357-372.

U.S. EPA - United States Environmental Protection Agency (1987). On-Site Meteorological Program Guidance for Regulatory Modeling Applications, EPA-450/4-87-013 Office of Air Quality Planning and Standards, Research Triangle Parks, North Carolina 27711

Voss, C.I., and Provost, A., (2010). SUTRA—A model for saturated-unsaturated variable-density ground-water flow with solute or energy transport: U.S. Geological Survey Water-

Resources Investigations Report 02–4231 (Version of September 22, 2010 (SUTRA Version 2.2).

Wang, Y. and Zhang, C. (2016). Project Final Report - 21st Century High-Resolution Climate Projections for Guam and American Samoa. Retrieved from: https://www.sciencebase.gov/catalog/item/583331f6e4b046f05f211ae6

Westenbroek, S. M., Engott, J. A., Kelson, V. A., & Hunt, R. J. (2018). SWB Version 2.0—A Soil-Water-Balance Code for Estimating Net Infiltration and Other Water-Budget Components (No. 6-A59). US Geological Survey.

Willis, R and Yeh W., (1987) Groundwater Systems Planning and Management, Prentice-Hall, Englewood Cliffs, New Jersey.

Wong, M.F. (1996). Analysis of streamflow characteristics for streams on the island of Tutuila, American Samoa. US Geological Survey Water Resources Investigations Report No. 95-4185, Honolulu, HI.

WMO - World Meteorological Organization (1983). Guide to Meteorological Instruments and Methods of Observation. World Meteorological Organization No. 8, 5th edition, Geneva Switzerland

Yin, D., Liu, Y., Padmanabhan, A., Terstriep, J., Rush, J., & Wang, S. (2017, July). A CyberGIS-Jupyter framework for geospatial analytics at scale. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact* (p. 18). ACM.

Young, R. A., & Bredehoeft, J. D. (1972). Digital computer simulation for solving management problems of conjunctive groundwater and surface water systems. *Water Resources Research*, *8*(3), 533-556.

Zheng, C., & Wang, P. P. (1999). *MT3DMS: a modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide*. Alabama Univ University.