Photovoltaic soiling loss modeling with *pvlib* and Python automation

Capstone in GIS: Final Paper

Manuel Larsen Santos
4/23/2024

Abstract

        The goal of this capstone project is to create a Python script tool for ArcGIS Pro that calculates photovoltaic soiling losses for a user-provided location. As part of the site development process for constructing a solar, or photovoltaic, power plant, system energy losses are estimated in to better understand power plant energy yields. A major contributor to system losses is the accumulation of material on solar panels, known as soiling. This soiling blocks the panel surface which reduces the amount of solar energy absorbed by the panel. To model these losses several approaches have been developed, including techniques that use rainfall and snow data from on-site data collection facilities or nearby weather stations. The PV Soiling Loss Modeler (PVSLM) tool accepts user input coordinates and several modeling parameters to then identify the closest weather station with the required data for soiling loss modeling, download the weather station data, download the corresponding satellite data, produce monthly Kimber soiling loss estimates, produce monthly Townsend snow loss estimates, and create a composite soiling loss estimate table for all modeling techniques. The results are generated in the ArcGIS Pro messages interface, as well as in an output .csv file.

Table of Contents

Introduction

The transition from fossil fuels to renewable energy sources is of mounting importance for the global economy. As solar, or photovoltaic energy, has increasingly become a viable solution for renewable energy demand, more accurate estimates on how much energy a proposed solar energy power plant will produce are of increasing importance. A major contributor to how much energy a solar power plant produces is the obstruction of a solar panel's surface by the accumulation of material like dirt and snow. The resulting reduction in energy output of a solar power plant due to material covering the solar panel surface is known as soiling loss. Soiling is quantified as a percentage of available solar irradiance that is not absorbed by the solar panels. Existing soiling models attempt to estimate the rate of soiling loss, but the process can be onerous, time consuming, and results vary depending on choice of model. The proposed solution is a Python script tool for ArcGIS Pro, called PV Soiling Loss Modeler (PVSLM) that utilizes the *pvlib* Python library to analyze solar irradiance and meteorological data with several soiling loss modeling techniques as a single interface. This tool produces monthly soiling loss estimates that solar resource analysts can use to populate their energy production estimate models.

Literature Review

Research and market studies show that renewable energy, particularly solar energy, is a growing part of the global energy portfolio. While energy demand worldwide is increasing, fossil fuels still serve as much as 80% of global energy consumption (Ali, 2019). Among the available renewable resources, solar energy, or photovoltaic (PV) energy, has been identified as an effective and increasingly efficient means of clean power generation (Georgiou, 2016). According to Solar Power Europe (2023), PV energy has reached nearly 1.2 Terawatts (TW) in installed global capacity, with an additional 341 Gigawatts (GW) built in the last year. Despite increasing PV energy deployment, PV systems make up less than 5% of global need (Solar Power Europe, 2023). In addition to building more PV generation, optimizing operational efficiency is significant for meeting global PV system energy generation goals.

Factors Contributing to PV Energy Production Loss

Many researchers are studying optimization of existing and planned PV site performance to improve energy yield. Performance efficiency losses of PV systems result from system engineering and environmental variables, primarily solar irradiance, temperature, and soiling (Borah, 2023). Studies show PV soiling losses, losses in PV system energy output due to accumulation of particulate matter on solar panels, or PV modules, will likely reduce global PV generation between 4-7% (Bessa, 2021). Research has identified the primary contributors to PV module soiling are weather patterns, pollution, and the attributes of the land where the site is located, like proximity to active farmland and soil texture. Metrics on daily mean particulate matter, precipitation quantities, and length of periods between rain events have been found to be effective indicators of soiling accumulation rates (Micheli & Muller, 2017). Solar monitoring stations (SMS) at PV power plants are used to collect data like particulate matter accumulation, temperature, and precipitation amounts, but this information is typically unavailable for PV sites in early stages of development (Micheli, et al., 2022). Without readily available information on particulate matter accumulation, modeling of soiling losses is done with historical weather data.

Existing Soiling Loss Models

To estimate soiling losses without meteorological data collected at the proposed PV site, models using data collected from local weather stations are in the process of development by researchers and industry professionals. Initial soiling loss estimates are calculated without the benefit of local particulate matter accumulation data, instead assuming standard rates of particulate matter accumulation. This build-up of material is reset by precipitation rain events, those greater than 6mm per *pvlib* default settings (F. Holmgren, 2018). Soiling models based on satellite-derived weather data can identify regional and global patterns, but coarse data resolution and atmospheric interference make the analysis less suitable for site specific soiling loss estimates (Li, et al., 2020). In addition to improving spatial resolution of weather data, smaller

temporal intervals like daily values available through local weather station data allow for the development of soiling mitigation plans (Bessa, 2021).

Snow Loss Modeling

While not an issue in many parts of the world, engineering of PV power plants in areas that may receive frozen precipitation must address this cause of reductions to energy output. Frozen precipitation, typically measured by weather stations as snowfall and snow depth, is an additional contributor to soiling loss that many models have not adequately addressed (Baldus-Jeursen, 2023). Current studies on PV site performance considering system design specifications and environmental variables, including snow measurements, have been found to be inadequate in characterizing soiling losses (Heidari, 2015). Current snow loss modeling uses data for snow accumulation, frequency of snow events, and rates of snow accumulation, but does not consider related meteorological phenomena like hail and ice build-up on modules (F. Holmgren, 2018). A popular composite approach is to calculate soiling loss estimates using soiling models that do and do not include snowfall data and choose monthly soiling estimates based on which model produces the higher soiling loss values for each month of the year.

Ongoing Efforts for Refining Soiling Loss Modeling

Developing a soiling loss model applicable across a wide range of global landscapes and with sufficiently meaningful meteorological data precision is an ongoing effort. Several common soiling loss models do not incorporate snowfall data but are appropriate for areas that do not experience regular snowfall. For areas that do experience regular snowfall, there are now soiling loss models designed for considering snowfall data. The Python library *pvlib* is a widely adopted open-source set of tools for estimating PV system performance and soiling loss modeling with snow loss variables (F. Holmgren, 2018). However, an assumption of *pvlib* modeling, and most soiling models, is that over the operational lifespan of a PV site, typically more than 30 years, average environmental conditions remain consistent (U.S. Department of Energy, 2022). This assumption allows the use of historical weather data to be representative of future conditions for modeling purposes. However, because weather stations are not evenly distributed, the selection of the closest weather station with sufficient historical data is significant to improving model results. To match up with available global solar irradiance data collected by the Landsat 7 Satellite, beginning in 1998, soiling loss modeling typically uses historic data going back to the year 2000. However, the U.S. Department of Energy and Solar Power Europe (2022, 2023) emphasize that results of soiling models that use historical weather data will experience increased uncertainty as global climate change progresses. Studies show that robust validation and comparative analysis of available soiling loss models is lacking, and further research is necessary to improve confidence (Muller & Rashed, 2023).

A New Soiling Loss Model Approach

To facilitate the process of soiling loss modeling for preliminary PV power plant energy production estimates, a Python-based script tool for ArcGIS Pro was developed. This tool, PVSLM, uses functions in the *arcpy* Python library, and the NOAA API to identify the closest weather station with the necessary data for soiling calculations, and downloads the required data. After identifying the weather station and downloading data, the tool uses functions in the *pvlib* library to calculate monthly soiling losses and present results. Solar resource analysts in commercial or research applications will be able to use the proposed Python script tool to analyze the potential energy output of a preliminary PV power plant design more efficiently. The needs of solar resource analysts were assessed by applying principles of Scenario-Based Design, examining how analysts perform soiling loss modeling analysis and how the proposed system would be used (Rosson & Carroll, 2002). In addition, an assessment of the needs of various user-types, known as Persona Mapping, informed the development of the proposed tool (Getto & Moore, 2017). Solar resource analysts within my organization were asked to assess the useability of the tool and provide feedback when evaluating the ongoing viability of the tool. The python tool produces soiling loss values using the Kimber and Townsend models available in *pvlib* and analyzes the data with user-selected modeling parameters. Users of the Python tool will be able to obtain site specific soiling loss estimates across the United States, enabling more effective comparative analysis of models and validation of results with on-site meteorological data that will be collected when a proposed PV site reaches later stages of development.

Summary Points

Despite shortcomings in current soiling loss modeling, for solar power deployment to advance, PV site energy generation estimates must continue with the best available data and methods. The proposed Python tool will automate the weather station selection process and use the *pvlib* Python library to synthesize meteorological data from local weather stations to output site specific soiling loss data necessary for estimating PV energy generation.

Data

Soiling loss modeling requires satellite data for solar irradiance information, and weather station data for information on local meteorological values like temperature, rainfall, snowfall, and snow depth. On-site meteorological data will be used to assess modeling results.

Weather Stations

Weather station location and identification data for all land surface stations in the Global Historical Climatology Network (GHCN) are listed by NOAA, the National Oceanic and Atmospheric Administration (NOAA, n.d.). Weather station location coordinates and station metadata will be used in the Python script to identify the nearest station to the user provided site coordinates.

Global Historical Climatology Network Daily Summaries (GHCNd)

The Global Historical Climatology Network Daily Summaries dataset includes daily statistics for temperature, precipitation, snow fall, and snow depth collected from weather stations across the globe (Menne, et al., 2012). These data are available through the National Centers for Environmental Information (NCEI) and will be used for Kimber soiling loss modeling with the *pvlib* Python library.

US Climate Normals Monthly (CNm)

The U.S. Climate Normals Monthly dataset is a 30-year (1981-2010) average of meteorological data collected from weather stations across the United States (Arguez, et al., 2010). Available by weather station, temperature and snowfall values from these data will be used when populating Townsend snow loss modeling parameters.

Typical Meteorological Year (TMY)

Typical Meteorological Year files, produced by the National Renewable Energy Laboratory (NREL) as part of the National Solar Radiation Database (NSRDB), contain half-hourly values for meteorological and solar irradiance data representing the most typical month of each year over the span of the available satellite data, going back to 1998 (NREL, n.d.). For example, the data for January values may be from 2012, and the data for February may be from 2002. These data are available globally at a resolution of 4 km x 4 km. Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI), temperature, and dew point data will be used for soiling loss models that use snow data, like the Townsend model.

On-Site Meteorological Measurements from Solar Monitoring Stations (SMS)

       To assess the success of the soiling loss modeling with *pvlib*, soiling loss calculations derived from on-site SMS will be used as a baseline for comparison. The six sample sites are in Texas, Arizona, Utah, Maine, Alabama, and Virginia, shown in Figure 1. Data was collected at each of these locations for 1 year. SMS data are made available by Longroad Energy and the locations are generalized for privacy concerns.
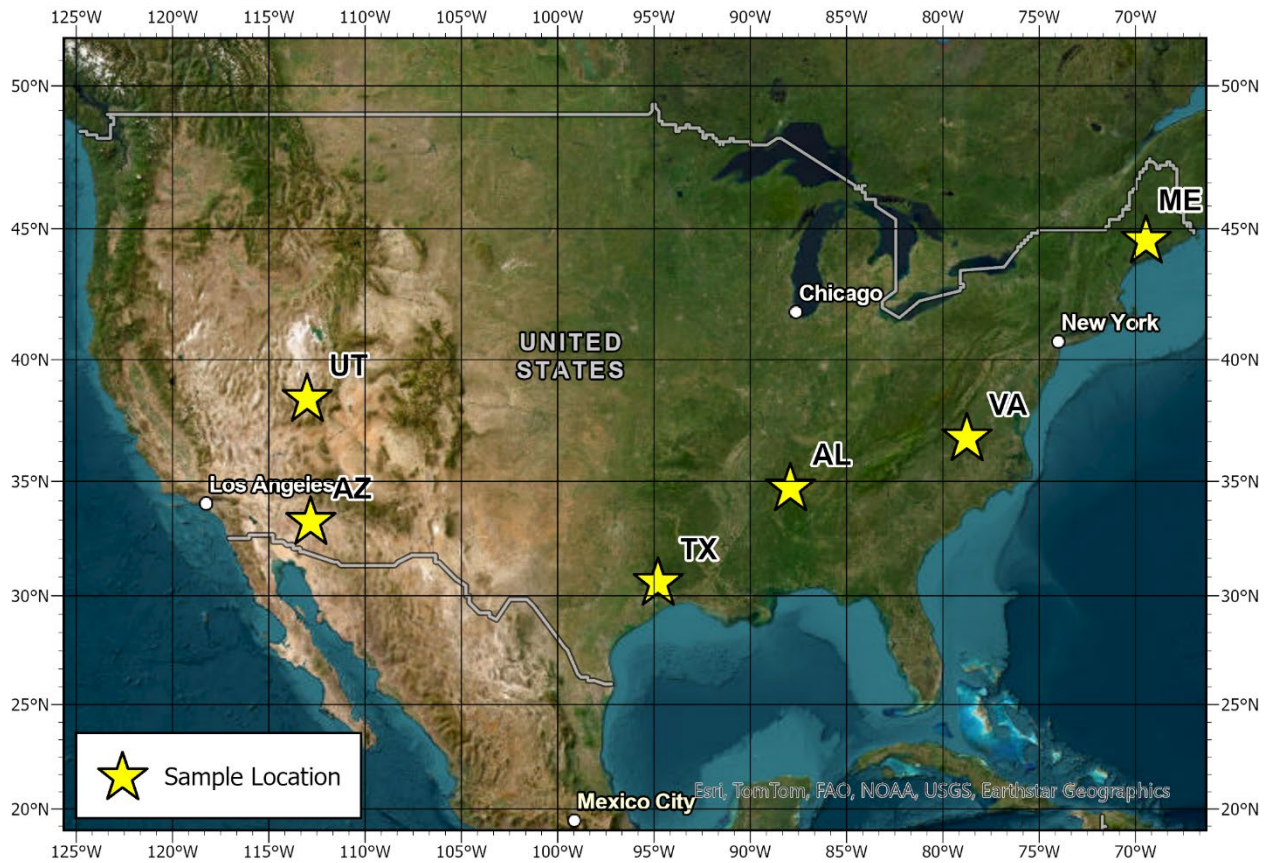

Figure 1. Locations of sample sites where ground data was available for soiling losses. The six sites shown were used as the basis of comparison for PVSLM modeling results.

Methodology

Needs Assessment

The first step in the process is conducting a needs assessment to better understand existing processes for soiling modeling and how the proposed Python script tool will be used. Needs assessment will be determined by communicating with colleagues involved in PV site energy production estimates to design PV soiling loss modeling scenarios. Scenario-based design involves creating a narrative description of how users will complete tasks with the proposed application design (Rosson & Carroll, 2002). Once the needs assessment is complete, development of the proposed Python script tool and its components can begin.

Python Script Development

A Python script will be developed to choose the best local weather station, access, and then process the necessary TMY data, GHCNd, and CNm data. The user will provide coordinates for the area of interest as an input parameter then choose from several parameters to modify the results. The script will use the *arcpy* library's Nearest function to identify the closest weather stations to the input coordinates. The script will then use functions in the *pvlib* library to analyze the TMY GHCNd, and CNm data (Holmgren, et al., 2018). First, the GHCNd and CNm weather station data must be checked to verify there is data coverage going back to 1998, and that all the necessary data types are available. If that weather station has insufficient data, the next closest weather station will be selected and analyzed until a suitable weather station has been identified. Next, the soiling loss parameter Plane of Array (POA) irradiance must be calculated using the TMY data. POA is how much solar irradiance a solar panel receives based on the angle of the solar panel and the position of the sun, measured in watts/m^2 (Sandia National Laboratories, n.d.). POA calculations are performed using either the Hay or Perez transposition models. The script will return monthly soiling values based on the user's selection of POA transposition models and soiling loss model parameters. The higher of each of the monthly values, between the soiling model and the snow loss model, will be used for the output values.

ArcGIS Pro Python Script Tool Development

After the Python script for soiling loss modeling has been developed, a soiling loss modeling Python script tool will be created using ArcGIS Pro. The soiling loss modeling tool will provide a graphic user interface for the user to provide location information, in the form of coordinate values. Then the user can review the Python script tool parameters and select the desired options for soiling loss modeling. The user will have the option to generate a formatted document with the soiling loss data tables and results. This tool will be accessible as a Python toolbox for ArcGIS Pro.

Useability Testing and Evaluation of Modeling Results

   To assess the outcome of the proposed Python script tool, useability testing and cost/benefit analysis will be conducted, as well as an evaluation of modeling results in comparison to ground measurements. Several photovoltaic analysts (4) within my organization will be invited to use the python tools and provide feedback. A cost/benefit analysis of the python script tools will examine the soiling loss modeling process with and without the use of the proposed new tools. To assess the results of the soiling loss modeling, six test sites were chosen based on availability of on-site soiling loss measurements. The six sites, located in Texas, Arizona, Utah, Maine, Alabama, and Virginia are representative of many different environments where PV energy is being developed in the United States. The Texas and Arizona sites are arid locations with little to no snow. The Utah site is higher elevation, arid, and experiences some snow. The Maine site experiences frequent rain and snow events. The Virginia and Alabama sites experience frequent rain and occasional snow events. Soiling loss calculations for these six sites, derived from on-site measurements, will be compared to modeling results.

Resources Required

   The tools required for completing this project involve computer hardware, software, and programming languages. A laptop computer using the Windows 10 Pro operating system will be used for all phases of the development process. ArcGIS Pro will be used to work with preliminary data and test results. PyCharm will be used to develop the Python script. Python will be the primary programming language. The *arcpy* Python library will be used for processing spatial data like the identification of closest weather stations. Several *pvlib* Python functions will be used for the analysis and modeling of soiling losses. The NSRDB and NCEI Application Programming Interfaces will be used to access satellite and weather station data, respectively.

Results

Needs Assessment

After interviewing solar resource analysts and developing scenarios for soiling loss modeling, several design priorities, and potential benefits of the PVSLM tool were identified. The two scenarios described in Table 1 illustrate some of the efficiencies and shortcomings with available soiling loss calculation methods. Scenario 1 describes an option where the analyst has some proprietary tools developed but requires several steps with direct analyst involvement. Scenario 2, while arguably more efficient than Scenario 1, does not have the capacity to readily analyze weather station data and demands a comfortable working knowledge of programming with Jupyter Notebooks. The PVSLM tool can address these concerns by simplifying the data retrieval and analysis processes.

Scenario 1. Soiling Loss Modeling with Excel Spreadsheets
The solar resource analyst needs to calculate soiling loss estimates for a proposed PV site design. The analyst does not have much programming experience but is comfortable working with Excel spreadsheets. They already have a preformatted Excel file with proprietary methods used for calculating soiling losses. The analyst goes to the NOAA Climate Data Online search website, searches by the name of the nearest city and browses nearby stations until they find one that has the precipitation data for the timespan they need. After downloading the weather station data, the analyst copies the meteorological data columns and pastes into a preformatted Excel worksheet, then updating pivot tables to get monthly Kimber soiling losses. Several parameters are available to modify Kimber soiling loss results. To calculate Townsend snow losses, the analyst goes to the NOAA website for US Climate Normals Quick Access, searches for the station, and browses to find a nearby weather station with snow data. The analyst then downloads the Climate normals data, copies the precipitation and snowfall tables into preformatted Excel worksheets. Next the analyst downloads satellite TMY data for the location through one of several public or private data portals. Then the relative humidity column from the TMY data is copied into the Excel worksheet. After this, the analyst prepares several site design components in PVsyst system modeling software to produce the POA Irradiance measurements for a site. The POA measurements are also copied into the Excel worksheet. After this, the analyst updates all pivot tables in the Excel workbook and checks the soiling loss estimates column.
Scenario 2. Soiling Loss Modeling with Jupyter Notebooks
The solar resource analyst needs to calculate soiling loss estimates for a proposed PV site design. The analyst is comfortable working with Python and Jupyter Notebooks. They are going to use Jupyter Notebooks to calculate soiling losses with available satellite data. The analyst opens a web browser, logs in to their Jupyter Notebooks account and and creates a project for the proposed site. The analyst prepares the Python environment by importing the necessary Python packages, like *pvlib*. Then the analyst downloads the satellite data for the place of interest with *pvlib* and creates a dataframe with the satellite data. Using the Kimber soiling loss model in *pvlib*, the analyst calculates soiling losses and moves on to prepare the data inputs for snow loss modeling. They calculate POA Irradiance with the *pvlib* function by first calculating airmass and extraterrestrial DNI with corresponding *pvlib* tools. With all the loss inputs prepared, the analyst then calculates snow losses with the *pvlib* Townsend snow loss model. The analyst then checks the two soiling loss tables in a new dataframe.

Table 1. Two scenarios describing how a solar resource analyst may accomplish PV soiling loss modeling.

Python Script Tool Development

The PVSLM Python script tool was created using functions in the *arcpy* and *pvlib* Python libraries to identify the nearest weather station with sufficient data, get satellite data for that location, then analyze the data to produce monthly soiling loss estimates with Kimber and Townsend modeling. The process for selecting the nearest weather station with complete data is working successfully. Figure 2 shows an operational version of the PVSLM tool, accessed using ArcGIS Pro software. The user provides latitude and longitude for a place of interest, as well as a folder path where downloaded data will be saved. The user must also provide their NOAA API token and indicate a starting year for requesting GHCNd data. If the user already knows which weather station they want to use for data requests, they can provide that station ID as an optional parameter. Providing the weather station ID will skip the weather station selection process and immediately download the GHCNd and CNm data. If the user already has the weather station or satellite data, they can provide the paths to these files and PVSLM will use the data in these files instead of downloading data. The soiling loss, plane of array, and snow loss parameters are all prepopulated with values from *pvlib* and typical design specifications based on user feedback. Tool parameters are subject to change per tech and location, at the user's discretion.

If the user does not provide a weather station Id value, all GHCN weather stations within 150 km are ranked by distance to the input coordinates. Then the weather station data are checked, in order of distance from the input coordinates, to confirm that GHCNd data has precipitation values, and CNm data has temperature and snow values. The tool proceeds through the list of weather stations until the data availability is confirmed. Then the GHCNd data are downloaded for all years from the user-defined start year to the present and the CNm data are downloaded as .csv files. The final step in the data retrieval process is downloading the TMY data from Geographical Information System (PVGIS).

After choosing the weather station and downloading data, PVSLM continues to the data analysis. The GHCNd dataset precipitation values are used in the *pvlib* Kimber soiling model parameters to produce the soiling loss estimates without snow. Next the inputs for the Townsend snow loss model are prepared. Snow loss modeling requires several additional steps to prepare data, specifically the calculation of POA Irradiance. To calculate the POA Irradiance, relative airmass and extraterrestrial DNI are first calculated using the TMY data and corresponding *pvlib* functions. Then the *pvlib* Townsend snow loss model is used to calculate monthly snow losses. The two loss models, Kimber and Townsend, are compared and the higher loss values for each month are used in a final monthly soiling losses table. The tool input parameters, soiling losses table, and POA Irradiance table are all exported to a .csv file so the analyst can review results.
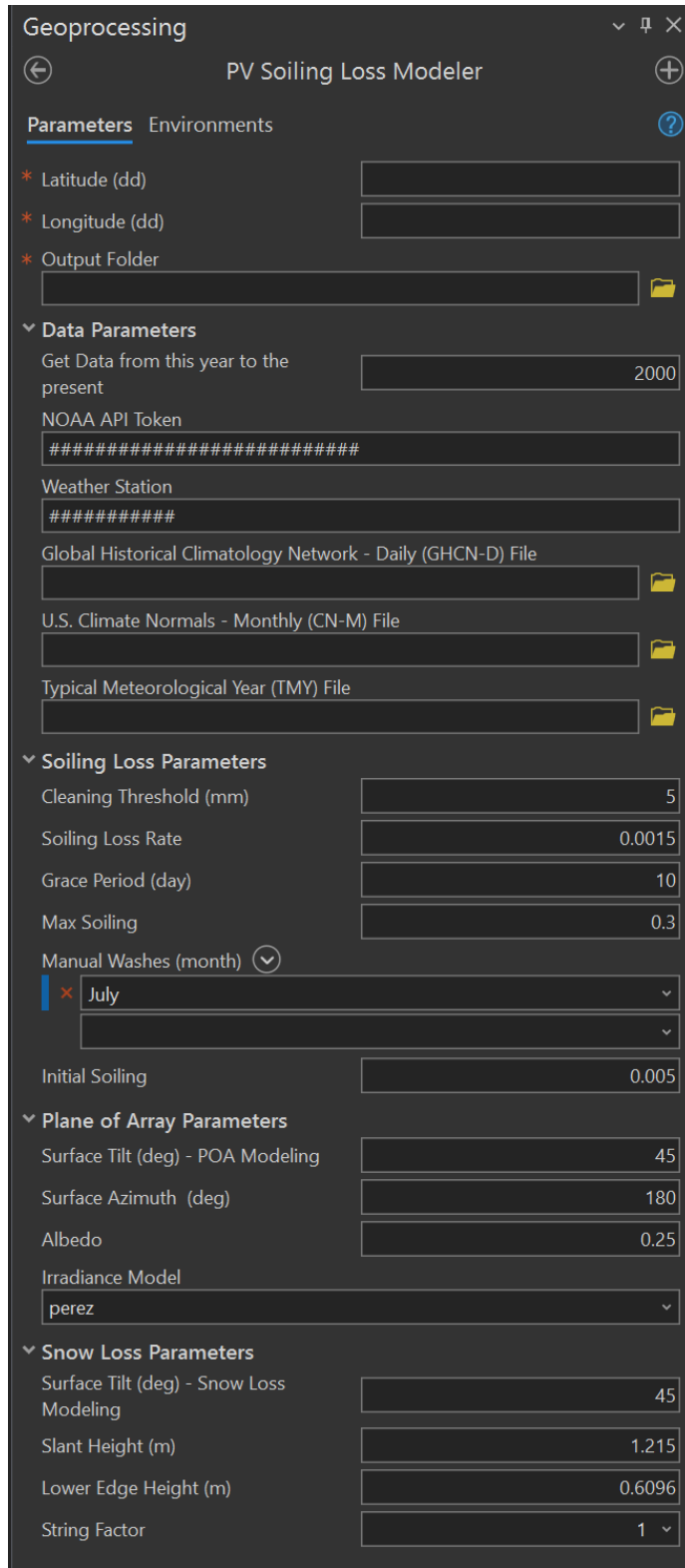
Figure 2. PV Soiling Loss Modeler tool as seen in ArcGIS Pro software with sample parameter values. Input parameters are categorized in collapsible groups corresponding to their tool analysis step.

User Testing

Once the PVSLM tool was operational, the tool was set up for use by 4 analysts. User testing over the course of two weeks resulted in feedback focusing on modifications to tool parameters and simplifying the tool options. Users requested the option to have an end date for GHCNd weather data searches, as opposed to the current functionality where the user provides the start year, and the tool retrieves data from the start year to the present. The option to skip the weather station search/download process by using available tool parameters to point to the .csv files for weather station and TMY data was confusing to users. Users also recommended setting a minimum monthly soiling loss to 0.5% and maximum monthly soiling loss to 50%.

Soiling Loss Model Results

Soiling loss modeling was completed for six sample sites across the conterminous United States. Table 2 shows the soiling loss modeling results, as well as the on-site ground station soiling loss measurements for those six test locations. The latitude and longitude of the SMS at each sample site was used for the input coordinates of the PVSLM modeling. Each site was processed with the PVSLM tool and successfully produced soiling loss results. Monthly and annual soiling losses for each site are compiled in Table 2. Average PVSLM processing time ranged from about 1-5 minutes, with the greatest time discrepancy dependent on the number of weather stations assessed before identifying the closet weather station with complete data availability.

Site	Measurement Method	Soiling Losses												
		Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Annual
Site A - TX	PVSLM Kimber Model (%)	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.66	0.50	0.50	**0.51**
	PVSLM Townsend Model (%)	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	**0.50**
	PVSLM Combined Model (%)	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.66	0.50	0.50	**0.51**
	On-Site SMS (%)	0.58	0.51	0.52	0.57	0.61	0.57	0.56	0.66	0.54	0.63	0.57	0.54	**0.57**
Site B - AZ	PVSLM Kimber Model (%)	3.09	3.34	2.99	5.22	7.89	2.36	1.29	0.98	1.27	3.42	5.25	4.06	**3.43**
	PVSLM Townsend Model (%)	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	**0.50**
	PVSLM Combined Model (%)	3.09	3.34	2.99	5.22	7.89	2.36	1.29	0.98	1.27	3.42	5.25	4.06	**3.43**
	On-Site SMS (%)	5.20	4.10	3.40	3.80	3.50	3.70	0.00	0.70	0.30	2.50	4.40	5.50	**3.09**
Site C - UT	PVSLM Kimber Model (%)	2.98	2.48	1.92	1.81	1.99	3.55	4.53	3.81	3.97	3.88	3.54	3.30	**3.15**
	PVSLM Townsend Model (%)	50.00	50.00	50.00	29.40	6.72	0.50	0.50	0.50	0.50	0.50	35.41	50.00	**22.84**
	PVSLM Combined Model (%)	50.00	50.00	50.00	29.40	6.72	3.55	4.53	3.81	3.97	3.88	35.41	50.00	**24.27**
	On-Site SMS (%)	5.90	6.50	2.70	1.80	1.10	1.50	2.30	1.80	1.80	1.90	3.90	7.80	**3.25**
Site D - ME	PVSLM Kimber Model (%)	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	**0.50**
	PVSLM Townsend Model (%)	50.00	50.00	50.00	50.00	29.06	0.50	0.50	0.50	0.50	5.53	50.00	50.00	**28.05**
	PVSLM Combined Model (%)	50.00	50.00	50.00	50.00	29.06	0.50	0.50	0.50	0.50	5.53	50.00	50.00	**28.05**
	On-Site SMS (%)	23.30	14.50	7.00	2.10	0.50	0.50	0.50	0.50	0.50	1.00	1.90	16.50	**5.73**
Site E - AL	PVSLM Kimber Model (%)	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.98	1.06	0.50	**0.59**
	PVSLM Townsend Model (%)	4.34	4.47	1.59	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	**1.24**
	PVSLM Combined Model (%)	4.34	4.47	1.59	0.50	0.50	0.50	0.50	0.50	0.50	0.98	1.06	0.50	**1.33**
	On-Site SMS (%)	5.5	5.6	2.7	1.3	0.9	0.1	0.2	0.5	1.8	2.4	3.9	5.9	**2.57**
Site F - VA	PVSLM Kimber Model (%)	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.83	0.50	**0.53**
	PVSLM Townsend Model (%)	50.00	39.50	14.41	2.95	0.50	0.50	0.50	0.50	0.50	0.50	0.50	16.06	**10.54**
	PVSLM Combined Model (%)	50.00	39.50	14.41	2.95	0.50	0.50	0.50	0.50	0.50	0.50	0.83	16.06	**10.56**
	On-Site SMS (%)	6.10	5.60	3.80	2.00	0.70	0.50	0.10	0.90	0.80	2.20	3.60	6.00	**2.69**

Table 2. Soiling loss results at six test sites for PVSLM Kimber, Townsend, Combined modeling losses, and soiling losses measured at on-site SMS facilities.

Discussion

The objective of this project was to create a Python script tool for ArcGIS Pro, PVSLM, that would produce PV soiling loss estimates for a proposed location based on publicly available weather station and satellite data. The main components of this task were to identify the closest weather station with all the required data for soiling loss modeling, download that data, download satellite data, produce Kimber Model soiling loss estimates, produce Townsend Model snow loss estimates, and combine the soiling loss estimates to get total monthly and annual loss estimates. The PVSLM tool has met all objectives and was successful in completing the soiling loss analysis for all test sites.

The results of the soiling loss modeling in comparison to the on-site SMS soiling loss measurements will require additional work to improve confidence in modeling outcomes. Kimber soiling losses for the six sample sites are mostly within a percentage point of the measured losses. These are acceptable results for soiling loss modeling. However, Townsend model snow losses are consistently higher than measured losses, exceeding differences of more than 40% higher modeled losses. The high values for the snow modeling will require further investigation to determine the cause. Data inputs and processes for the Townsend model were reassessed to identify any potential issues but no clear evidence of problems in methodology was identified. Some of the model parameters may need to be revised to better match measurement conditions for the SMS. It is yet to be determined whether the inconsistent snow losses result from issues with methodology in building the PVSLM tool, limitations of modeling methods within the *pvlib* functions, or the Townsend snow loss model itself. Given the challenges of snow loss modeling in general, it is no surprise that this portion of the modeling is the least reliable.

A limitation of the weather station selection and data retrieval process is that the characteristics of the terrain are not considered, only distance and availability of data. The tool does not account for if the weather station is in a dramatically different terrain or landcover type than the input coordinates. Differences in landcover can result in dramatically different soiling phenomena, like the difference between a site adjacent to industrial facilities and one surrounded by woodland. In addition, landcover can be a consideration when choosing the best satellite data tile for a place of interest. If the 4 km x 4 km data tile where the input coordinates are located has a water body covering a significant portion of the tile, it can yield meteorological measurements inconsistent with the terrain of the site. In those cases, it may be more appropriate to select an adjacent satellite data tile. Incorporating a landcover dataset as another reference layer for analysis could improve the weather station and satellite data selection process.

Cost/Benefit Analysis

To implement the PVSLM tool, certain expenses are required. Software necessary to operate PVSLM includes ArcGIS Pro Basic, about $765 annually. The computer hardware required to work with PVSLM would likely already be available to the resource analyst as part of their normal work functions.

Despite the expense of software and hardware, PVSLM will make up for the initial investment with greater efficiency of data retrieval and analysis. Bringing several work processes into a single interface will offer the end-user greater access to soiling loss datasets and the ability to easily modify outputs. Further cost benefit analysis, shown in Table 3, provides insight into the long-term outcome of PVSLM implementation. Based on analysis of the current processes as compared to existing workflows, the proposed GISystem design will result in an approximate efficiency improvement of 2.5%, or roughly 1 hour per work week.

Year	Cost ($)	Benefit ($)	Net ($)
1	765	12,000	11,235
2	765	24,000	23,235
3	765	36,000	35,235
4	765	48,000	47,235
5	765	48,000	47,235
6	765	48,000	47,235
7	765	48,000	47,235
8	765	48,000	47,235
9	765	48,000	47,235
10	765	48,000	47,235

Table 3. Cost/Benefit analysis assuming standard hourly expense for 4 end-users improving from 0.5 hours per week in the first year, to 1 hours per week in the third year.

Future Work

Ongoing development of the PVSLM tool will involve refinements to the snow loss modeling, expanded options for data sources, and updates to weather station selection. The most urgent issue to resolve is the high Townsend snow loss modeling results. This will require a careful review of all data inputs and processes informing the loss results. I will be at a conference in Salt Lake City next month where some of the individuals from NREL who are involved in the development of *pvlib* will be in attendance. This will be a good opportunity to discuss the Townsend snow loss model and get some expert insight. I would also like to incorporate the option to choose between additional soiling loss models in PVSLM. The Humboldt State University (HSU) soiling loss model is an alternative to the Kimber model but requires particulate matter data in addition to precipitation data. While particulate matter data is not generally available in public satellite data, it is sometimes available from private sources like SolarAnywhere or SolarGIS. The ability to request data from the aforementioned private data sources already exists within *pvlib* but requires a paid subscription to those data services. Incorporating analysis of a landcover dataset could improve the results of the weather station data selection process by identifying stations that are a closer match to the input coordinates. Inclusion of the landcover element in the weather station selection process would require additional research to understand the best practices involved in this type of analysis.

Conclusion

        The goal of this project is to create a Python script tool for ArcGIS Pro that generates monthly soiling loss estimates for a proposed PV site using publicly available data. The data produced by the tool is used for populating variables in the estimation of the long-term energy generation of a PV site design. Existing methods for compiling and analyzing the necessary data for estimating soiling losses are tedious and introduce opportunities for user error. The PVSLM tool simplifies the process by automatically analyzing solar resource and meteorological data to produce soiling loss estimates based on user provided site coordinates and a selection of modeling options. While Kimber soiling loss results are consistent with measured values, Townsend snow loss estimates are unexpectedly high. With further refinement of snow loss modeling with the PVSLM tool, solar resource analysts will be able to quickly produce soiling loss estimates for proposed PV sites using the modeling parameters they determine to be most appropriate for their site.

References

Ali, S., Taweekun, J., Techato, K., Waewsak, J., & Gyawali, S. (2019). GIS based site suitability assessment for wind and solar farms in Songkhla, Thailand. *Renewable Energy*, *132*, 1360–1372. https://doi.org/10.1016/j.renene.2018.09.035

Arguez, A., Durre, I., Applequist, S., Squires, M., Vose, R., Yin, X., & Bilotta, R. (2010). *U. S. Climate normals product suite (1981-2010)*. NOAA National Climatic Data Center. https://doi.org/10.7289/V5PN93JP

Baldus-Jeursen, C., Petsiuk, A. L., Rheault, S.-A., Pelland, S., Côté, A., Poissant, Y., & Pearce, J. M. (2023). Snow losses for photovoltaic systems: Validating the Marion and Townsend models. *IEEE Journal of Photovoltaics*, *13*(4), 610–620. https://doi.org/10.1109/JPHOTOV.2023.3264644

Bessa, J. G., Micheli, L., Almonacid, F., & Fernández, E. F. (2021). Monitoring photovoltaic soiling: Assessment, challenges, and perspectives of current and potential strategies. *iScience*, *24*(3), 102165. https://doi.org/10.1016/j.isci.2021.102165

Borah, P., Micheli, L., & Sarmah, N. (2023). Analysis of soiling loss in photovoltaic modules: A review of the impact of atmospheric parameters, soil properties, and mitigation approaches. *Sustainability*, *15*(24), 16669. https://doi.org/10.3390/su152416669

F. Holmgren, W., W. Hansen, C., & A. Mikofski, M. (2018). Pvlib python: A python package for modeling solar energy systems. *Journal of Open Source Software*, *3*(29), 884. https://doi.org/10.21105/joss.00884

Georgiou, A., & Skarlatos, D. (2016). Optimal site selection for siting a solar park using multi-criteria decision analysis and geographical information systems. *Geoscientific Instrumentation, Methods and Data Systems*, *5*(2), 321–332. https://doi.org/10.5194/gi-5-321-2016

Getto, G., & Moore, C. (2017). Mapping personas: Designing UX relationships for an online coastal atlas. *Computers and Composition*, *43*, 15–34. https://doi.org/10.1016/j.compcom.2016.11.008

Heidari, N., Gwamuri, J., Townsend, T., & Pearce, J. M. (2015). Impact of snow and ground interference on photovoltaic electric system performance. *IEEE Journal of Photovoltaics*, *5*(6), 1680–1685. https://doi.org/10.1109/JPHOTOV.2015.2466448

Li, X., Mauzerall, D. L., & Bergin, M. H. (2020). Global reduction of solar power generation efficiency due to aerosols and panel soiling. *Nature Sustainability*, *3*(9), 720–727. https://doi.org/10.1038/s41893-020-0553-2

Menne, M. J., Durre, I., Korzeniewski, B., McNeill, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). *Global historical climatology network—Daily (Ghcn-daily), version 3*. NOAA National Centers for Environmental Information. https://doi.org/10.7289/V5D21VHZ

Micheli, L., & Muller, M. (2017). An investigation of the key parameters for predicting PV soiling losses. *Progress in Photovoltaics: Research and Applications*, *25*(4), 291–307. https://doi.org/10.1002/pip.2860

Micheli, L., Smestad, G. P., Bessa, J. G., Muller, M., Fernandez, E. F., & Almonacid, F. (2022). Tracking soiling losses: Assessment, uncertainty, and challenges in mapping. *IEEE Journal of Photovoltaics*, *12*(1), 114–118. https://doi.org/10.1109/JPHOTOV.2021.3113858

Muller, M., & Rashed, F. (2023). Considering the variability of soiling in long-term pv performance forecasting. *IEEE Journal of Photovoltaics*, *13*(6), 825–829. https://doi.org/10.1109/JPHOTOV.2023.3300369

National Oceanic and Atmospheric Administration. (n.d.). *GHCND-Stations*. https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt

National Renewable Energy Laboratory. (n.d.). *National Solar Radiation Database* [Typical Meteorological Year (TMY)]. https://nsrdb.nrel.gov/data-sets/tmy/

Rosson, M. B., & Carroll, J. M. (2002). Scenario-Based Design. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. 1032–1050). Lawrence Erlbaum Associates.

Sandia National Laboratories. (n.d.). *Plane of array (Poa) irradiance*. PV Performance Modeling Collaborative (PVPMC). Retrieved February 6, 2024, from https://pvpmc.sandia.gov/modeling-guide/1-weather-design-inputs/plane-of-array-poa-irradiance/

Solar Power Europe. (2023). *Global Market Outlook For Solar Power 2023—2027*. https://www.solarpowereurope.org/insights/market-outlooks/global-market-outlook-for-solar-power-2023-2027-1

United States Department of Energy, Office of Energy Efficiency and Renewable Energy. (2022, May). *Modeling of Photovoltaic Systems: Basic Challenges and DOE-Funded Tools*. https://www.energy.gov/sites/default/files/2022-05/Modeling%20of%20Photovoltaic%20Systems%20White%20Paper.pdf