SOFTWARE METAPAPER

CERF – A Geospatial Model for Assessing Future Energy Production Technology Expansion Feasibility

Chris R. Vernon[1], Nino Zuljevic[1], Jennie S. Rice[1], Timothy E. Seiple[1], Michael C. W. Kintner-Meyer[1], Nathalie Voisin[1], Ian P. Kraucunas[1], Jin Chunlian[1], Jarrod Olson[1], Laurel Schmidt[1], Scott L. Morris[1] and Pralit Patel[2]

[1] Pacific Northwest National Laboratory, Richland, WA 99352, US

[2] Joint Global Change Research Institute, Pacific Northwest National laboratory, College Park, Maryland, US

Corresponding author: Chris Vernon (chris.vernon@pnnl.gov)

The Capacity Expansion Regional Feasibility (CERF) model is an open-source geospatial model, written in Python and C++, that is designed to determine the on-the-ground feasibility of achieving a projected energy technology expansion plan. Integrated human-Earth systems models and grid expansion models typically do not have sufficient spatial, temporal, or process-level resolution to account for technology-specific siting considerations—for example, the value or costs of connecting a new power plant to the electric grid at a particular location or whether there is sufficient cooling water to support the installation of thermal power plants in a certain region. CERF was developed to specifically examine where power plant locations can be feasibly sited when considering high spatial resolution siting suitability data as well as the net locational costs (i.e., considering both net operating value and interconnection costs), at a spatial resolution of 1 km^2. The outputs from CERF can provide insight into factors that influence energy system resilience under a variety of future scenarios can be used to refine model-based projections and be useful for capacity expansion planning exercises. CERF is open-source and publicly available via GitHub.

(1) Overview
Model Description

The Capacity Expansion Regional Feasibility (CERF) model is written in Python and C++ and was built to determine the on-the-ground feasibility of achieving energy system expansion plans generated by integrated human-Earth systems models or regional capacity expansion models. Energy expansion projections from such models are typically limited to fairly large regions (e.g., balancing authorities or control regions) and do not account for on-the-ground barriers such as protected lands, potentially hazardous areas, highly populated areas, water availability, proximity to Class 1 airsheds, and other siting constraints that may impede the ability to achieve a planned expansion. CERF was developed to evaluate the feasibility of siting a mix of energy technologies over the contiguous United States of America (CONUS) at a resolution of 1 km^2 while considering current and future socioeconomic conditions, energy demands, land use, environmental regulations, and water availability. The information gleaned from determining the feasibility of technology expansion plans can be used to provide insight into factors that may create or exacerbate vulnerabilities related to energy system resilience.

CERF is unique in that it determines feasible siting locations using a combination of on-the-ground suitability constraints (e.g., protected lands) with simulated economic competition between energy technologies using an algorithm that minimizes net locational cost (NLC) to choose specific siting locations within suitable areas.[1] The NLCs are calculated for each technology and are influenced by the distance to existing transmission infrastructure, technology-specific marginal operating costs, and technology- and location-specific marginal energy values. In effect, the algorithm posits the existence of a regional planner who determines the costs and benefits of having new generation in different locations and sites power plants in order from lowest to highest NLC.

Technologies that share the same suitable areas compete for siting locations based on the locational economics.

CERF was designed to receive input from other models regarding the amount and type of power plants to be sited in a particular geographic area (e.g., in each balancing authority) as well as other variables such as water availability and locational marginal energy prices. CERF also "inherits" the assumptions of these models; for example, water availability constraints can be static or based on the output from an Earth system model using a specific scenario of future Earth system changes and corresponding time-evolving water availability.

CERF can currently site 17 energy technologies.[2] Each technology has spatial suitability rasters and technology-specific parameters that can be set by the user. A comprehensive literature review[3] was conducted to determine which suitability constraints should be applied for each technology. A suitability raster containing suitable and unsuitable areas that are common to all technologies is provided with CERF.[4] Technology-specific suitability data are also provided.[5] **Figure 1** demonstrates the combined suitable area that is common to all technologies as mapped onto CERF's 1 km² grid over the CONUS. This common suitability raster is then combined with technology-specific suitability rasters (**Figure 2**; example for nuclear power) to yield grid cells that are suitable for siting. Suitability constraints can be added by the user to specific technologies or to those that are common to all technologies, depending on research needs. The NLC of each grid cell is then used to determine economically feasible locations for siting in grid cells that are suitable for siting.[6]

Implementation and architecture

CERF is designed as a Python package that can be called in script after the user has installed the package on a Windows 7 OS. The CERF model is parameterized using an input configuration file, a required set of input XML files, and spatial data files (detailed in the Configuration section below). Ensure that local paths for the spatial data files in the technology_suitabilitymask_paths.xml have been set before a run. After these files have been prepared, CERF can be called from other Python scripts once installed as demonstrated in the example.py script[7] which is included with the package. This can be conducted as follows:

```python
from cerf.model import Cerf
if __name__ == "__main__":
    # <pth> is the full path to the configuration file
    ini = "<pth>/config.ini"

    # instantiate CERF
    m = Cerf(ini)

    # run CERF
    m.execute()
```

After `m.execute()` is called, CERF will process each state and technology sequentially to determine the geographic coordinates, net locational costs, interconnection costs, and net operational value of each power plant. This data is saved to an XML file that is ingested by CERF's outputs module to create summary graphics and tabular data as described in the outputs section below. A log file is also created during runtime which contains any pertinent runtime information.

The SAGA GIS [1] C++ API was used to provide spatial functionality related to ensuring consistent geographic projection use, calculating interconnection distance, and extracting masks of raster data. We extended SAGA GIS by developing a custom C++ module to calculate a local sum of a stacked group of rasters. This allows us to determine suitability for each 1 km² grid cell while maintaining the spatial integrity of the data (**Figure 3**). The C++ component of CERF was compiled for use in Windows 7 only and is called and parameterized through CERF's Python model.py wrapper.
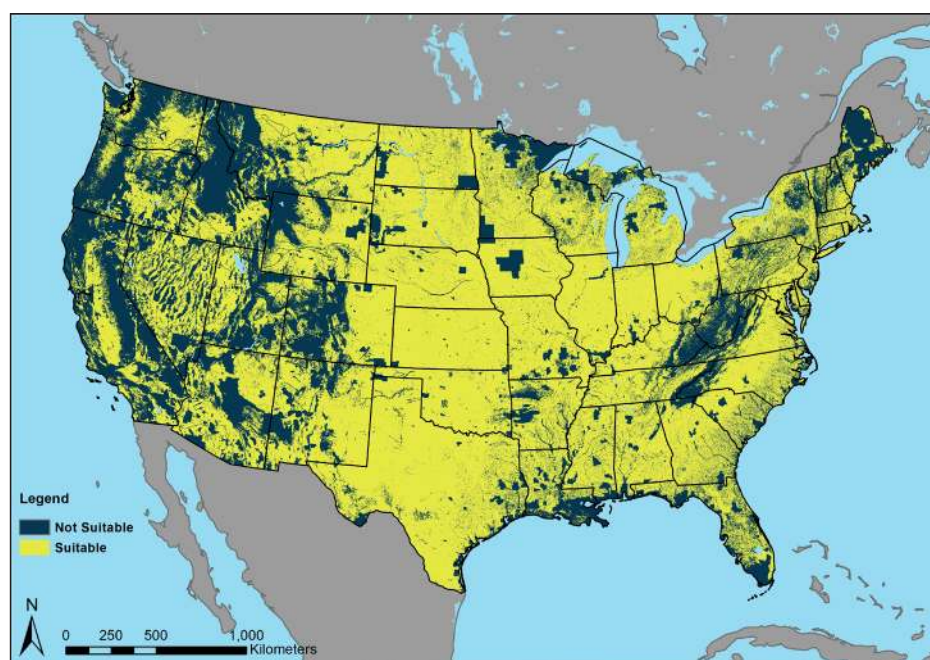

Figure 1: Barriers to siting that are common to all thermo-electric technologies. Unsuitable area where no siting can occur is represented by dark blue; suitable area where siting can potentially occur is represented by yellow.

Figure 2: Barriers to siting that are common to all thermo-electric technologies combined with technology-specific barriers for nuclear plants. Unsuitable area where no siting can occur is represented by dark blue; suitable area where siting can potentially occur is represented by yellow.
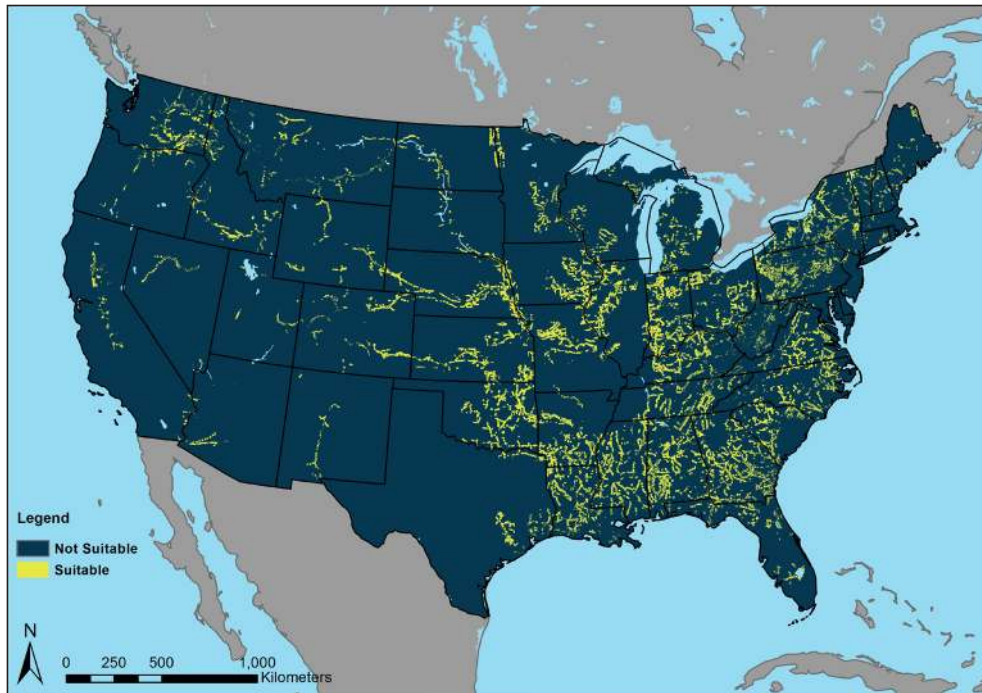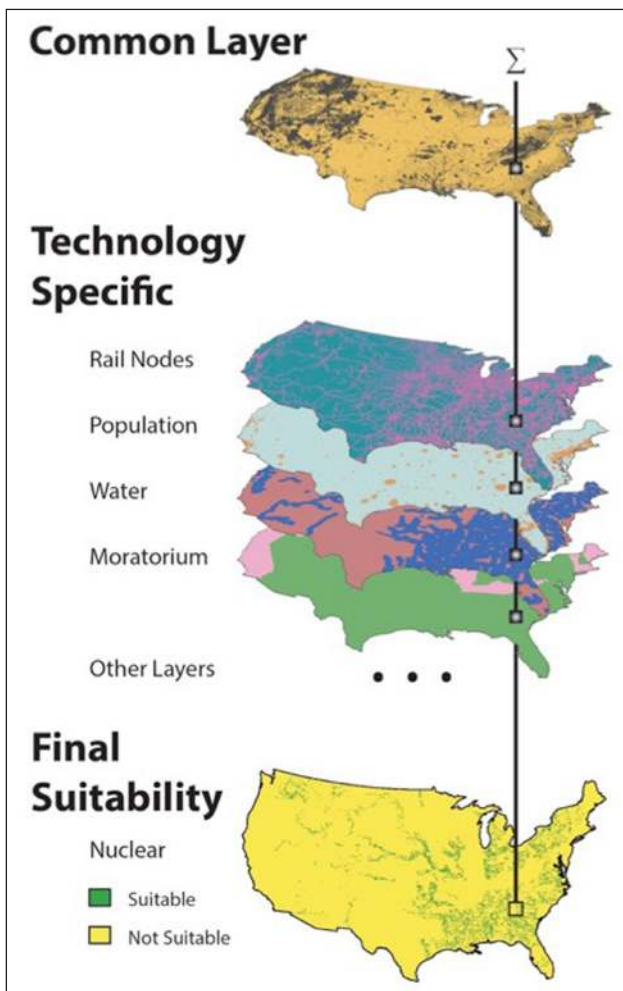

Figure 3: Custom SAGA function to compute grid cell suitability.

Example input datasets

CERF is designed to test the suitability and feasibility of "proposed" capacity expansion plans; therefore, it requires a range of input data both in terms of those plans as well as how the suitability and feasibility criteria might change over time. The example input dataset provided with CERF was built to demonstrate the uniqueness and functionality of CERF. A modeled energy technology expansion plan and technology-specific assumptions, in the context of a specific future scenario of environmental and sociological change (namely, the Representative Concentration Pathway, or RCP, 8.5 [2]), were generated by the US 50-state version of the Global Change Assessment Model (GCAM) [3–4]. Gridded water availability data was derived from a large-scale Water Management (WM) model [5–6] driven by an Earth system model using the same RCP 8.5 scenario assumptions. Locational marginal prices were provided by the PROMOD IV production cost model [7].

Publicly available datasets for transmission lines and gas pipelines were used in this demonstration as provided by the U.S. Energy Information Administration (EIA) and the Homeland Infrastructure Foundation-Level Data (HIFLD) [8–9]. These datasets influence the interconnection cost for each site and play a role in determining NLC; therefore, we recommend obtaining more detailed representations of both transmission lines and gas pipelines if they are available. The solar (CSP) suitability raster simply adopted the common suitability criteria that applies to all thermo-electric technologies and should be enhanced to capture the idiosyncrasies of siting CSP.

Figure 4 demonstrates illustrative results for nuclear siting feasibility from the example data for the Eastern Interconnection in reference to a combination of
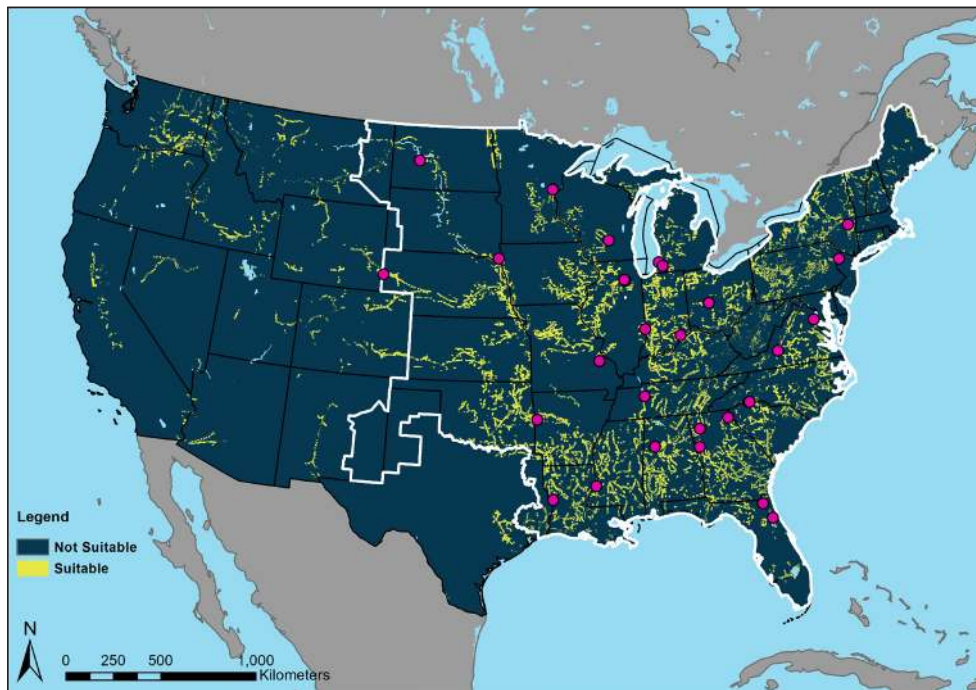
Figure 4: Nuclear siting feasibility in the Eastern Interconnection (white polygon) in reference to suitability. Unsuitable area where no siting can occur is represented by dark blue; suitable area where siting can potentially occur is represented by yellow, and nuclear sites feasible based on econometrics and technology-specific requirements for the Eastern Interconnection are in magenta.

suitability constraints that are common to all technologies and those that are specific to nuclear.

Configuration
CERF's setup occurs in a configuration file and a set of XML files that are used to define constants, planned expansion, locational marginal pricing and carbon factors per utility zone, state zones, technology specific information, and the location of suitability rasters per technology. Examples of all input files can be found in CERF's example directory.[8]

Figure 5 illustrates CERF's workflow once the input files have been prepared.

The following describes the requirements and format of each input.

Constants
This file contains constant assumptions that are applied in CERF. The file name should be constants.xml. **Table 1** describes the required parameters for the constants.xml file.

Expansion Plan
This file contains the expansion plan expected capacity for each technology per state. **Table 2** describes the required parameters for the expansionplan.xml file.

Utility Zone Data
This file contains specifics related to each utility zone. **Table 3** describes the required parameters for the powerzones.xml file.

States
This file contains information for the states referenced in the expansion plan. **Table 4** describes the required parameters for the states.xml file.
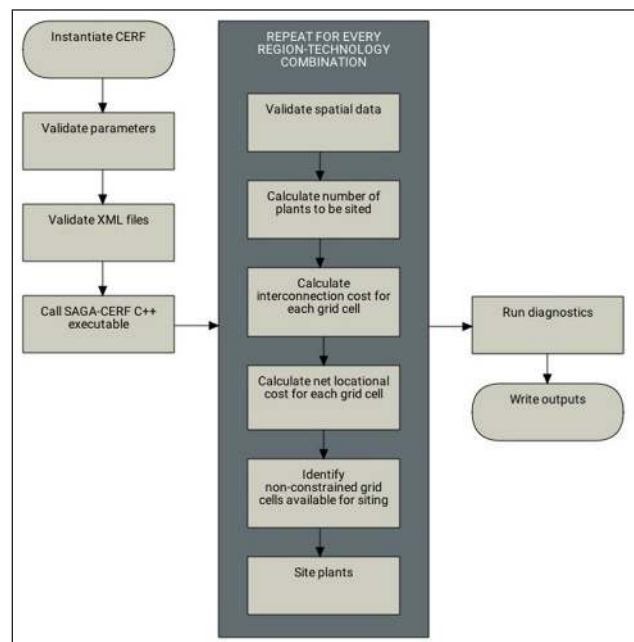

Figure 5: CERF's workflow.

Technologies
This file contains information for each technology. This information is usually derived from the technology assumptions utilized in the model providing the technology expansion plans. **Table 5** describes the required parameters for the technologies.xml file.

Suitability Data
This file contains the full path reference to each technologies suitability raster. **Table 6** describes the required parameters for the technology_suitabilitymask_paths.xml file.

Table 1: Parameters and descriptions for CERF's constants.xml input file.

Name	Description
discount_rate	Float from 0.0 to 1.0. The time value of money in real terms.
carbon_tax	Float. The fee imposed on the burning of carbon-based fuels in $/ton.
carbon_tax_escalation	The annual escalation rate for carbon tax. From 0.0 to 1.0.
tx_lifetime	Integer in years of the expected technology plant lifetime.
interconnection_cost_gas	Float for the capital cost of gas interconnection in $K/km.

Table 2: Parameters and descriptions for CERF's expansionplan.xml input file.

Name	Description
zoneid	Corresponding state id as referenced in the states.xml file.
techid	Corresponding technology id as referenced in the technologies.xml file.
<value>	The expected capacity in MW.

Table 3: Parameters and descriptions for CERF's powerzones.xml input file.

Name	Description
zone	Integer ID of the utility zone.
shapeid	Integer ID of the spatial reference referred to in states.xml.
name	The abbreviated name of the power zone as a string.
lmp	Float of the annualized locational marginal price for the target zone $/MWh.
description	String description of the utility zone.
<lmps><cf>	Float capacity factor for each LMP percentile.

Table 4: Parameters and descriptions for CERF's states.xml input file.

Name	Description
id	Unique integer ID of the state.
shapeid	Unique integer ID of the feature in the input shapefile.
<value>	String name of the state.

Configuration file

This file is the main configuration used to run CERF. **Table 7** describes the required parameters for the configuration file.

Preparing Suitability Rasters

Rasters for spatial suitability are required to conform to the format referenced in **Table 8**. Suitability rasters can be prepared using any GIS.

CERF outputs

CERF outputs are generated to help the user to interpret, summarize, and visualize CERF's results. The main categories of CERF's outputs can be analyzed at the site, technology, state, or project level for: siting feasibility, interconnection cost, net locational cost, and net operating value. A geographic reference (x, y coordinates) is also included for each site so the user may plot these in a geographic information system to conduct locational analysis.

The outputs may be generated using CERF's *outputs* module. Usage is as follows to create a heatmap showing all states and technologies relative to one another for interconnection cost.

```python
from cerf.outputs import Outputs
if __name__ == "__main__":
    # <pth> is the full path to the configuration file;
    ini = "<pth>/config.ini"
    # instantiate CERF outputs
    res = Outputs(ini)
    # create a heatmap of interconnection cost
    res.heatmap(metric = 'ic')
```

The Python help command may be used with any method in the class to view parameterization possibilities and a description of the method. Currently supported methods are described in **Table 9**.

Installation

CERF can be installed as follows:

1. CERF's GitHub repository uses the Git Large File Storage[9] (LFS) extension. Run the following command before cloning if you do not already have Git LFS installed: `git lfs install`.

Table 5: Parameters and descriptions for CERF's technologies.xml input file.

Name	Description
id	Unique integer ID for each technology.
unit_size	Integer value for power plant unit size in MW.
capacity_factor	Float value from 0.0 to 1.0 for the average annual power generated divided by the potential output if the plant operated at its rated capacity for a year.
variable_om	Float value representing the variable operation and maintenance costs of yearly capacity use in $/MWh.
variable_cost_escalation_rate	Float from –1.0 to 1.0 for the escalation rate of variable costs.
heat_rate	Float for the amount of energy used by a power plant to generate one kilowatt-hour of electricity in Btu/kWh.
fuel_price	Float for cost of fuel per unit in ($/MBtu)($10^6$ Btu/MBtu)(10^3 kWh/MWh).
fuel_price_escalation	Float from –1.0 to 1.0 for fuel price escalation.
fuel_co2_content	Float for CO_2 content of the fuel and the heat rate of the technology in Tons/MWh.
interconnection_cost_per_km	Float for the capital cost of interconnection in $K/km.
full_name	Full technology name.
lifetime	Integer for asset life time in years.
category	Type of fuel (e.g., gas, coal, etc.).
fuel_index	String reference for fuel index type.
variable_om_2005	Float value for variable operation and maintenance costs of capacity use in $/MWh.
siting_buffer	Buffer to place around a plant once sited.
carbon_capture_rate	Float for the rate from 0 to 1 of carbon capture.

Table 6: Parameters and descriptions for CERF's technology_suitabilitymask_paths.xml input file.

Name	Description
techid	Integer ID of the corresponding technology.
<value>	Full path with file name and extension to the input SAGA grid raster.

Table 7: Parameters and descriptions for CERF's configuration file.

Name	Description
exe_path	Full path with file name and extension of saga_cerf.exe.
xml_path	Full path to the directory containing the input XML files.
out_path	Full path to the directory where the output files will be written.
yr	Integer target four digit year (e.g., 2005).
primary_zone	Full path with file name and extension to the input states raster.
utility_zones	Full path with file name and extension to the input utility zones.
common_exclusion	Full path with file name and extension to the input suitability raster common to all technologies.
transmission_230kv	Full path with file name and extension to the input raster for transmission lines > = 230 Kv.
transmission_345kv	Full path with file name and extension to the input raster for transmission lines > = 345 Kv.
transmission_500kv	Full path with file name and extension to the input raster for transmission lines > = 500 Kv.
gasline_16in	Full path with file name and extension to the input raster for gas pipelines that are > = 16 inches.
buffer	Integer buffer in grid cells to place around each site.
distance_method	Integer from 0 to 2 to select type of distance method used when calculating interconnection costs [0: Chessboard, 1: Manhattan, 2: Euclidean distance].
direction_method	Integer from 0 to 3 to select the directional pattern used when siting a technology in a region [0: left, right, top, bottom; 1: RLBT; 2: LRBT; 3: RLTB].

Table 8: Requirements for spatial suitability rasters.

Attribute	Description
Number of columns, Number of rows	4693, 2999
Coordinate system	PROJCS["USA_Contiguous_Albers_Equal_Area_Conic", GEOGCS["GCS_North_American_1983", DATUM["North_American_Datum_1983", SPHEROID["GRS_1980", 6378137.0, 298.257222101]], PRIMEM["Greenwich", 0.0], UNIT["Degree", 0.0174532925199433]], PROJECTION["Albers_Conic_Equal_Area"], PARAMETER["false_easting", 0.0], PARAMETER["false_northing", 0.0], PARAMETER["longitude_of_center", –96.0], PARAMETER["standard_parallel_1", 29.5], PARAMETER["standard_parallel_2", 45.5], PARAMETER["latitude_of_center", 37.5], UNIT["Meters", 1.0]]
Origin	(–2405552.835500000044703, 1609934.799499999964610)
Pixel Size	(1000, –1000)
Upper Left	(–2405552.836, 1609934.799)
Lower Left	(–2405552.836, –1389065.201)
Upper Right	(2287447.164, 1609934.799)
Lower Right	(2287447.164, –1389065.201)
Center	(–59052.836, 110434.799)
Type	Byte

Table 9: Method name and description of for output options in the *Outputs* class.

Method	Description
heatmap	Creates a heatmap figure (**Figure 6**) showing all states and technologies relative to one another for a target metric. Can produce a heatmap for interconnection cost (ic), net locational cost (nlc), or net operating value (nov).
export_planned_vs_sited	Exports a CSV file of sites for each state and technology that were planned but not feasible to site. Returns a Pandas Data Frame.
plot_planned_vs_sited	Creates a bar plot (**Figure 7**) of either planned versus sited number of site or capacity for a target state.
eval_metric_per_tech	Creates a jittered strip plot (**Figure 8**) of sites per technology for a target state. Can produce a plot for interconnection cost (ic), net locational cost (nlc), or net operating value (nov).

2. Clone CERF into your desired location: `git clone https://github.com/IMMM-SFA/cerf.git`.
3. Install the Python package setuptools[10] if the package is not already on your machine.
4. From the directory you cloned CERF into and your setup.py file exists run `python setup.py install` which will install CERF as a Python package on your machine and install of the required Python dependencies.

Quality control

Due to the complex nature of the technology expansion plan data, CERF was built to provide detailed feedback to the user upon exception when validating the configuration file and input XML files. Parameters have set limits and type requirements which are strictly enforced and the code is designed to help users navigate this as easily as possible. CERF also builds a timestamped log file and output directory structure during runtime.

A validation of CERF's suitability constraints that are common to all technologies was conducted for the CONUS using existing operable power plant locations as provided by the EIA [8] (**Table 10**). The technology that was listed as the primary producer at each site was declared as the technology for that location. Locations from EIA were represented by geographic coordinates for each site and do not take into account the entire footprint of the site. For this reason, this validation quantifies how many times each plant's geographic coordinate falls directly on an area that was suitable as well as the times the plant fell within 5 km of an area that was suitable to account for the potential variability in the site's actual location. This validation also assumes that the sites from EIA are correctly georeferenced. Technologies were lumped into their EIA technology name (i.e., coal CC, coal IGCC, conventional coal

Table 10: Validation results for CERF's spatial suitability for siting as compared to EIA current power plant locations.

Technology	EIA sites	Able to be sited by plant coordinate	Percent sited by plant coordinate	Able to be sited within 5 km	Percent sited within 5 km
Nuclear	54	44	81.5%	53	98.1%
Coal	332	276	83.1%	328	98.8%
Natural Gas	1648	1382	83.9%	1602	97.2%
Biomass	725	590	81.4%	712	98.2%
Wind	993	952	95.9%	980	98.7%
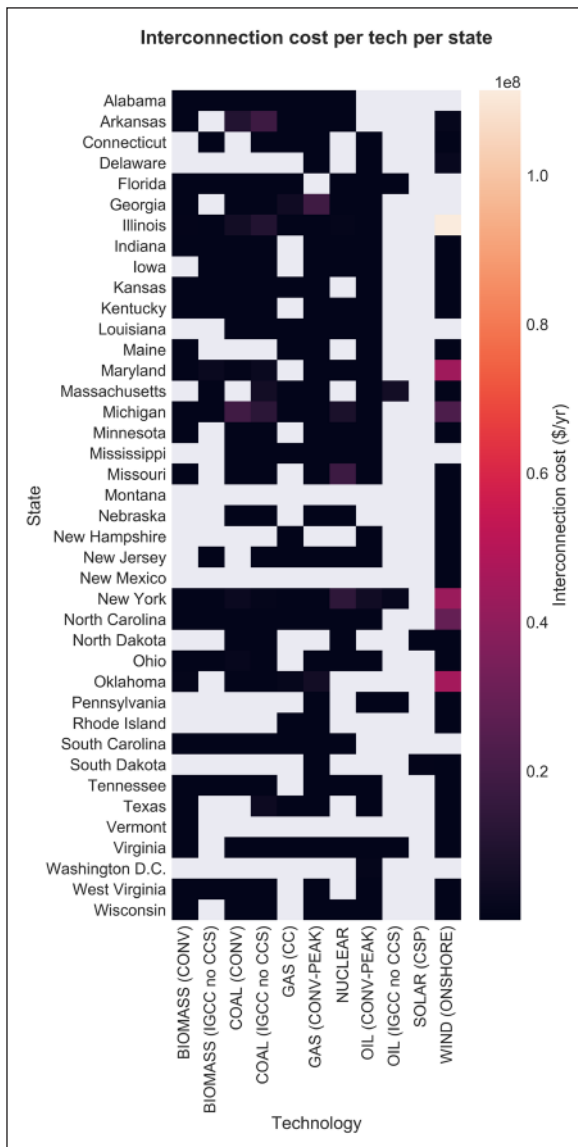
Figure 6: A heatmap showing comparative interconnection cost per technology per state from CERF's example data. The color bar ranges from 0 \$/year to 100 million \$/year. Grey in the plot represents no sites for the state-technology combination.

all lumped into a category named "coal"). The validation for wind also utilized the EIA power plant locations but compared them to the technology specific suitability layer for wind since the constraints that are common to all other technologies do not always apply to wind.

(2) Availability
Operating system
Exclusively Windows 7

Programming language
C++ and Python 2.7.13

Additional system requirements
It is recommended that a minimum of 8 GB of memory be available for runs due to the ingestion of large amounts of spatial data.

Dependencies
For C++:
· All included

For Python:
· setuptools (version 18.5)
· configobj (version 5.0.6)
· untangle (version 1.1.0)
· pandas (version 0.18.1)
· matplotlib (version 1.5.1)
· seaborn (version 0.8.1)

Software location
Archive
  Name: GitHub
  Persistent identifier: https://doi.org/10.5281/zenodo.1321313
  Licence: BSD 2-Clause
  Publisher: Chris R Vernon
  Version published: v1.0.0
  Date published: July 25, 2018

Code repository
  Name: GitHub
  Identifier: https://github.com/IMMM-SFA/cerf
  Licence: BSD 2-Clause
  Date published: December 28, 2017

Language
English

(3) Reuse potential
CERF can be used to complement integrated human-Earth systems models and grid expansion models, toward evaluating the feasibility of customized power system
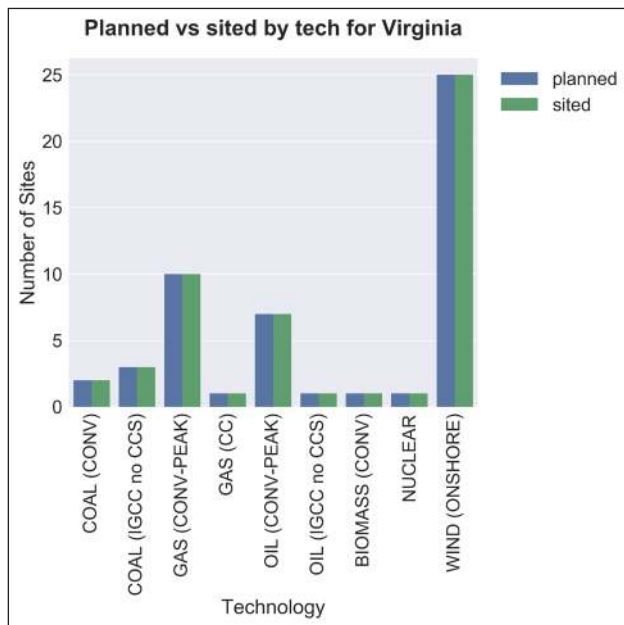
Figure 7: Bar plot showing planned versus achieved siting for each technology for Virginia from CERF's example data. This plot is used to detail when siting is not feasible for a target state.
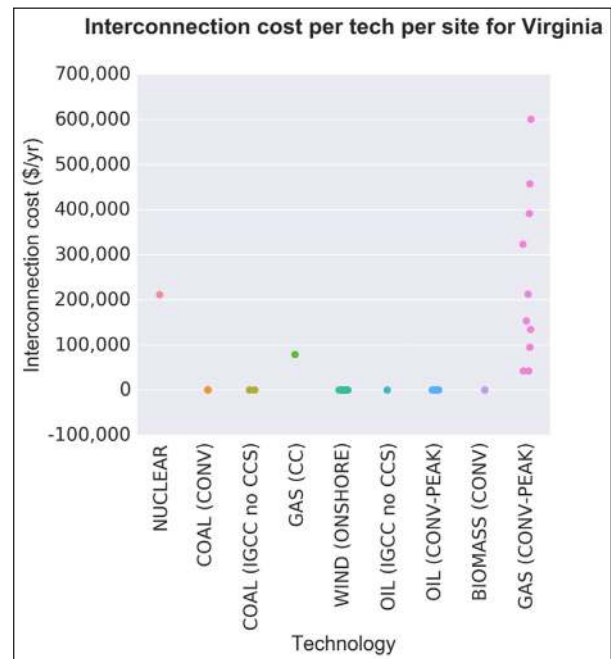

Figure 8: Jittered strip plot showing interconnection cost per site for each technology for Virginia from CERF's example data.

technology expansion scenarios. CERF can be called by other Python models through its Python wrapper on a time-step by time-step basis. CERF's Python output module can also be extended to address custom inquiries about the outputs; this includes multi-year analysis and comparison. CERF's structure is also responsive to the addition of new technologies by simply adding nodes into the XML structure containing technology specifics. Updates to CERF's technology assumptions and underlying spatial data can also easily be integrated by modifying the XML node values for the target technologies and updating the SAGA rasters, respectively.

Notes
1 See https://github.com/IMMM-SFA/cerf/blob/master/docs/cerf_econmomic_algorithm.pdf.
2 See https://github.com/IMMM-SFA/cerf/blob/master/docs/cerf_electric_gen_techs.pdf.
3 See https://github.com/IMMM-SFA/cerf/blob/master/docs/cerf_siting_criteria_literature.pdf.
4 See https://github.com/IMMM-SFA/cerf/blob/master/docs/cerf_common_suitability_rasters.pdf.
5 See https://github.com/IMMM-SFA/cerf/blob/master/docs/cerf_techspecific_suitability.pdf.
6 See https://github.com/IMMM-SFA/cerf/blob/master/docs/cerf_siting_criteria_development.pdf.
7 See https://github.com/IMMM-SFA/cerf/blob/master/example/example.py.
8 See https://github.com/IMMM-SFA/cerf/tree/master/example.
9 See https://git-lfs.github.com.
10 See https://pypi.python.org/pypi/setuptools.

Acknowledgements

Competing Interests
The authors have no competing interests to declare.

References
1. **Conrad, O, Bechtel, B, Bock, M, Dietrich, H, Fischer, E, Gerlitz, L, Wehberg, J, Wichmann, V** and **Böhner, J** 2015 System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.,* 8: 1991–2007. DOI: https://doi.org/10.5194/gmd-8-1991-2015
2. **Riahi, K, Shilpa, R, Krey, V, Cho, C, Chirkov, V, Fischer, G, Kindermann, G, Nakicenovic, N** and **Rafaj, P** 2011 RCP 8.5 – A scenario of comparatively high greenhouse gas emissions. *Climatic Change,* 109(33). DOI: https://doi.org/10.1007/s10584-011-0149-y
3. **Edmonds, J** and **Reilly, J M** 1985 *Global Energy: Assessing the Future,* 317. Oxford University Press, New York.
4. **Edmonds, J, Wise, M, Pitcher, H, Richels, R, Wigley, T** and **Maccracken, C** 1997 An integrated assessment of climate change and the accelerated introduction of advanced energy technologies-an application of MiniCAM 1.0. *Mitigation and adaptation strategies for global change,* 1(4): 311–339. DOI: https://doi.org/10.1023/B:MITI.0000027386.34214.60
5. **Voisin, N, Li, H, Ward, D, Huang, M, Wigmosta, H** and **Leung, L R** 2013a On an Improved Sub-Regional Water Resources Management Representation for Integration into Earth System Models. *Hydrology and Earth System Sciences Discussions,* 10(3): 3501–3540. DOI: https://doi.org/10.5194/hessd-10-3501-2013
6. **Voisin, N, Liu, L, Hejazi, M, Tesfa, T, Li, H, Huang, M, Liu, Y** and **Leung, L R** 2013b One-Way Coupling of an Integrated Assessment Model and a Water Resources Model: Evaluation and Implications of Future Changes over the U.S. Midwest. *Hydrology and Earth System*

Sciences, 17(11): 4555–4575. DOI: https://doi.
org/10.5194/hess-17-4555-2013

7. **ABB** 2017 *PROMOD*. [ONLINE] Available at: http://
new.abb.com/enterprise-software/energy-portfolio-
management/market-analysis/promod [Accessed 5
October 2017].

8. **U.S. Energy Information Administration** – Eia –
Independent Statistics and Analysis. https://www.eia.
gov/maps/layer_info-m.php.

9. **Homeland Infrastructure Foundation-Level Data**
https://hifld-geoplatform.opendata.arcgis.com/
datasets/electric-power-transmission-lines.