Geoscience Data Journal

WILEY

Geoscience Data Journal RMetS

DATA ARTICLE OPEN ACCESS

WRFtailor: A Toolkit for Tailoring the WRF Model Input Data

Amirhossein Nikfal

Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, Jülich, Germany

Correspondence: Amirhossein Nikfal (a.nikfal@fz-juelich.de)

Received: 30 September 2024 | **Revised:** 17 July 2025 | **Accepted:** 19 August 2025

Keywords: emissions data | WPS geographical data | WRF model | WRF-chem

ABSTRACT

WRFtailor is an open-source toolkit offering a set of distinct functionalities to customise and tailor the Weather Research and Forecasting (WRF) model input data, such as WPS geographical data or WRF/Chem emissions data. Before running the toolkit, the user should specify an area of interest (AOI) from the WRF input file, a variable to be tailored within the AOI, and a polynomial of variables that will replace the specified variable. WRFtailor is a Linux-based toolkit, written in Shell and NCAR Command Language (NCL) scripts, and is available at https://github.com/anikfal/wrftailor.

1 | Introduction

One of the key applications of atmospheric and climate models is simulating the impacts of the changes in the ground-based variables on the atmosphere and climate. Specifying the boundary conditions is one of the main preprocessing steps for the atmospheric models. The bottom boundary of a simulation domain requires knowledge of ground surface variables. Some of the most important variables are topographic heights and terrain, land use and vegetation cover, surface water bodies, etc. There are also anthropogenic ground variables, such as emissions of the atmospheric particulate matter, which are needed for air quality modelling. These applications can even be used for climate forecasts to predict the impacts of anthropogenic modifications on the environment over the next decades. Some cases demonstrating the role of ground surface variables in atmospheric modelling are discussed in this section.

In a study (Li et al. 2018) in the Kolkata Metropolitan Development area, the WRF model is used to project microclimate impacts of future land cover changes in the Kolkata Metropolitan Development area under two scenarios: urbanisation and irrigation expansion. Urbanisation significantly increased regional temperatures and altered rainfall patterns, while expanded irrigation had a cooling effect and redistributed precipitation. The findings offer key insights for land use, water and agriculture policy planning in response to climate change. Another study (Lu et al. 2012) used high-resolution simulations to analyse atmospheric conditions of the wildfire events during 2007. They found that smoke was funnelled through mountain passes and elevated by chimney effects for regional transport. Mountain regions in this study outperformed coastal regions in model accuracy for temperature and moisture. Considering the water bodies data, a study (Zhao et al. 2012) over the Great Lakes region significantly improved the simulation of lake-effect precipitation by the integration of satellite-derived Lake Surface Temperature (LST) and Lake Ice Cover (LIC) into the WRF model.

As mentioned previously, providing emissions data in addition to other ground-based data is necessary for air quality modelling. In a study (Parajuli et al. 2019) over the Middle

Abbreviations: AOI, area of interest; CMAQ, Community Multiscale Air Quality; EDGAR, Emissions Database for Global Atmospheric Research; LIC, Lake Ice Cover; LST, Lake Surface Temperature; NCL, NCAR Command Language; NEI, National Emissions Inventories; PM10, Particulate Matter; SEBAL, Surface Energy Balance Algorithm for Land; TEB, Town Energy Balance; WASF, West Asia Source Function; WPS, WRF Preprocessing System; WRF, Weather Research and Forecasting.

East and North Africa, a high-resolution dust source function in WRF-Chem was developed to simulate dust emissions. The model effectively identified key dust sources and their impacts on air quality, showing that dust from the basin exceeds Particulate Matter (PM10) standards in downwind cities. In another similar study (Nabavi et al. 2017), the West Asia Source Function (WASF) to improve dust forecasts in the WRF-Chem model by correcting errors in the widely used Ginoux source function (Ginoux et al. 2001) was introduced. Implementing WASF significantly improved dust simulation accuracy, increasing Spearman correlation by 12%–16% compared to standard models. Emissions data have a crucial role in simulating gaseous pollution concentration. The available aerosol modules in WRF-Chem have been compared in another study (Zhang et al. 2016). The study highlights the need for accurate aerosol–cloud interaction modelling and shows that anthropogenic emissions significantly influence ozone, PM2.5, and secondary organic aerosols, impacting climate and air quality across the United States.

2 | Methods

WRFtailor has the same software structure as implemented in the PostWRF software (Nikfal 2023) for the visualisation and analysis of WRF output data. However, WRFtailor is designed specifically to tailor WRF input data. The main components of the software consist of Linux bash scripts, NCL (www.ncl.ucar.edu) codes and a text file known as a namelist, which must be modified by the user to carry out a specific tailoring task. NCL is a specialised programming language for the analysis of scientific data, particularly NetCDF data, and includes functions specifically designed for WRF model data. To run the WRFtailor tool, the user must modify a declarative configuration file called `namelist.tailor` and specify the desired tailoring task along with the required information and possibly additional input data (in addition to the WRF input data). After executing *wrftailor.sh*, the contents of `namelist.tailor` are read by the background shell scripts and passed to the NCL codes, which are responsible for modifying the WRF input data accordingly.

2.1 | WRF Model

The WRF model (Skamarock et al. 2019) is among the most recognised atmospheric regional models and is widely used in various atmospheric and environmental research and operational sectors. As of 2015, the number of scientific publications related to the WRF model exceeded 3500 articles, and this number is continually growing (Powers et al. 2017). The WRF model is not limited to atmospheric modelling; it can also be applied to other environmental sectors, such as air pollution modelling (WRF-Chem (Grell et al. 2005)) and hydrological simulations (WRF-Hydro (Gochis et al. 2018)). The WRF model can be coupled with other geoscientific models to provide meteorological input data. Many of the models coupled with WRF are air pollution models, such as Community Multiscale Air Quality (CMAQ) (Wong et al. 2011). Radiation transfer and biophysical simulations, such as Town Energy Balance (TEB) (Meyer et al. 2020) and Surface Energy Balance Algorithm for Land (SEBAL) (Nikfal and Karimi 2024), are other modelling systems that have been coupled with the WRF model. The aforementioned models, which do not belong to the WRF modelling system, can be coupled in an offline manner. This means that the WRF model is run first, and then the WRF output is ingested as input data for the coupled model. However, for WRF-Chem and WRF-Hydro, the simulations are online, meaning that the WRF model and the air pollution or hydrological models run simultaneously.

The general structure and main components of the WRF modelling system are depicted in Figure 1. The first stage involves preparing the input data to provide the initial and boundary conditions for the simulation domains. In the second stage, the main simulation is carried out by the WRF dynamic core, applying atmospheric physics and boundary layer options set by the user. Simulation results are stored in one or more output data sets, which can subsequently be post-processed and visualised to analyse the results.

2.1.1 | WRF Ground-Based Input Data

Depending on the simulations run by the WRF main meteorological model or by WRF coupled with another geoscientific model, several input data, such as topography height, land use, soil temperature and surface albedo, may be required. For running the WRF model alone, two categories of input data are required: meteorological data to provide initial and boundary conditions and geographical static data to provide ground surface conditions. For the online coupled simulations (e.g., WRF-Chem), there are additional required input data, such as emissions data for air pollution modelling or specific ground surface data for hydrological simulations. Nevertheless, regardless of the models coupled with WRF, meteorological and geographical static data are necessary input data for any simulations based on the WRF model. Direct and arbitrary modification of meteorological variables (e.g., changing wind speed or temperature independently in gridded input files) is generally not appropriate in operational workflows, due to the interdependence of atmospheric variables. However, techniques such as PV inversion (Davis and Emanuel 1991) explicitly account for these relationships and are well-established in case study design and controlled perturbation experiments. That said, the WRFtailor tool is designed for a different scope: it focuses specifically on non-meteorological data, including static geographical inputs (e.g., land use, vegetation, albedo) and emissions data. It does not support the modification of meteorological initial or boundary conditions, nor does it include dynamical consistency mechanisms like the PV

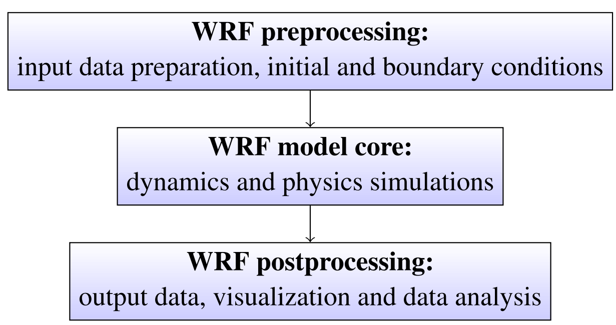


FIGURE 1 | Main components of the WRF modelling system.

TABLE 1 | Examples of ground-based WRF input data.

Variable name	Number of dimensions	Time dimension size	Categorical dimension name	Categorical dimension size
Albedo	4	Single	Month	12
Landuse	4	Single	land_cat	24
Green fraction	4	Single	Month	12
Erodibility	4	Single	dust_erosion_dimension	3
Sand fraction	3	Single	—	—
Topography height	3	Single	—	—
CO emissions	3	Single or multiple	—	—
SO2 emissions	3	Single or multiple	—	—

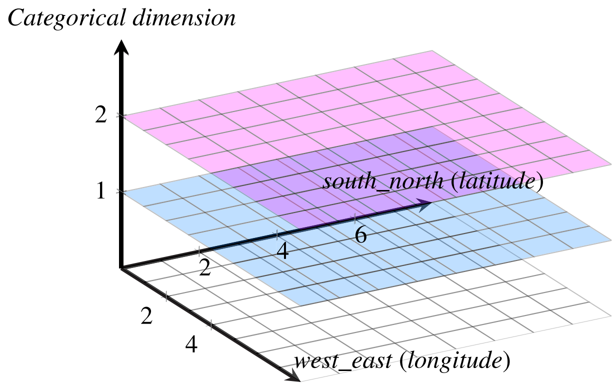


FIGURE 2 | The typical WRF input geographical data structure based on the NetCDF format, with one single time.

technique. The geographical and emissions WRF input data can be directly modified by the user to be ingested in the WRF simulations with more accurate values. Table 1 shows some examples of the geographical static data and emissions data as the most important ground-based (non-meteorological) WRF input data.

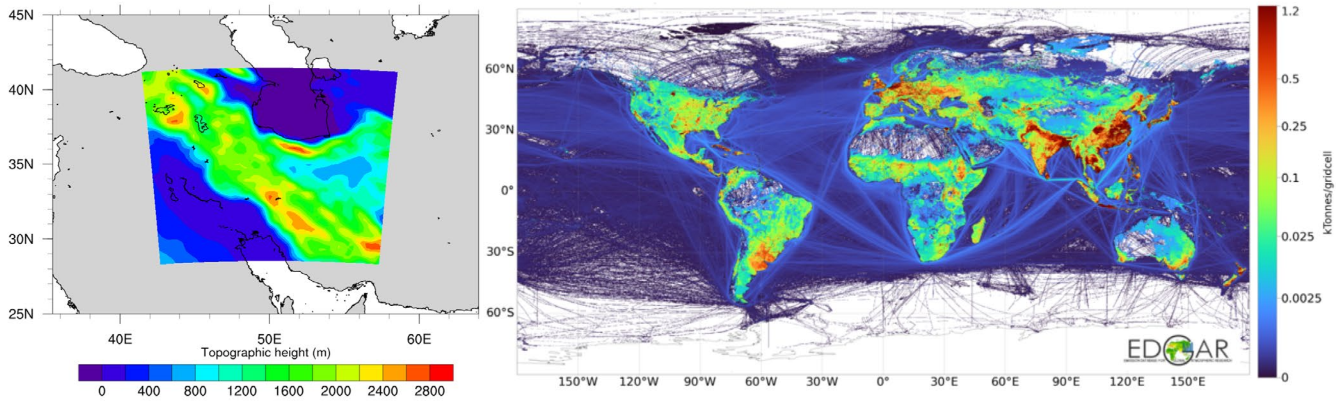
2.1.1.1 | Geographical Static Data. Ground surface acts as the bottom boundary for most atmospheric simulation domains. Some of the key atmospheric processes, such as heat and moisture fluxes, occur in this area as the interaction boundary between the atmosphere and the Earth's surface. Among the geographical data are terrain height, soil type, vegetation fraction and land use. The temporal variations of these geographical variables are not significant in comparison to atmospheric data (temperature, pressure, etc.), which exhibit hourly variations. Therefore, they are called static data. Nevertheless, many of the geographical data sets are provided in monthly resolution, which can show significant changes in the values of some static variables, such as vegetation and land use. Among the geographical data, some variables, such as topography, albedo and soil type, are mandatory, whereas other optional data (e.g., erodibility and groundwater data) are required for specific applications, such as dust and irrigation modelling. All of the WRF geographical data are available at https://www2.mmm.ucar.edu/wrf/users/download/get_sources_wps_geog.html.

WRF geographical static data are usually stored in NetCDF format, with gridded values in the north–south and east–west directions as the x–y plane. Such dimensions are shown in Figure 2 as a 3D coordinate axis system. The categorical dimension can provide some spatial, temporal or quality information about the variable, such as soil category or monthly variation of green fraction. We can omit the (hourly) time dimension since it is usually a single size dimension representing the simulation start time. Figure 3a shows an example map of topographic height for a sample WRF domain.

2.1.1.2 | WRF/Chem's Emissions Data. Preparing emissions data and incorporating the emission data into the model is a prerequisite step for air quality modelling using the WRF/Chem or any other atmospheric numerical model. Various emissions data types, such as anthropogenic (man-made) or biogenic (natural-origin) emissions, can be provided from several data sources. For example, Emissions Database for Global Atmospheric Research (EDGAR) (Crippa et al. 2016) is a well-known global data set of anthropogenic emissions. As shown in Figure 3b, which illustrates the EDGAR data set for CH4 emissions, higher values are clearly evident in more industrialised regions. There might be more accurate emissions data for specific countries. For example, National Emissions Inventories (NEI) (Lin et al. 2005) is a specialised emissions data set for the United States. Since global emissions data are usually monthly, they can be ingested into the WRF/Chem model at a single time. However, they are also structured as NetCDF data, similar to that in Figure 2, but without a categorical dimension. In other words, many preprocessed emissions data for WRF/Chem can be considered 2D grid data on the ground surface. More information about the WRF/Chem model and the emissions data is available in the model's user guide (Peckham 2012).

2.2 | WRFtailor Software Structure

The main components of the WRFtailor toolkit are displayed in Figure 4. The software consists of shell scripts, NCL codes and a namelist to configure the setup for the run process. The only file that needs to be modified by the user is `namelist.tailor`. Shell scripts, as Linux terminal commands, manage the



(a) Geographical static data

(b) EDGAR emissions data

FIGURE 3 | Examples of the WRF input data for (a) topographic height and (b) CH4 emissions.

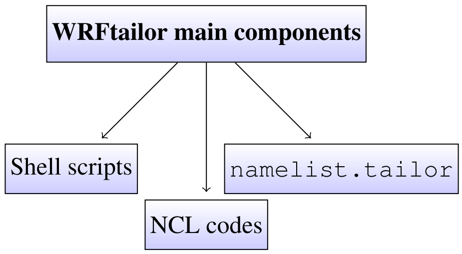


FIGURE 4 | WRFtailor building blocks and main software components.

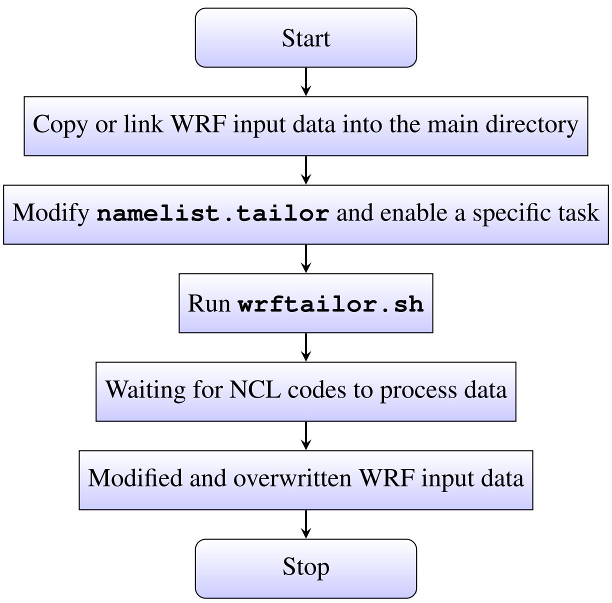


FIGURE 5 | Flowchart of the WRFtailor run process.

NCL code(s) associated with the task specified by the user in `namelist.tailor`.

WRFtailor can modify the WRF or WRF/Chem input data through five different approaches. These approaches include using certain methods, such as a shapefile, a bounding box, a list of coordinate points, the entire smallest domain and a GeoTIFF file. The user can specify a specific approach in `namelist.tailor` and then run the tool to modify and overwrite the WRF input data. According to Figure 5, WRFtailor features a straightforward and simple run process. After cloning the software from https://github.com/anikfal/wrftailor, the user is required to copy or link some WRF or WRF/Chem input data in the main software directory. The next step is to modify and enable the specific task (approach) in `namelist.tailor` that the user wants to use to modify the WRF input data. Then, the code wrftailor.sh must be run. This code calls the NCL codes, which are responsible for modifying the input data.

The grid points that lie within the areas defined by a bounding box, a shapefile or the entire subdomain are directly modified in the original file. Therefore, interpolation is not applicable for these three methodologies. However, when tailoring WRF input data using pairs of coordinates, the nearest neighbour interpolation method is used. This method is applied to the grid points in the original WRF input file. The NCL function *getind_latlon2d* is used for this tailoring approach based on coordinate pairs. In the raster replacement methodology, once the point locations of each raster pixel are determined, they are interpolated within the grid points of the original WRF input file using the NCL function *rgrid2rcm_wrap*. This is a true interpolation method, not merely based on nearest neighbour grid points.

2.2.1 | Configuration File

As mentioned before, WRFtailor must be set up by the user via a declarative configuration file called `namelist.tailor`. This file consists of several sections, each associated with a specific task. In other words, each task corresponds to a particular type of AOI, such as a shapefile area, a subdomain or a bounding box. Therefore, the parameters in each section may differ from those in other sections. The first parameters in `namelist.tailor` are displayed in Table 2.

3 | Results

To demonstrate the capabilities of the WRFtailor software, we discuss examples of each approach to modifying WRF input data. Since the WRF model follows a standard input and output data structure, users can apply the same tasks to any other WRF input data. In Figure 6, two sets of WRF domain configurations are presented as examples of how WRFtailor functionality can be utilised. Figure 6a shows the domain configurations used for modifying geographical static data, whereas Figure 6b displays WRF/Chem domain configurations to use emissions data.

3.1 | Tailoring With a Shapefile Mask

WRFtailor can modify WRF input data within the area defined by a shapefile. Figure 7 (WRF domain configuration according to Figure 6a) shows a sample modification applied using

TABLE 2 | First parameters of `namelist.tailor`.

WRF files and input data	
number_of_domains	=2
domain_1	=/home/anikfal/wrftailor/geo_em.d01.nc
domain_2	=/home/anikfal/wrftailor/geo_em.d02.nc
domain_3	=/home/anikfal/emissions/wrfchemi_d03_2021-01-22_00:00:00
domain_4	=
domain_5	=
Modify by shapefile mask	
shapefile_ON_OFF	=1
target_variable1	=GREENFRAC
target_var_level1	=4
substitute_variable1	=GREENFRAC*2+ALBEDO12M/100
substitute_var_levels1	=4, 4
path_to_shapefile	=/home/anikfal/wrftailor/tehranbasin.shp
inverse_mask_on_off	=0
...	

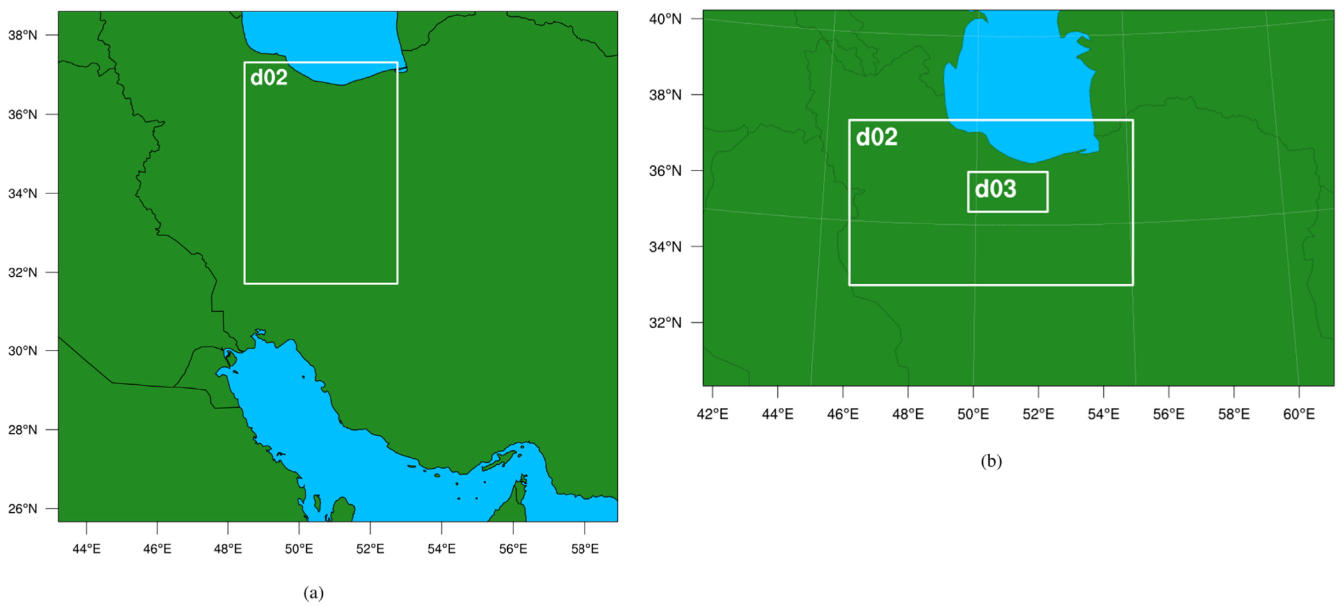
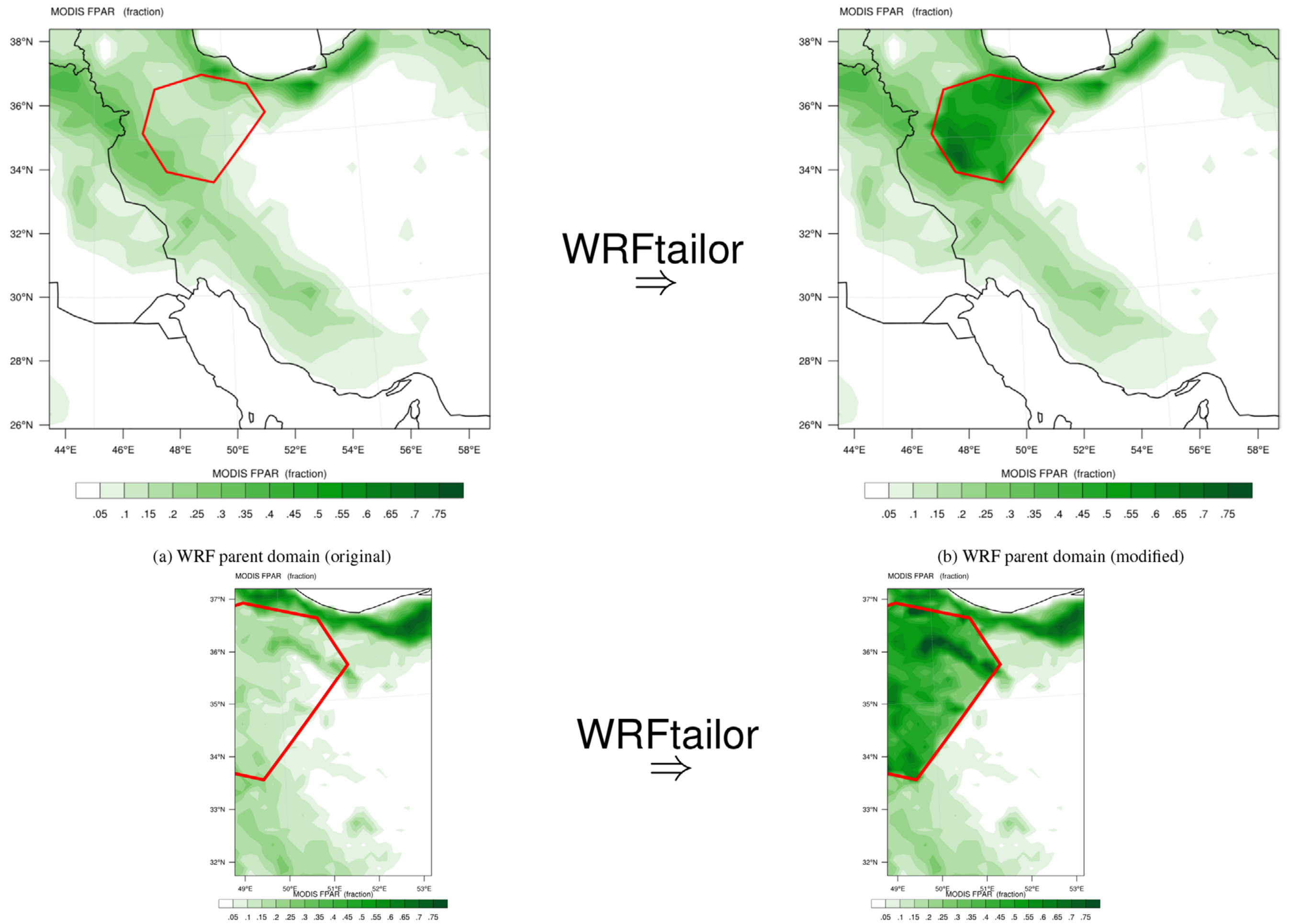


FIGURE 6 | WRF domain configurations used to showcase the WRFtailor's capabilities.



(a) WRF parent domain (original)

(b) WRF parent domain (modified)

(c) WRF subdomain (original)

(d) WRF subdomain (modified)

FIGURE 7 | WRF original parent and subdomains (a and c) for the green fraction variable, modified within the shapefile (b and d).

Equation 1 within the shapefile (red polygon). In this case, the variable *green_fraction* for the fourth month is replaced by the sum of *green_fraction* and *albedo* divided by 100 for the fourth month at each grid point inside the shapefile border. As a result, the new values inside the shapefile, shown in Figure 7b,d, are noticeably larger than the original values displayed in Figure 7a,c.

$$GREENFRAC^{month\_4} = \left(GREENFRAC^{month\_4} * 2\right) + ALBEDO12M^{month\_4} / 100 \quad (1)$$

3.2 | Tailoring With a Bounding Box

A rectangular area can be specified in `namelist.tailor`, allowing WRFtailor to modify a specific variable over that area. The rectangular region is depicted in blue in Figure 8, with a portion of the region covering the subdomain. The primary variable modified by Equation 2 is soil erodibility. This variable has a categorical dimension with three levels, corresponding to soil layers. The erodibility values in the first layer of the categorical dimension have been divided by three, as shown in Figure 8b,d.

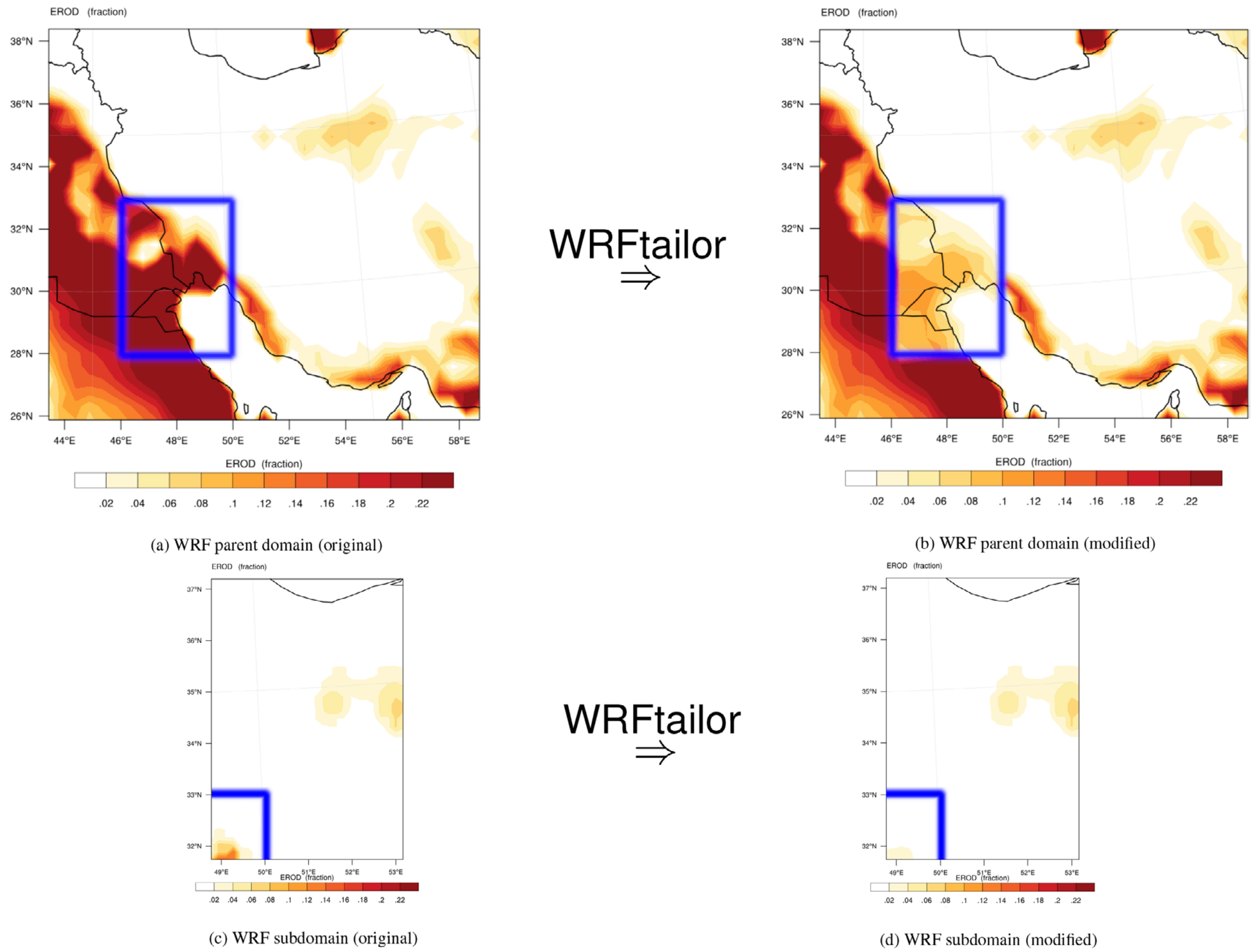
$$EROD^{layer\_1} = EROD^{layer\_1} / 3 \quad (2)$$

3.3 | Tailoring Using Pairs of Coordinates

Users can specify a list of coordinates (latitude and longitude pairs) for WRFtailor to modify WRF input data at those specific points. Figure 9 shows the WRFtailor results based on the points listed in Table 3. Circles in Figure 9b,d,f represent the new values at the specified points.

3.4 | Tailoring Using the Whole Nested Domain

In some research studies, it is desirable to modify a specific variable across an entire subdomain. A good example is shown in Figure 10, where the emissions values for the NO variable in the last subdomain (second nested domain) have been set to zero. WRFtailor can assign any new values, which can even be a function of other variables, as provided in other tasks of the WRFtailor toolkit. The zero values for NO emissions can be seen in Figure 10b,d,f.



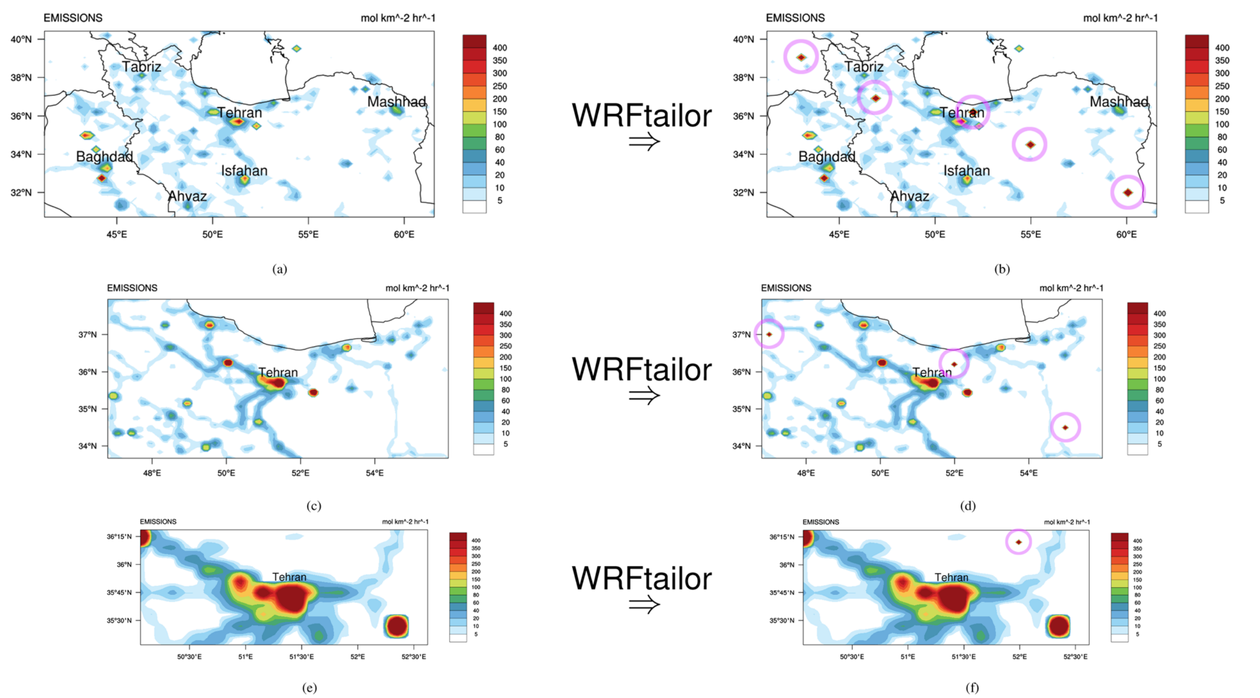
(a) WRF parent domain (original)

(b) WRF parent domain (modified)

(c) WRF subdomain (original)

(d) WRF subdomain (modified)

FIGURE 8 | WRF original parent and sub domains (a and c) for the erodibility variable, modified within the bounding box (b and d).



(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 9 | WRF/Chem parent and subdomains (a, c, e) for the NO emissions, modified over 5 points (b, d, f). Circles are depicted for clearer visualisation of the changes.

TABLE 3 | Examples of a list of coordinates modified by WRFtailor for the NO emissions values.

Latitude	Longitude	New values
39	43	700
37	47	800
34.5	55	900
36.2	52	1000
32	60	1100

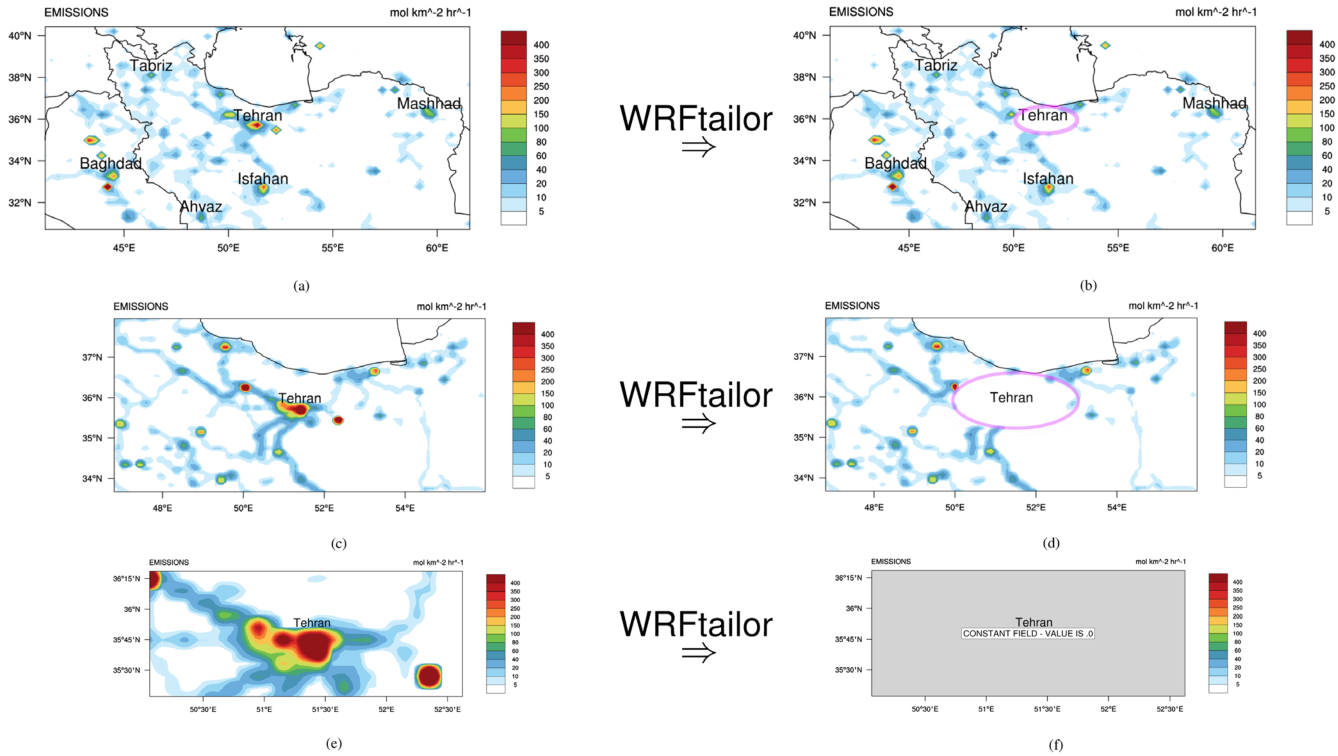


FIGURE 10 | Making the NO emissions equal to zero for the nested domain (domain d03 in Figure 6b). The d03 domain is encircled.

3.5 | Tailoring Using Raster Replacement

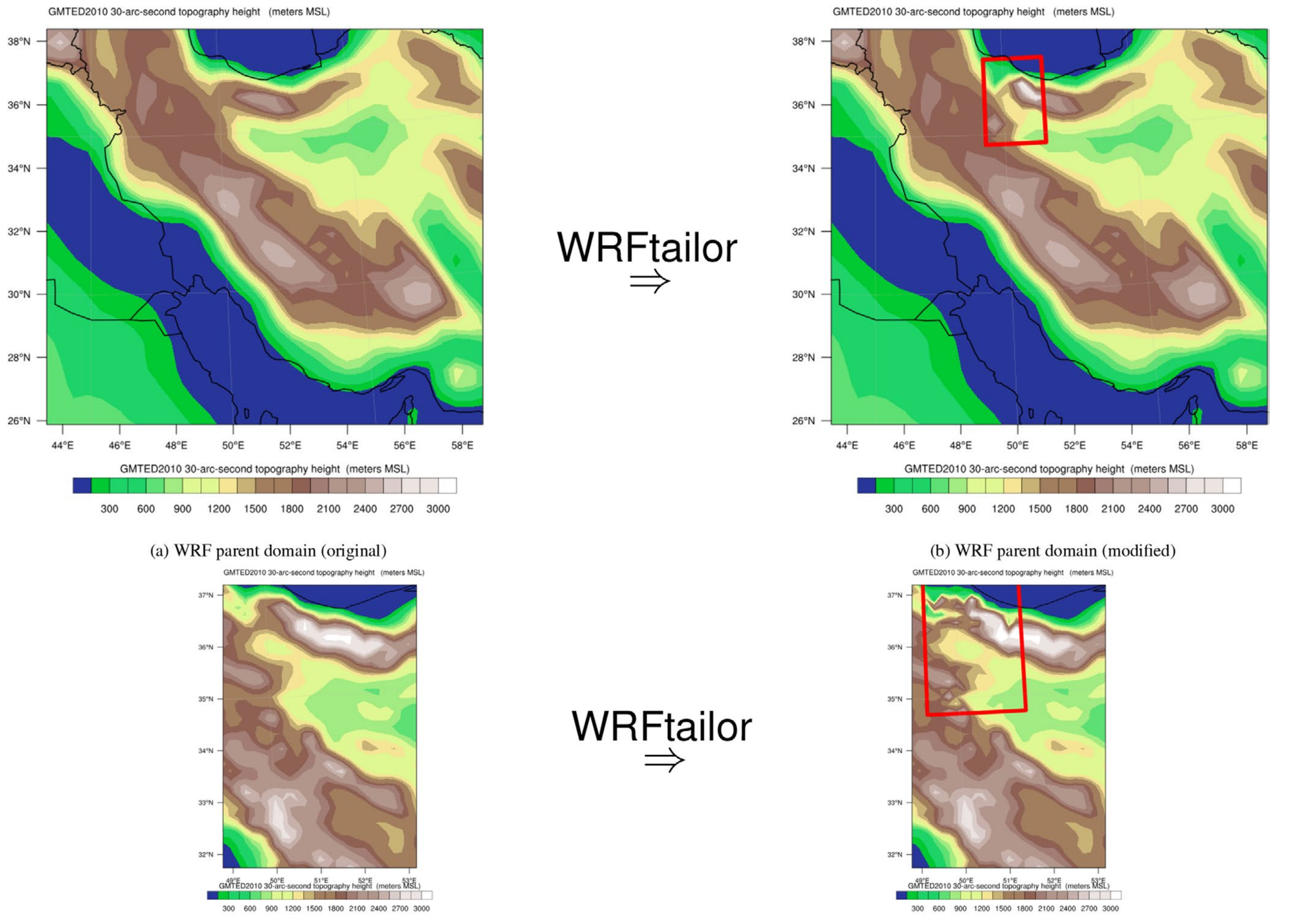
One of the most common tasks in modifying WRF input data is replacing the default values of a specific variable with an updated version. The updated values can be obtained from satellite data or other raster sources, which are often based on the GeoTIFF data format. Figure 11 shows the topographic heights over the domain configurations depicted in Figure 6a, modified using a sample GeoTIFF file with higher spatial resolution. The final modifications are visible in Figure 11b,d, within the red rectangle representing the GeoTIFF file boundary.

WRFtailor includes an optional mass conservation feature designed for use when tailoring mass-related variables, such as emissions, with raster-based data. When enabled, the tool calculates the area of each grid cell in both the input raster and the WRF domain, adjusts the values to preserve total mass during interpolation and applies a scaling factor to ensure that the final modified field maintains physical consistency. This option can be activated via the `enable_mass_conservation` directive in the `namelist.tailor` file, allowing users to apply it when appropriate for their specific use case.

4 | Conclusions

WRFtailor offers five distinct methods to tailor and customise WRF model input data, which were discussed in detail in previous sections. Since WRFtailor does not require a complex installation process and operates through a simple configuration file, it is a user-friendly and efficient solution for modifying WRF model input data. The five approaches offered by WRFtailor cover most of the methods commonly used by researchers to tailor model data. However, more approaches can be added in future software developments.

WRFtailor is open-source software available at https://github.com/anikfal/wrftailor under an MIT licence. While the WRFtailor toolkit is very easy to use, online documentation is



(a) WRF parent domain (original) (b) WRF parent domain (modified)

(c) WRF subdomain (original) (d) WRF subdomain (modified)

FIGURE 11 | WRF original parent and subdomains (a and c) for the green fraction variable, modified by being replaced with the values of a raster datafile (b and d).

available at https://wrftailor.readthedocs.io/en/latest/, along with training videos (https://youtube.com/playlist?list=PL93HaRiv5QkA8uzFzcZkyTqkKPweJajrJ&si=kCJ3UNvnJlrGX-FWd) for each approach in the toolkit.

Acknowledgements

Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in WRFtailor at https://github.com/anikfal/wrftailor.

References

Crippa, M., G. Janssens-Maenhout, F. Dentener, et al. 2016. "Forty Years of Improvements in European Air Quality: Regional Policy-Industry Interactions With Global Impacts." *Atmospheric Chemistry and Physics* 16, no. 6: 3825–3841.

Davis, C. A., and K. A. Emanuel. 1991. "Potential Vorticity Diagnostics of Cyclogenesis." *Monthly Weather Review* 119, no. 8: 1929–1953.

Ginoux, P., M. Chin, I. Tegen, et al. 2001. "Sources and Distributions of Dust Aerosols Simulated With the GOCART Model." *Journal of Geophysical Research: Atmospheres* 106, no. D17: 20255–20273.

Gochis, D., M. Barlage, A. Dugger, et al. 2018. The WRF-Hydro Modeling System Technical Description Version (5.1.1).

Grell, G. A., S. E. Peckham, R. Schmitz, et al. 2005. "Fully Coupled "Online" Chemistry Within the WRF Model." *Atmospheric Environment* 39, no. 37: 6957–6975.

Li, X., C. Mitra, L. Dong, and Q. Yang. 2018. "Understanding Land Use Change Impacts on Microclimate Using Weather Research and Forecasting (WRF) Model." *Physics and Chemistry of the Earth, Parts a/b/c* 103: 115–126.

Lin, C. J., T. C. Ho, H. Chu, et al. 2005. "A Comparative Study of US EPA 1996 and 1999 Emission Inventories in the West Gulf of Mexico Coast Region, USA." *Journal of Environmental Management* 75, no. 4: 303–313.

Lu, W., S. Zhong, J. Charney, X. Bian, and S. Liu. 2012. "WRF Simulation Over Complex Terrain During a Southern California Wildfire Event." *Journal of Geophysical Research: Atmospheres* 117, no. D5: 2011JD017004.

Meyer, D., R. Schoetter, M. Riechert, et al. 2020. "WRF-TEB: Implementation and Evaluation of the Coupled Weather Research and Forecasting (WRF) and Town Energy Balance (TEB) Model." *Journal of Advances in Modeling Earth Systems* 12, no. 8: e2019MS001961.

Nabavi, S. O., L. Haimberger, and C. Samimi. 2017. "Sensitivity of WRF-Chem Predictions to Dust Source Function Specification in West Asia." *Aeolian Research* 24: 115–131.

Nikfal, A. 2023. "PostWRF: Interactive Tools for the Visualization of the WRF and ERA5 Model Outputs." *Environmental Modelling & Software* 160: 105591.

Nikfal, A., and M. A. Karimi. 2024. "Operational Workflow to Simulate Biophysical Variables, Based on the Coupled WRF/SEBAL Models." *Computers and Electronics in Agriculture* 222: 109003.

Parajuli, S. P., G. L. Stenchikov, A. Ukhov, and H. Kim. 2019. "Dust Emission Modeling Using a New High-Resolution Dust Source Function in WRF-Chem With Implications for Air Quality." *Journal of Geophysical Research: Atmospheres* 124, no. 17–18: 10109–10133.

Peckham, S. E. 2012. WRF/Chem Version 3.3 User's Guide.

Powers, J. G., J. B. Klemp, W. C. Skamarock, et al. 2017. "The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions." *Bulletin of the American Meteorological Society* 98, no. 8: 1717–1737.

Skamarock, W. C., J. B. Klemp, J. Dudhia, et al. 2019. "A Description of the Advanced Research WRF Version 4." *NCAR Techical Note NCAR/Tn-556+ Str* 145.

Wong, D., J. Pleim, R. Mathur, et al. 2011. "WRF-CMAQ Two-Way Coupled System With Aerosol Feedback: Software Development and Preliminary Results." *Geoscientific Model Development Discussions* 4, no. 3: 2417–2450.

Zhang, Y., J. He, S. Zhu, and B. Gantt. 2016. "Sensitivity of Simulated Chemical Concentrations and Aerosol-Meteorology Interactions to Aerosol Treatments and Biogenic Organic Emissions in WRF/Chem." *Journal of Geophysical Research: Atmospheres* 121, no. 10: 6014–6048.

Zhao, L., J. Jin, S. Y. Wang, and M. B. Ek. 2012. "Integration of Remote-Sensing Data With WRF to Improve Lake-Effect Precipitation Simulations Over the Great Lakes Region." *Journal of Geophysical Research: Atmospheres* 117.