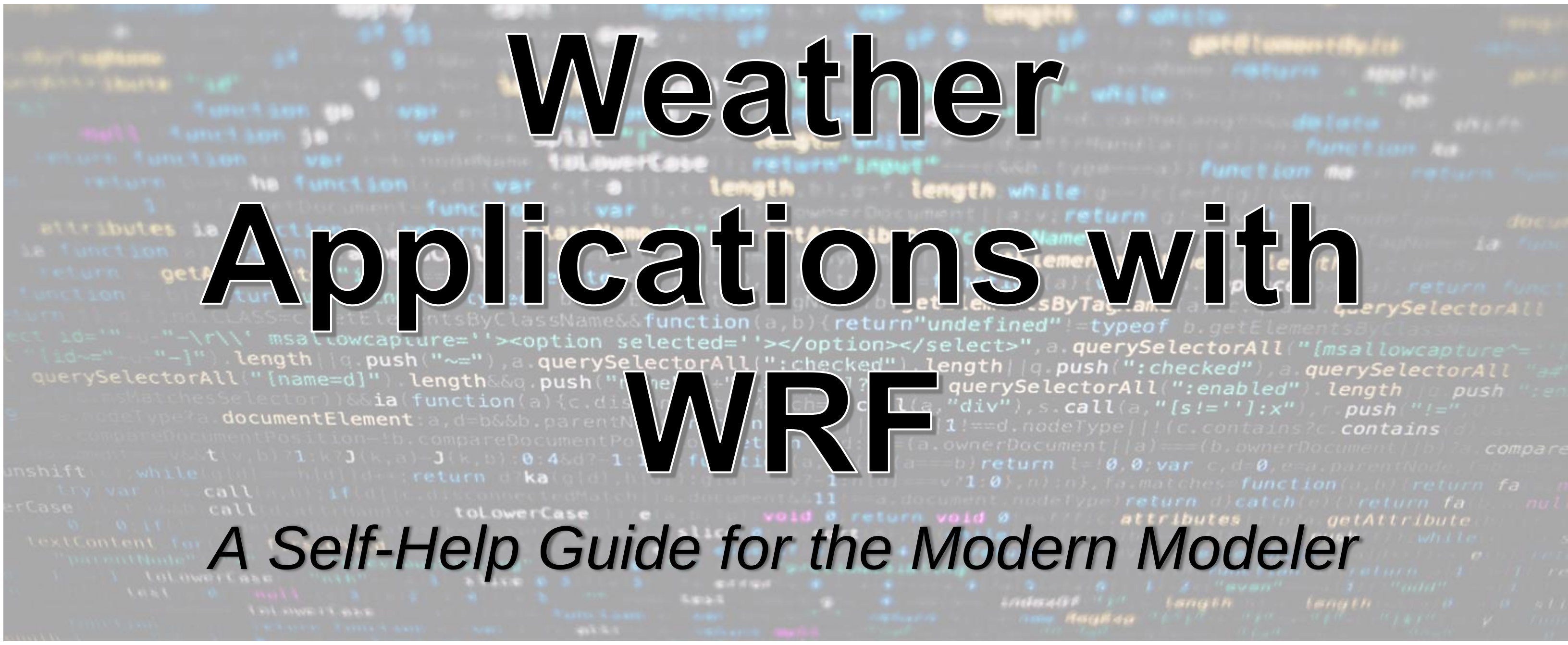
YComm

YARKER CONSULTING
ATMOSPHERIC SCIENCE & EDUCATION

m²lab
.org



Weather Applications with WRF

A Self-Help Guide for the Modern Modeler

Michel dos Santos Mesquita, PhD

Morgan B. Yarker, PhD, CCM

Michelle Serino

Chapter 2: Experimental Design

Introduction

This chapter will focus on how to design a regional climate modeling experiment. We believe this is one of the most important components of understanding weather and climate models. Warner (2011, p.3) agrees, stating, "...it is the author’s experience that using the model early in the process only prolongs the amount of time required to complete a research project, or a thesis."

A common mistake novice modelers tend to make is to run the model first and then try to make sense of the results. An immediate drawback to this approach is that it will likely take you an necessary long time to interpret your results. Or worse, you may discover your model set-up was not ideal. For example, maybe the horizontal resolution was too coarse to capture the physical processes; or perhaps the domain size was too small to capture the entire event; or perhaps there are too many biases because it was not set up appropriately. In all these cases, you would need to start all over again, wasting hours of work, not to mention the computational resources, as well.

This is why Warner (2011, p.3) is so blunt about the fact that it will take the modeler more time to complete a task if he or she does not make a plan beforehand. We call this plan **experimental design**. If an experiment is well-planned and the work is put in up-front, then not only will it be much easier to evaluate and interpret the results, it will also be:

- unnecessary to re-run the model
- easier to identify why something may have gone wrong
- completed in time
- much more insightful, which is imperative if you want to publish meaningful results

So, this is why we have dedicated an entire chapter towards designing experiments. By the end of this chapter, you should have a better understanding about:

- how to design a scientific experiment with respect to a weather/climate model
- how to establish a scientific question with respect to a weather/climate model, as well as design an experiment to answer the question
- how to decide the appropriate resolution for your experiment
- how to choose a proper domain size for your experiment
- how to select the appropriate projection for your experiment

Self Reflection

Before continuing, consider the following questions to begin thinking about your experimental design options:

1. What region interests you?
2. Are you interested in analyzing a weather or climate event?
3. How much time do you have to devote to this experiment?

Application Activity 2.1: Design an Experiment to Solve a Problem

The best way to learn how to design a science experiment is to practice doing it. So, let's get right into it!

Scenario

Suppose you have received funding to work on a project for your city. The funding agency would like you to study whether the city will experience more or less rainfall in the next 20 years. They need information for different areas of the city, and so there is a need for high-resolution modeling. The funding you received is enough for you to complete the work within 10 months. You have decided to use the WRF model for this work, and your research institute has installed WRF on a Linux cluster. You have made a few tests and found out that you can run the equivalent of 1 year of your simulation in 72 hours at 30-km horizontal resolution.

Discussion

The scenario we presented to you may sound complicated. You may have noticed that it takes a lot of time to run the model. But this scenario is a common one when dealing with climate and high-resolution modeling.

Try it on Your Own

There are lots of ways to go about planning an experiment. But first, it is important to think about what your ultimate goals are and how you anticipate being able to achieve them. For this reason, we have devised a set of questions that you should consider asking yourself as you contemplate what the design of your experiment should be. In terms of the given scenario, consider these questions:

A. Imagine this is the city you came from or the city you are living in now. What are the dimensions of the city? Do you think 30-km horizontal resolution would be suitable for this experiment?
B. Before you run WRF for the future simulations, you need to validate the model. Validating the model means that you confirm that it can simulate the present climate well and that its results seem reasonable. So, you decide to use the period from 1981 to 2011 to validate your simulation. How long would it take you to run this experiment?

C. Then you spend 1 month post-processing the output of the model and analysing your data. Now, you want to compare your model with observations. Where would you obtain the observations for your city? Are they easily accessible? If they are not, which alternatives would you use?
D. After spending another month validating your experiment, you discover that it is not able to reproduce the climate of your city well. What would you do?
E. After determining why your simulation did not work, you decide to make a new simulation. How long would that take again, including the postprocessing and validation?
F. You then work on the future simulations. How long will those take you?
G. Do you think you would have enough time to complete your project within the target 10 months?

As far as modeling projects go, this is a fairly straightforward scenario -- we have not even considered more complex situations where ensembles (i.e, running the model several times to measure the uncertainty in the model) are necessary. If you run ensemble simulations, your research would take much longer to complete. Working with time constraints is very difficult under any circumstance, but with modeling, the computational time is a factor that needs serious consideration, especially if it also costs money to do. This is why you need to plan your projects carefully before you start.

Check Your Reasoning

A. **Smaller regions of interest require finer resolutions.** Before making any decisions, you must determine the size of the region you are interested in. For example, the city of Bergen, Norway, is 466 square kilometers, which is roughly equivalent to 20 km by 20 km in dimensions. If we would like to run WRF at 30-km resolution, it would be far too coarse to resolve this city adequately. Rather, this would be a much better option if we were interested in the entire country of Norway. For a smaller region, like Bergen, it would be more useful to use a much finer resolution, such as a 3-km resolution.
B. **Use the general calculation.** In the given scenario, it would take 72 hours to run the equivalent of 1 year of your simulation. While this would vary from computer to computer, there is a general calculation that will help you estimate simulation run-time: number of simulation years x number of hours to simulate 1 year. So, if you want to simulate 30 years, the calculation would be: 30 years x 72 hours/year = 2160 hours (90 days) to complete the simulation.
C. **Observation data should come from local resources.** A great place to start will be the country's (or countries') national weather organization. For example, data in and around the city of Bergen, Norway, can be accessed through the Norwegian Meteorological Institute or in the United States from the National Oceanic and Atmospheric Administration (NOAA). However, if no data are available, a fallback can be to use reanalysis output, such as those from the European Centre for Medium-Range Weather Forecasts (ECMWF; ERA Interim) or jointly from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) in the United States. Sources for these and other data can be found on the resources webpage.

D. **Check your parameterizations.** Choosing the best parameterizations for a model simulation is, by far, the most critical component, as well as the most difficult to do. There is never one single best approach -- it is highly dependent on your region of interest, your grid resolution, domain size, and your research question. Therefore, if your simulation does not capture the area well, we suggest starting here. We also recommend running several simulations for a very short period of time using different configurations to determine which combination of parameterizations best captures the observation data for your region. Then you can run your finalized experiment.
   a. If you are in a time crunch, it is possible to correct for the bias in your data (e.g., if your model consistently underestimates high temperatures by a factor of two, you could adjust your output appropriately to account for this bias). However, we recommend using this approach *with extreme caution*. It is possible that correcting for bias will mask certain processes that can impact the quality of your findings.

E. **What's the rule of thumb here?** This would take: 90 days to complete my simulation + 1 month post-processing + 1 month validating = 5 months. Recall that the 90 days came from Question B. The time it takes for post-processing and validation will depend on the computer capability, as well as the type of analysis you want to perform with the output. There are other tools available for post-processing and validation, which may make it easier and faster for you to work with the results and makes it difficult to give a rule of thumb here. Regardless, you still need to account for this time if you are working on a project with a deadline.

F. **Use the general calculation.** The timeframe to run the future simulations would take: 20 years x 72 hours/year = 1440 hours (60 days).

G. **Model simulations can take several months to set up and run.** If nothing goes wrong, you can assume the following:
   a. It will take approximately 5 months for the present climate simulation, including post-processing and validation.
   b. It will take approximately 4 months for the future climate simulation, including post-processing and validation.
   c. This yields a total project time of 9 months. However, I am sure you can imagine that if you need to re-run your simulation, it will be very difficult to finish in the time frame provided.

This is why careful planning is crucial! We discuss the preceding points in more detail throughout the rest of this chapter.

Model Uncertainty

Chaos

When thinking about chaos and its role in atmospheric processes, an excellent comparison is to think about the Butterfly Effect. The Butterfly Effect is a thought experiment that was first posed by Edward Lorenz back when computer models were just starting to become commonplace for forecasting the weather. He suggested that something as large and destructive as a tornado could have been caused by something as small and seemingly insignificant as the flap of a butterfly's wing several weeks prior. He argued that the flap of the wing could influence tiny processes that grow and grow in magnitude until it leads to the conditions that are ideal for thunderstorm development and tornadogenesis. His point was that computer models could not capture such small detail, let alone keep up with the series of random outcomes that would eventually lead to larger atmospheric phenomena.

In recent years, the Butterfly Effect has become a more generalized description of any chaotic process, where individual events are nearly impossible to predict. But it still holds true for atmospheric computer models to this day. It is important that we as modelers understand that, while we have made great strides in our understanding of atmospheric processes, as well as computational power and speed over the years, models still have their limitations. That is, atmospheric processes are chaotic, and the model cannot always predict the pathway of a chaotic system. This is what we refer to as uncertainty in modeling, which is the extent to which the model may be off due to the chaotic nature of our atmosphere.

Application Activity 2.2: Understanding Chaos in the Atmosphere

Please see the resources webpage for an explanatory video about the Butterfly Effect and consider the following:

A. How can we measure uncertainty?
B. Why do we need to consider uncertainty in atmospheric models?

Check Your Reasoning

A. There are different statistical methods to estimate how far your results are from the observations. It is also a good practice to create an ensemble of simulations so that you can estimate how far from the observations your model results are. This will allow you to put a quantitative error bar in your visual results. We will discuss ensembles in the following section. You can look at Chapter 9, "Verification methods," and Chapter 7, "Ensemble methods," in Warner (2011) for more details or at the other information on the resources webpage.
B. Atmospheric models can be considered uncertain systems. If you start your simulation with slightly different initial conditions or on different days, you will have different solutions.

Ensembles

Earlier, we mentioned the idea of ensembles; but why is this useful? Ensemble forecasting is a numerical technique, which allows you to understand the uncertainty of a future prediction.

Instead of running just one simulation into the future, one would run several, where in each simulation one would change the initial conditions, or other parameters, to create the simulation.

The ensemble output can then be analysed and a final prediction can be made, which would include the uncertainty measure. By using ensembles, one would take into account different sources of uncertainty and a more informed prediction could be made.

Ensemble forecasting is used by the major forecasting centers, such as the United States' NOAA and Europe's ECMWF.

For more information about ensembles, including an interview with an operational meteorologist who uses them daily, please see the resources webpage.

Designing Experiments

In this section, we will be going through the process of designing an experiment. All of the ideas learned here will be important in upcoming chapters, where we will be applying these concepts to construct our design in WRF. As stated previously, a well-constructed research design can help minimize mistakes in your initial simulations, saving you lots of time. Additionally, we argue that well-designed research experiments are essential for achieving better results in your simulations, which means improved reports and/or better chance to have your project funded or published. If you publish your results in a peer-reviewed journal or are using them to obtain funding, you may be criticized for the type of design you have chosen. Spending time designing and thinking about your experiments before you actually run them will help avoid these criticisms.

It is also worth noting that designing experiments is directly related to experience working with a model. That is, the more experienced you are with a model, the more you will understand its strengths and limitations, which is key to designing a well-thought-out research experiment.

There are five sources that will be very important for you in your journey to learning best practises and getting better at working with weather or climate modeling. You can find more information about all of these, and more, on our resources webpage. To supplement this book, we highly recommend the following free resources:

- **The WRF Users' Guide**: Since e-WRF is a version of WRF, we recommend using the WRF Users' Guide to better understand how to run the model. This guide is updated each year when a new version of the WRF model is released and, like WRF itself, is free to use. The Users' Guide provides in-depth details you will need to design experiments and create sound simulations with WRF. It contains the options for parameterization schemes, as well as other details on parameters you need to set up weather and climate simulations. Guides for the most recent versions of WRF can be found here: https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/contents.html.
- **The WRF Technical Description**: This document explains the technical details on the model, such as the equations used and the numerical methods applied to WRF. This is an essential resource for setting up the configuration of your model and choosing parameters. It is in this document you will find the equations and details about how WRF

actually works. The most recent versions of the description can be found here: https://www2.mmm.ucar.edu/wrf/users/docs/technote/contents.html.

- **The Done et al. (2012) NCAR Technical Note**: If you plan to work with WRF in a tropical region, this note will be helpful in configuring your simulations. It can be found here: https://opensky.ucar.edu/islandora/object/technotes%3A535/datastream/PDF/view.

If you are looking for additional resources, these two books take a much deeper look at the conceptual and mathematical processes that make up atmospheric models and would be a great next step once you have finished this book and are ready to learn more.

- ***Numerical Weather and Climate Prediction*** by Thomas Tomkins Warner (2011): You may have already noticed, but we will reference Warner often throughout this textbook. Warner was ahead of his time when he argued for proper education of future climate modelers, and we are inspired by his perspective. This book provides the theoretical underpinnings of limited-area modeling, as well as several examples on real applications with models, such as WRF.
- ***Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*** by David Stensrud (2007): This book is one of the most complete references to parameterization schemes that is available. We will discuss parameterizations in WRF in this book, but it will be highly specific to WRF. For a more detailed look at parameterizations for a broader range of models, this is a good resource.

Always remember that the best way to learn how to run a model is to practise, practise, and practise. According to Malcolm Gladwell in the book, *Outliers: The Story of Success*, if we want to be good at something, we need to spend about 10 000 hours working on that. This is true for modeling the same way it is to become a proficient writer or learning to speak a new language. Study is important, but at the end of the day, the very best modelers, writers, and speakers have spent a lot of time doing it and making mistakes. For this reason, we emphasise that you *must* practice to become a proficient modeler. Use your computing resources wisely, but practice your simulations as much as possible.

Even if you end up not using WRF specifically in the future, remember that the conceptual understanding will transfer, and you will be much better prepared to learn a new model if the foundations have already been laid. This book is designed to do just that!

Model Parameterizations

To understand parameterizations, we first must discuss what a model does. According to Schwarz et al. (2009), models are *representations* that *explain* and *predict* a natural phenomenon. In the atmospheric sciences, accuracy of the model's prediction is directly related to how well it *represents* atmospheric processes, which can be adjusted by choosing the best parameterization schemes for your research question. Every choice made changes the outcome of the model simulation. As a result, parameterization schemes are one of the most important aspects to consider when setting up a computer model.

Unfortunately, choosing the best parameterization schemes is also one of the most difficult steps of the model setup procedure because there is no single "best combination;" it depends entirely on the research question you chose, the location and size of your domain, and the resolution of your domain. Even after considering these options, some choices may still not be clear. It is possible to make an informed decision on which parameterizations to use if you understand how the different schemes influence the model.

What do parameterization schemes do?

From our current understanding of atmospheric processes, atmospheric scientists have derived several equations that describe dynamic and physical processes. In a basic sense, weather is created when there is uneven heating, creating regions of relatively warm and cool temperatures, which causes air to move or be transported. On a spherical planet with no water or vegetation, this simple model description is probably pretty representative of how the atmosphere behaves. However, on a planet like Earth, more equations are needed in order to take into account things like oceans, mountains, ice, plants, and animals.

In fact, there are so many different influences and interactions to consider, that many equations are needed. Therefore, parameterization options are generally broken down into categories. In general, parameterizations use mathematical formulas, derived from theoretical understanding of atmospheric processes, to calculate values for variables of interest. Stensrud (2007) summarizes most schemes into the categories of: land surface, atmosphere interaction, water-atmosphere interaction, planetary boundary layer and turbulence, convection, microphysics, and radiation. WRF uses similar distinctions, but organizes its parameterization schemes using the following structure:

1. Physics Options
    a. Microphysics (mp_physics)
    b. Longwave Radiation (ra_lw_physics)
    c. Shortwave Radiation (ra_sw_physics)
    d. Cloud fraction option
    e. Surface Layer (sf_sfclay_physics)
    f. Land Surface (sf_surface_physics)
    g. Urban Surface (sf_urban_physics)
    h. Lake Physics (sf_lake_physics)
    i. Planetary Boundary Layer (bl_pbl_physics)
    j. Cumulus Parameterization (cu_physics)
2. Diffusion and Damping Options
3. Advection Options
4. Lateral Boundary Conditions

As you can see, there are many categories with at least two to three scheme options per category. The result is literally hundreds of thousands of potential combinations to choose from.

A Note about Terminology

We defined **parameterization** schemes in the previous section. These are different from model **parameters** that define your model experiment. These parameters are things like resolution, geographic projection, and number of nests, all of which will be discussed later in this chapter.

Advantages and Disadvantages of Parameterization Schemes

The equations that make up parameterization schemes in WRF range from very simple to very complex. In general, the more complex options are much more comprehensive and provide more precise and accurate results. This begs the question: Why not use the most complex, in-depth parameterizations every time, if they yield more precise results? There are several reasons, but the best explanation is: computation time is expensive.

In order to get an accurate representation of the region you are modeling, you want to create a domain that is as large as possible with the highest resolution possible; meaning, as many grid points as possible. The problem is that the model solves all the equations at every grid point for every time step. Therefore, if the model runs for a long time period, for a large domain, or for a high resolution, the computer is doing a lot of calculations. This can take a very long time to do or requires lots of processors, which are expensive. Therefore, sacrifices have to be made based on the question being researched.

Consider, for example, a climate model versus a weather model. Weather models tend to have complex parameterizations that focus on small-scale processes and, therefore, also tend to have high resolutions. The goal of weather models is to precisely and accurately forecast day-to-day weather. Since the resolution is high and parameterizations are complex, the models are run over a very small domain in order to be sure the model run-time is reasonably short. After all, what good is a 24-hour forecast if the model takes 48 hours to run?

Climate models, however, have a different goal. The goal is to accurately indicate *trends*, rather than specific values, over a very long period of time and a very large domain, often the entire planet. Therefore, parameterization schemes focus on large-scale processes and resolution is generally quite low. In the case of climate models, it is okay if the model takes several days to complete a run, since the model is usually looking at several decades into the future.

Regional climate modeling, as is frequently done using WRF, requires some combination of both processes and is why nesting is such an important part of setting up the model domain. The larger, coarser nest calculates the climate influence on the region (using large-scale, less-complex processes) and the smaller, high resolution nest calculates the weather as influenced by the climate for that region (using small-scale, more-complex processes). Therefore, parameterization schemes must be chosen with both of these ideas in mind. We will discuss nests in more detail later in this chapter.

Consider the following hypothetical research question: Does urban development impact the wind over a wind turbine farm in my domain?

In order to choose the best parameterizations for the model run, you have to think about the question being asked. First, the dynamic processes that calculate small-scale wind and speed relating to radiative heating are important and complex, so in-depth parameters should be chosen. Similarly, small-scale dynamical processes and accurate representation of orography and urban structures require small grid spaces, so a high resolution is also required. However, small-scale microphysical processes that determine different concentrations of various water particles in the atmosphere is not an important component of this question. So, using a less-complex microphysics scheme will save computing power and likely will not impact the answer to your research question.

While choosing the most complex and comprehensive parameterization option may seem like the best choice, it is important to consider the resolution and size of your domain as well as the time period of your model run. When you consider your research question, be sure to ask yourself:

- What length of time am I interested in studying: long-range or short-range time scales?
- What variables am I interested in looking at: are they small-scale or large-scale?
- Based on my variables of interest and the size of my region, do I need to use a fine resolution or is a coarse resolution okay? Is nesting an appropriate option?
- Are my variables of interest the result of primarily dynamic processes or microphysical processes?
- Which parameterization category options are most important? Which are least important?

For more information about parameterizations, please see the resources webpage.

Scientific Question Considerations

Before starting any research experiment, the first thing you need is to think about the question you are trying to answer. This is extremely important because your entire experimental design will be based on this question and how you will go about answering it. Not only does it help you design the experiment, it also tells you how (and even if) a model run is necessary.

Since your question is the central part of your entire experimental design, it is imperative that you develop a quality question. How do you know if you have a good scientific question? There are a few things to consider.

First, the topic must be of interest to you. This means you should think about topics that are valuable to you and that you have an invested interest in.

Second, they should not have straightforward answers. Questions whose answers can be solved by simple observation are not necessarily the best questions to design an entire experiment around. While questions like those may be important and have value to the scientific community, they are not scientific in nature because they do not require the process of collecting data and then drawing evidence from that data. The process of collecting evidence and drawing conclusions is what makes a question scientific and is important for a quality research experiment.

Third, science questions should be specific with only one possible answer. A great example of a question that is too broad is, “What makes a seed grow?” There are lots of possible answers to this question and many different ways to go about answering it, so it needs to be broken down into something more specific. This is where we think about what topic is specifically of interest to us. A better, more specific, science question might be “How do variations in light sources impact how a seed grows?” Ask questions that consider impact, or effect, or relationships. These help narrow your questions to something that can lead to a well-designed experiment.

Finally, a good science question considers only one variable at a time. What is the one element of this natural process that we want to study? What adjustments should be made? How can we control other variables so that they are consistent and do not compromise the data? This is where modeling is extremely useful! So long as we ask a good question and choose our model setup appropriately, it is a great way to study the impact of one particular variable on the atmospheric system.

For more information on the scientific process, science inquiry, and specifically model-based science inquiry, we recommend taking a look at Yarker (2013) on the resources webpage.

The process for determining a scientific question can be summarized and referenced in Table 2.1 below. Now that you have thought about it, we need to begin designing the experiment, which is how we will set up the model to answer this question.

Table 2.1: Scientific Question Considerations

1. What is the specific process or event I am interested in studying?
    a. What do I currently know about this process or event?
    b. Where can I go to learn more about it?
    c. What were or what could be the impacts from this process or event?
2. What is the one variable I am most interested in studying?
    a. What is the relevance of this specific variable on this process or event?
3. Do I need a model to study the impact of this variable?
    a. If so, what type of model is best?
    b. Am I analyzing the influence of this variable on past observations or future predictions?
    c. Is it directly observable through other means?
4. What is my scientific question?

Experimental Design Considerations

Limited-Area Models (LAMs)

So far in this book, we have talked quite a lot about experiment design and how to prepare yourself for a model run, but we have not yet actually defined or described the type of model WRF is. In Warner (2011, p.96), he refers to most kinds of sub-climate models as limited-area models or LAMs. He defines them as "everything but global models" (Warner, 2011). The difference between an LAM and a global climate model lies entirely in the equations used to run the model. As the name suggests, a global climate model solves a set of equations that describe large-scale processes that encompass the entire planet. This is different from an LAM, which processes a different set of equations that describe much smaller, more detailed processes in the atmosphere. Why the difference? It all has to do with computation time.

As we have mentioned before, running numerical models is expensive computationally. It is for this reason that global models have a very coarse resolution, on the order of hundreds of kilometres. While newer generations of climate models have higher resolutions, the resolution is still large enough that smaller-scale processes are still too small for the model to see because they happen within one grid cell and are essentially invisible to the model itself.

As we make grid sizes smaller, the amount of computational resources needed increases, until eventually it takes so long to run the model, it is essentially useless. Imagine if we tried to run a climate model with an extremely fine resolution -- it would take years to finish computing. How useful would a 30-year projection be in a case like this? Not very useful at all. This is, therefore, why we are constrained by the availability of our computing resources.

However, this does not mean that we cannot use global climate models to provide more detailed information about a smaller region. There are a few ways we can use global climate model simulations to inform more detailed processes at a smaller scale. Two common examples are:

- **Statistical LAMs**: This is when statistical equations are used in order to build high spatial resolution of a variable of interest.
- **Dynamical LAMs**: This is essentially a numerical model, where a set of defined dynamic and thermodynamic equations are used to simulate the weather or climate in high resolution.

Since WRF is a dynamical model, a **dynamical LAM** is what is used when running a regional climate simulation. To get around the computational limitations, global equations are used for the entire planet, but nested within the global model is a limited, defined area where resolution is much higher and the more detailed equations are used. This is what is commonly referred to as **regional climate modeling**.

Domain Size

Now that we know the kind of model we are using to answer this science question, we know that we are working with WRF, an LAM. Since it is not a global model, this means we have to define how big the region is we are studying; that is, what is the size of the model domain?

To define the size of your domain, you must first consider the **scale** of the process or event you want to study. Your domain needs to be big enough to not only fit the entire process within it, but also big enough to capture a reasonable amount of area around it, as well. For example, to study Cyclone Nargis, which made landfall in Myanmar in 2008, you need a domain that is not only big enough to fit the entire cyclone inside the domain but also the area around it that is of interest to you. If you are only interested in the moment Cyclone Nargis made landfall, you can keep your domain fairly small and focused on the region where landfall occurred. However, if you want to study something about the path it took before it made landfall, you must increase the size of your domain to fit the entire path of the cyclone.

In addition to the scale of the event, we also must take into consideration **model limitations**. If your domain is northern India, for example, you can estimate the centre of northern India and find a latitude/longitude point there using Google Earth. You would then use that location as the centre of the domain when defining it in WRF, which we will detail later in this book. One of the questions that always arises when working with India is: Should you include the whole Himalayan region or not? This is something you would need to test and would be dependent on the scientific question you are trying to answer. However, you should not select your domain boundary in a way that it cuts across any mountain range because this could potentially cause instability in the model simulation. Either you include the whole mountain range or you exclude it.

Alternative to Designing a Domain

There are other alternatives to designing the domain. For instance, you could borrow WRF scripts from well-tested experiments. Or you could easily create a domain using the WRF Domain Wizard: https://esrl.noaa.gov/gsd/wrfportal/DomainWizard.html.

However, it is our belief that the best way to learn how to create a design is by trying it yourself, creating the scripts on your own. Our activities were designed to help you achieve that. This will also make you a more independent user of the model.

Vertical and Horizontal Resolution

Once the domain is chosen, the next thing to consider is resolution. This is an important decision that requires careful consideration of:

- the size and scale of your event and the variable you are studying
- the size of your domain
- the amount of computing power you have access to

While we usually think about resolution as being an East/West and North/South configuration (i.e., horizontal), resolution must also be considered in the up/down (i.e., vertical) orientation, as well. One aspect many people forget is that if you increase the horizontal resolution, you most likely need to increase the vertical resolution, too. However, the main problem with this is that increasing the vertical resolution will also increase the amount of time it will take to run the model. Your project will determine if both your model resolution and run-time are appropriate.

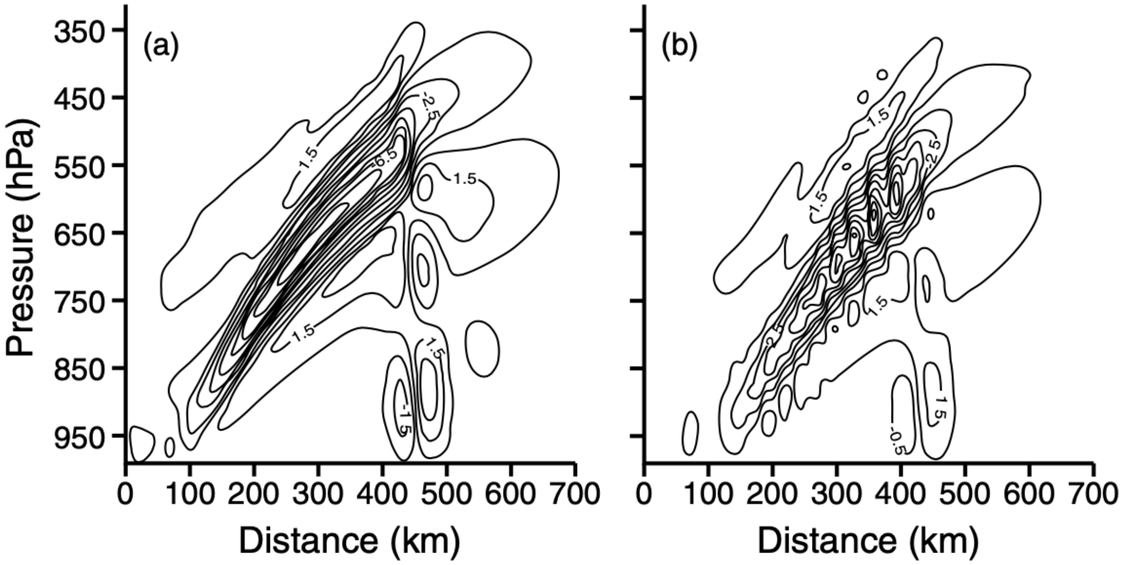


Fig. 3.16 Vertical cross sections of vertical velocity, ω (solid lines, $\mu b\ s^{-1}$) after 24-h simulations of conditional symmetric instability with a 10-km horizontal grid increment and 75 layers (a) and 25 layers (b). From Persson and Warner (1991).

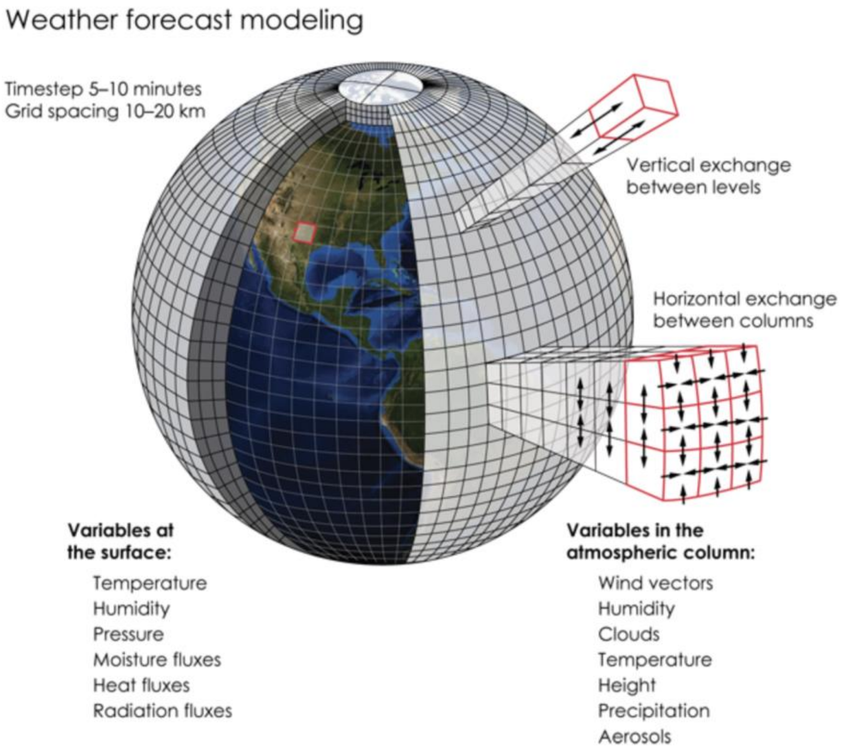
Figure 2.1. Comparison of vertical velocity simulations using (a) 75 vertical layers and (b) 25 vertical layers (Warner, 2011).

Take a look at Figure 2.1. Which panel gives a more realistic representation of the process: panel a or panel b? The “ripples” in panel b indicate some sort of error in the model simulation due to how the domain was set up. Panel a seems to be a much better representation of the atmosphere because it shows a smooth, more realistic, and expected visualization of the process. Warner (2011) explains that these figures are the result of a model simulation run at a 10-km horizontal resolution. However, panel a was run with 75 vertical layers, and panel b was run with only 25 vertical layers. This is an excellent depiction of how crucial the vertical resolution can be for simulating certain processes more accurately.

How do you know what resolution is best for your study? It depends on a number of different factors, some of which are illustrated in Figure 2.2. Questions you should ask yourself are:

- **What type of system do you want to model?** Examples include a hurricane, lake effect snow event, or next week’s precipitation outlook.
- **What physical processes are most important to understand your study?** Dynamical processes? Microphysical processes? Synoptic flow?
- **What computational resources are available to you?** Do you have access to a supercomputer or several processors? What is the time scale of your simulation, and how much time do you have to run it?

Figure 2.2. This diagram illustrates some of the variables to consider when creating an



experiment, paired with both horizontal and vertical grid boxes (From The Weather Guys at the Cooperative Institute for Meteorological Satellite Studies/Space Science and Engineering Center of the University of Wisconsin-Madison: http://wxguys.ssec.wisc.edu/2019/03/04/models/).

If your goal is to understand the intensity of a specific hurricane, you will probably need a higher resolution in order to capture the necessary details. You will obviously want as fine a resolution as possible, but it still must be within the bounds of: 1) how much computing power you have access to and 2) the scale of the physical process that you are trying to study. For example, most synoptic processes are quite large, and, therefore, can easily be captured with a resolution of 20 kilometers. With reasonable computing power, simulations of these processes can run over the period of several days to over a week. However, if you are studying a lake-effect snow event, now your interest is on microphysical processes of cloud development, which will need a much smaller scale to be seen. Here, it is necessary to choose a resolution less than 1 kilometer, and, therefore, the time scale of the simulation should be much smaller -- about a day -- to be within reasonable use of computational power.

Once you have chosen the boundaries of your domain, as discussed in the previous section, you can also use the following figure to determine the ideal resolution for your domain (Warner 2011).

Identify the curve that best matches the latitude of the centre of your region of interest. **The best**

match is when the curve approximates a scale equal to or less than 1.0 (the scale is located on the y-axis).

Tip: Remember that when you choose your horizontal and vertical resolutions, you will need to test the model to make sure that those resolutions are able to capture the features you want to study and that your model will run in a reasonable amount of time.

For more information about model resolution, please see the resources webpage.

Lateral Boundary Conditions (LBCs)

As mentioned previously, WRF is an LAM. At this point, we have discussed how to set physical boundaries of our domain. Now, we need to establish the conditions along the boundaries so that the model has data to initialize the simulation.

Initial conditions are the current atmospheric conditions in the domain, which the model uses to solve its equations and integrate forward in time. These are the starting point for the model, or how the model knows the state of the system at the start of the simulation. This is not just for a point in time; the model also needs information at the boundaries of the domain. This is because WRF will start solving equations at a grid point at the edge, then use that information to solve the equations for the grid point next to it, and so on until it has solved all the equations for all the grids in the entire domain. That is just the first timestep. Then it integrates forward in time and repeats the process.

Where do we get LBC data from? For regional climate modeling, the data come from global climate models. For example, if we want to run a high-resolution weather forecast for the city we live in, we should do the following:

- Download data from a global model that was run for the next few days.
- Use these data to provide the initial conditions and LBCs to the WRF model.

This is exactly what many weather centres around the world do. In fact, if you want to open your own weather forecasting centre using WRF, you can.

For all other types of model simulations, reanalysis data from satellites are frequently used for initial conditions, which are essentially observation data. For more information about reanalyses, please see the resources webpage.

Tip: In order to avoid errors due to LBC interpolation, it is recommended that you place your region of interest in the centre of your domain

For more information about LBCs, you can read Warner (2011, pp.96-113) or see the resources webpage.

Projections

Another important consideration is your choice of projection for your simulation. In fact, it can be a contributing source of any error in your simulation if inappropriately chosen. For example, if you are using WRF in a polar region, between 60 and 90 degrees latitude, you might choose a **polar stereographic** projection. This type of projection would reduce errors for high latitudes. If you are working in a tropical region, between 0 and 25 degrees latitude, you might choose a **Mercator** projection. If your region of interest is the middle-latitudes, between 25 to 60 degrees latitude, you might use the **Lambert conformal** projection. Figure 2.3 compares these common geographic projections for atmospheric applications.

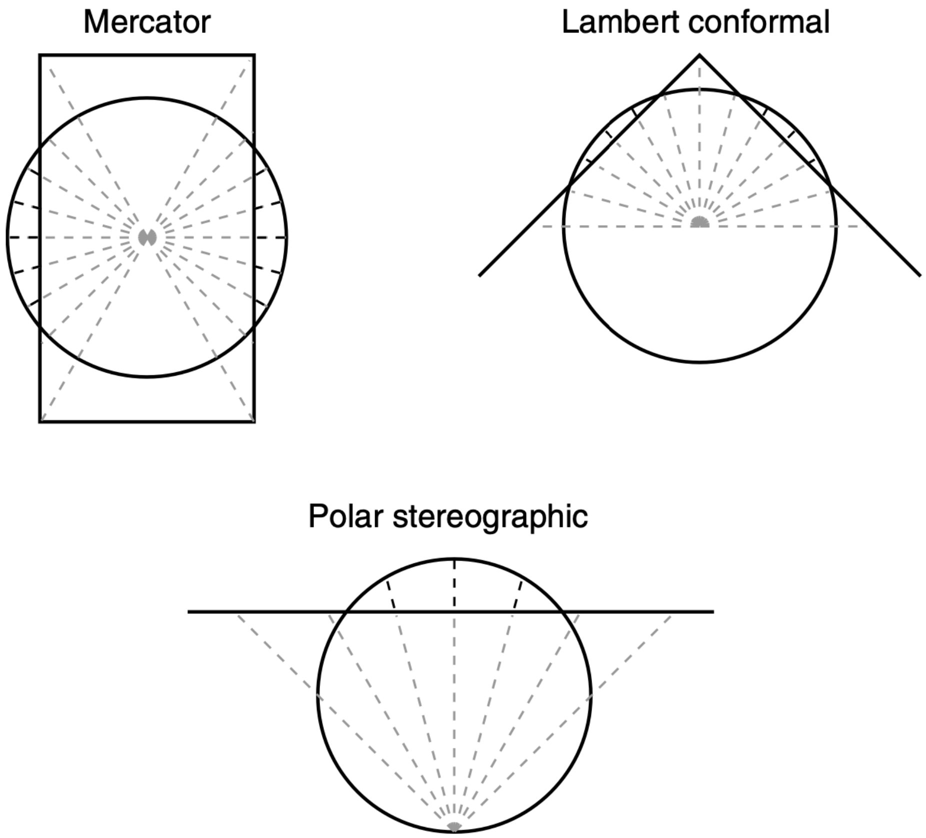


Fig. 3.3 Three map projections commonly used in atmospheric modeling. The cylinder (Mercator), right-circular cone (Lambert conformal), and plane (polar stereographic) are the surfaces on which the information on the sphere is projected. The radial lines connect points on the sphere and points on the projection surface. The axes of the cylinder and the cone, and the perpendicular to the plane, are parallel to Earth's axis of rotation. In these images, we are thus viewing Earth from over the Equator.

Figure 2.3. Comparison of the common geographic projections used in atmospheric applications (Warner, 2011).

For more information about these and other types of projections, please see the resources webpage.

Nesting and the 3:1 Ratio

There are instances where your question requires a large domain but also requires a fine resolution to capture smaller atmospheric processes. Computationally speaking, it is likely impossible to run an LAM at a fine resolution across a large domain. It will either require too much power or will take so long that the data will be useless. However, to deal with this problem, we can nest a smaller domain (with higher resolution) inside a larger domain (with a coarser resolution). This is essentially how regional climate models work and is illustrated by Figure 2.4.

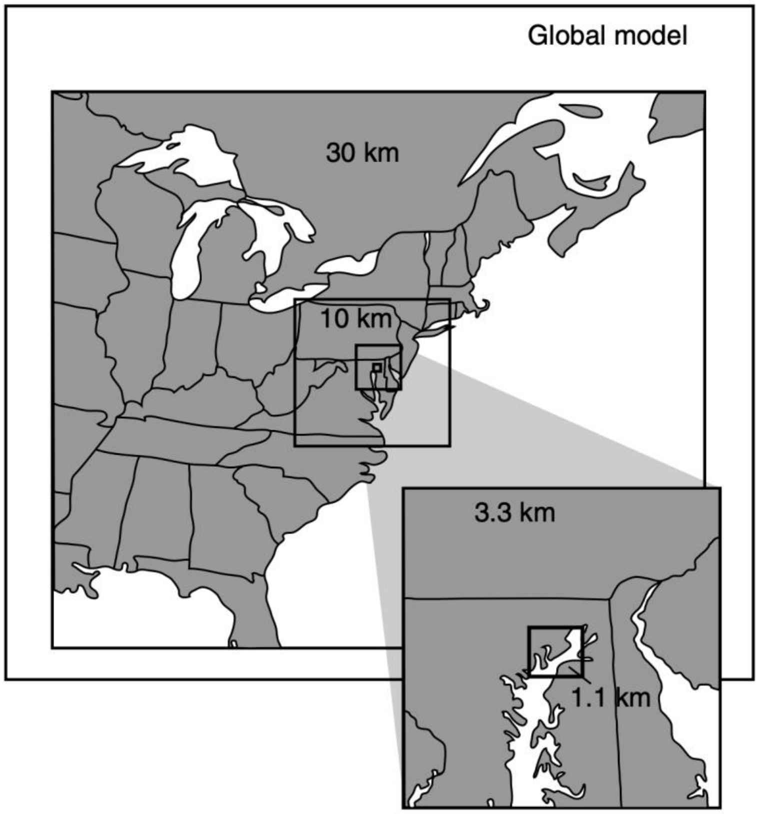


Fig. 3.13 An example of a nested-grid model used for operational prediction. The model with the two-way interacting grids is embedded within a global-model prediction. Model grid increments are indicated. From Liu *et al.* (2008a).

Figure 2.4. Diagram demonstrating nesting options for an LAM (Warner, 2011).

Suppose you get some data from a global model with horizontal resolution of 100 km. You now want to use WRF to produce high-resolution information for your city. You need information at 10-km resolution. So, you want to use the global model output as LBCs in your WRF simulation. **Can you nest the smaller 10-km domain directly into the larger 100-km domain?** This is an important question to consider because answering it requires that we understand how WRF gets information from the global model and passes it to its domain. When using WRF, it will pass that 100-km information coming from the global model into the smaller domain by interpolating the information from the larger domain at the boundaries of the smaller domain.

So, what would you expect your results to be if you interpolated information from 100 km to 10 km? Well, this is a large "jump" from coarse to fine resolution, and the results may not be very accurate. In addition to that, the interpolation process introduces errors, shown in Figure 2.5 (Warner, 2011).

Figure 2.5. This figure illustrates the errors introduced by the interpolation from a global model into WRF (Warner, 2011).

This figure was constructed by first running a global model at a certain horizontal resolution. Then, the authors ran an LAM (like WRF) at the same resolution. After that, they calculated the difference between the two, which shows the errors due to the interpolation process. As you can see, these errors propagate in the domain, but normally the centre area is not affected that much compared to the northern or southern parts of the domain.

In order to avoid these errors we discussed here, it is recommended that you use the 3:1 ratio, which is how you will determine the resolution of each nest. If you have global climate model information at 100 km and you would like to downscale information to 10 km, you would need to do this in steps:

1. Downscale from 100 km to 33.33 km (using a ratio of 3:1 -- 100 km divided by 3).
2. Then, downscale from 33.33 km to 11.11 km (using a ratio of 3:1 -- 33.33 km divided by 3).

If you needed to use a finer resolution, you could continue dividing by 3 until you reach your desired resolution. For example:

3. Continue downscaling from 11.11 km to 3.70 km (using a ratio of 3:1 -- 11.11 km divided by 3).
4. Downscale again from 3.70 km to 1.23 km (using a ratio of 3:1 -- 3.70 km divided by 3).

Tip: If you want to use an exact resolution of 30 km, 10 km, 3 km, or 1 km, you could round these numbers. For instance, you could downscale to 30 km and then to 10 km.

Some examples of when you would likely want to use nests include:

- Hurricane simulations
  - Nests will allow you to resolve both the smaller- and larger-scale processes crucial to hurricane evolution and still be efficient with computing resources.
  - Moving nests are used in forecasting the potential tracks and intensities of hurricanes more accurately.
- Turbulence simulations
  - Nests will allow you to resolve processes at multiple scales when studying turbulence (Daniels, 2016).

- These events occur in one or more specific layers of the atmosphere, and nests can be used in the vertical dimension, as well (Daniels, 2016).

You will learn more about nesting later in this course, especially when it comes to setting up nests in WRF. For more information about nesting, please see the resources webpage.

Table 2.2: Experimental Design Considerations

1. How large is my region and event of interest?
    a. How big can I reasonably make the domain so that I fully capture both the area of interest as well as the event?
    b. Are there any geographical features I should consider including or excluding in the domain?
2. What resolution and geographic projection will best represent my data?
3. Based on the resolution of the global boundary data and the resolution of my model, do I think I should embed nests into the domain?
    a. How many?
4. How will I validate the model?
    a. How will I obtain observation data, and how will I compare it with the model output?
    b. What kind of statistical analysis will I use to evaluate how well the model captured reality for my region?
5. What will I do if the model setup is not a good representation of my region?
    a. What changes may I need to make to the domain, resolution, input data, or parameterizations?
6. Once I have set up the model in a way that is a good representation of the domain, how will I set up the experimental simulation?
    a. How will I process the output data to obtain my results?
7. How long will it take me to perform the above steps?
    a. Is it reasonable to perform this experiment within any time limitations I have?

Table 2.2 summarizes the experimental design considerations that you should think about at the beginning of and throughout your experiment. These can be easily referenced along with the scientific considerations in Table 2.1.

Tips

Be sure to spend extra time researching the set up of your model simulation. Do not be afraid to run a few very short runs to do quick and dirty evaluations of how well your configurations capture reality.

Sometimes it is not necessary to do your own model evaluation if it has already been done. Be sure to do your background research to see if a team has published results for the model you plan to use on your region of interest. If so, see if it is possible to set up your model to fit their domain and parameterizations so that you can save yourself a lot of time! Additionally, you can take the time to evaluate several different model configurations.

One of the most thorough climate studies using the WRF model was the work by Mooney et al. (2013), published in the Journal of Climate: "Evaluation of the Sensitivity of the Weather Research and Forecasting Model to Parameterization Schemes for Regional Climates of Europe over the Period 1990–95." The authors made a solid experimental design and tested the model for different configurations. They were then able to make decisions on which of these they would choose as representative for their region of interest. This paper is also an excellent example of a very well-designed experiment and it offers several great research questions, which you can use as inspiration for your own research questions.

To learn more about experimental design, we recommend Chapter 10 in Warner (2011): "Experimental design in model-based research."

Application Activity 2.3: Design Your Experiment

Choose an event for a region of interest to you. Consider the scientific questions in Table 2.1 to determine a science question that you can answer using e-WRF. Go through the experimental design considerations in Table 2.2 to start designing your experiment. It will be helpful if you record your answers and choices somewhere, as you will use this information moving forward to set up your experiment in e-WRF.

Self Reflection

Now that you have taken time to think about your scientific question and design your experiment, it is time to start running the model. Before we move on to Chapter 3, please ask yourself the following questions:

1. How might your study produce meaningful results for the scientific community?
2. Are there any aspects of your experimental design that you cannot yet determine until you know more about how e-WRF works?