

Utilización de datos satelitales para la evaluación y mejora de los pronósticos
numéricos en alta resolución a muy corto plazo

A Thesis
Presented to
The Division of Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

In Partial Fulfillment
of the Requirements for the Degree
Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en
el área de Ciencias de la Atmósfera y los Océanos

Paola Corrales

Algún momento de marzo 2023?

Approved for the Division
(Departamento de Ciencias de la Atmósfera y los Océanos)

Victoria Galligani

Juan Ruiz

Acknowledgements

I want to thank a few people.

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

Table of Contents

Introducción	3
0.1 Pronostico de eventos severos	3
0.2 Asimilación de datos como posible solución	4
0.3 Resultados previos de asimilación de distintas fuentes de observaciones	6
0.3.1 Asimilación de radianzas de satélites	8
0.4 Asimilación de datos en Sudamérica	11
0.5 Motivación y objetivos	12
 Chapter 1: R Markdown Basics	 13
1.1 Lists	13
1.2 Line breaks	14
1.3 R chunks	14
1.4 Inline code	14
1.5 Including plots	15
1.6 Loading and exploring data	16
1.7 Additional resources	18
 Chapter 2: Mathematics and Science	 21
2.1 Math	21
2.2 Chemistry 101: Symbols	21
2.2.1 Typesetting reactions	22
2.2.2 Other examples of reactions	22
2.3 Physics	22
2.4 Biology	22
 Chapter 3: Graphics, References, and Labels	 23
3.1 Figures	23
3.2 Footnotes and Endnotes	26

3.3	Bibliographies	26
3.4	Anything else?	28
	Conclusion	29
	Appendix A: The First Appendix	31
	Appendix B: The Second Appendix, for Fun	35
	References	37

List of Tables

1.1 Max Delays by Airline 17

List of Figures

1	Esquema de un ciclo de asimilación típico. El tiempo de las observaciones y el campo preliminar deberá coincidir.	4
3.1	Reed logo	23
3.2	Mean Delays by Airline	25
3.3	Subdiv. graph	26
3.4	A Larger Figure, Flipped Upside Down	26

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Dedication

You can have a dedication here if you wish.


```
knitr::opts_chunk$set(  
  echo = FALSE,  
  message = FALSE,  
  warning = FALSE  
)  
library(here)
```

here() starts at /home/paola.corrales/tesis_doctorado

Introducción

0.1 Pronostico de eventos severos

La simulación numérica de la atmósfera, es decir, la integración de las ecuaciones que rigen la evolución del sistema atmosférico es la base para la predicción del tiempo en diversas escalas temporales desde horas a semanas.

La predicción de fenómenos meteorológicos extremos es de particular importancia ya que pueden producir cuantiosas pérdidas humanas y materiales. En Argentina, una gran cantidad de estos fenómenos están asociados a la ocurrencia de convección profunda entre los que se cuentan tornados, ráfagas intensas, precipitaciones extremas en cortos períodos de tiempo, granizo de gran tamaño y actividad eléctrica. Es por tal motivo necesario avanzar en el conocimiento de estos fenómenos y en la capacidad de pronosticar la ocurrencia de los mismos.

Si se cuenta con condiciones de borde apropiadas, es decir, una correcta representación de las características de la superficie terrestre y el tope de la atmósfera, la integración de un modelo atmosférico es un problema de condiciones iniciales. La generación de pronósticos de calidad dependerá entonces, de la capacidad del modelo para representar los procesos atmosféricos y la exactitud de las condiciones iniciales usadas (Kalnay, 2002).

El pronóstico de los fenómenos severos es a su vez un desafío científico y tecnológico muy complejo debido a la predictibilidad limitada en la mesoescala y debido a la dificultad de conocer o diagnosticar el estado de la atmósfera en escalas espaciales pequeñas y tiempos cortos (por ejemplo de 1 a 10 km y del orden de los minutos).

Uno de los métodos que pueden utilizarse para el pronóstico de fenómenos meteorológicos severos es la utilización de modelos numéricos de la atmósfera que resuelvan explícitamente la convección profunda. Diversos estudios, han comprobado que estos modelos agregan valor al pronóstico a corto plazo y que en muchos casos proveen información sobre el modo de organización de las celdas convectivas y su intensidad (Aksoy, Dowell, & Snyder, 2010; Stensrud et al., 2013). No obstante, la capacidad de

los modelos numéricos en anticipar la ubicación y tiempo de ocurrencia de eventos extremos asociados a convección es muy limitada si no se cuenta con una detallada información sobre el estado de la atmósfera en la escala de las tormentas en el momento en el que se inician los pronósticos numéricos (Clark, Gallus, Xue, & Kong, 2009).

0.2 Asimilación de datos como posible solución

Por otro lado es posible aplicar técnicas de asimilación de datos para generar una mejor estimación de las condiciones iniciales necesarias para integrar un modelo numérico. La asimilación de datos combina de manera optima un pronóstico numérico o campo preliminar en un tiempo t con las observaciones disponibles para ese mismo tiempo, generando un análisis. Esta combinación optima toma en cuenta el error asociado a al modelo meteorológico (errores de pronóstico) y el error de las observaciones (instrumental, de representatividad) y si ambos tienen una distribución Gaussiana, el error resultante será menor a los errores originales. Por esta razón el análisis es considerado la *mejor aproximación* disponible del estado real de la atmósfera.

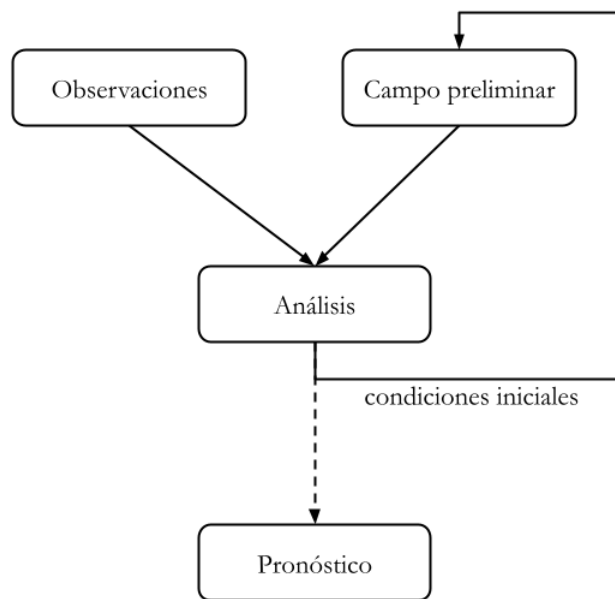


Figure 1: Esquema de un ciclo de asimilación típico. El tiempo de las observaciones y el campo preliminar deberá coincidir.

En el caso de modelos globales, típicamente cada ciclo de asimilación de 6 horas utiliza el campo preliminar previo, es decir el pronóstico a 6 horas inicializado a partir del análisis anterior y las observaciones disponibles para las 6 horas previas o en

un periodo similar centrado en la hora del análisis. Para poder comparar el campo preliminar con las observaciones, este es interpolado a la ubicación de las observaciones. En determinados casos, por ejemplo cuando se trabaja con observaciones de satélite o radar), será necesario transformar las variables del modelo para obtener las variables observadas. En la siguiente ecuación H es el operador de las observaciones que se encarga de las interpolaciones y transformaciones necesarias sobre el campo preliminar x_b .

$$x_a = x_b + W[y_o - H(x_b)]$$

La diferencia entre las observaciones y_o y el campo preliminar se denomina innovación. El análisis x_a se obtiene aplicando las innovaciones al campo preliminar teniendo en cuenta un peso W que incluye información sobre los errores del pronóstico y de las observaciones.

Existen diferentes metodologías para obtener x_a . Los métodos variacionales, 3D-Var y 4D-Var, definen una función de costo que es proporcional a la distancia entre el análisis y simultaneamente, el campo preliminar y las obvservaciones. Esta función de costo J es minimizada para obtener el análisis.

$$J = \frac{1}{2}[y_o - H(x_a)]^T R^{-1}[y_o - H(x_a)] + (x_a - x_b)^T B^{-1}(x_a - x_b)$$

En la ecuación, el primer término corresponde a la distancia entre el campo del análisis y las observaciones, pesado por la covarianza de los errores de las observaciones R . El segundo término a la distancia entre el campo del análisis y el campo preliminar pesado por la covarianza de los errores del pronóstico B . Para el caso más simple, es decir, una variable de modelo y una observación, R y B son escalares. Para el caso real, serán matrices de covarianza de dimension XXX que deben ser estimadas. Las ecuaciones 1 y 2 son análogas si se define a W como:

$$W = BH^T(HBH^T + R^{-1})^{-1}$$

El método 4D-Var extiende el uso del método 3D-Var para incluir la distancia temporal a las observaciones dentro de la ventana de asimilación en la misma función de peso.

Los métodos secuenciales y en particular el filtro de Kalman extendido, actualizan el análisis a medida que las observaciones están disponibles. Este método tiene la ventaja de actualizar la matriz B junto con el análisis. En este caso la matriz W toma el nombre de K o ganancia de Kalman y se actualiza en cada ciclo de asimilación t_i .

$$K_i = B(t_i)H^T(HB(t_i)H^T + R^{-1})^{-1}$$

La estimación de B es particularmente costosa en terminos computacionales por lo que en la práctica se utiliza el filtro de Kalman por ensambles o EnKF. Un ensemble consiste en un conjunto de simulaciones ligeramente diferentes que se resuelven simultaneamente para incluir los posibles estados de la atmósfera y provee información dependiente de la dinámica durante la ventana de asimilación. A partir del ensemble, la matriz B se estima como:

$$B \approx \frac{1}{m-1} \sum_{k=1}^m (x_b^k - \bar{x}_b)(x_b^k - \bar{x}_b)^T$$

donde $k \in [1, m]$ el miembro k -ésimo del ensemble. Esta estimación será buena si el ensemble logra capturar los posibles estados futuros o en otras palabras el spread se mantiene estable a lo largo de los ciclos de asimilación. Sin embargo, este método no es aplicable a menos que el tamaño del ensemble sea comparable a los grados de libertad de un modelo que resuelve 10^9 variables de estado, lo que resulta computacionalmente inviable.

El método Local Ensemble Transform Kalman Filter (LETKF) busca resolver los problemas anteriores restringiendo el área de influencia de las observaciones a un determinado radio de localización reduciendo el costo computacional necesario. Además, calcula el análisis para cada punto de retícula uno a uno, incorporando todas las observaciones que puedan tener influencia en ese punto al mismo tiempo. De esta manera este método es hasta un orden de magnitud más rápido comparado con otros métodos desarrollados previamente (Whitaker, Hamill, Wei, Song, & Toth, 2008).

Independientemente de la metodología aplicada, el modelo cumple un rol fundamental en la asimilación de datos ya que *transporta* información de regiones donde existe mucha información disponible (por ejemplo, los continentes) a regiones donde las observaciones son escasas (zonas oceánicas) manteniendo los balances físicos que rigen los procesos atmosféricos.

0.3 Resultados previos de asimilación de distintas fuentes de observaciones

Para que los métodos de asimilación de datos tengan éxito, deben utilizarse redes de observación con suficiente resolución temporal y espacial capaces de captar la

variabilidad en las escalas que se quieren resolver, por ejemplo, la mesoescala.

Wheatley & Stensrud (2010) investigó el impacto de la asimilación de datos de presión de superficie en un sistema de asimilación de datos basado en conjuntos de mesoescala, pero encontró un impacto limitado en dos estudios de caso relacionados con sistemas convectivos de mesoescala. Ha & Snyder (2014) demostró que la asimilación de la temperatura y la temperatura del punto de rocío de las redes de estaciones meteorológicas de superficie de alta resolución mejoraba sistemáticamente la estructura de la capa límite planetaria simulada y mejoraba la previsión de precipitaciones de corto alcance sobre los Estados Unidos. Chang, Jacques, Fillion, & Baek (2017), Bae & Min (2022) y Chen, Zhao, Sun, Zhou, & Lee (2016) informaron sobre los efectos beneficiosos de la asimilación de observaciones de estaciones meteorológicas de superficie en un sistema de asimilación de datos de alta resolución utilizando las metodologías de EnKF, 3D-Var y 4D-Var, respectivamente, encontrando impactos positivos en el pronóstico de la temperatura y la humedad en la capa límite planetaria y en la localización de los sistemas de precipitación. Sobash & Stensrud (2015) demostró en un sistema de asimilación de datos de mesoescala que el impacto sobre la iniciación de la convección y el pronóstico de la precipitación de corto alcance es positivo si los datos se asimilan con frecuencia (en el orden de minutos, en lugar de en el orden de horas). Maejima, Miyoshi, Kunii, Seko, & Sato (2019) investigó el impacto de la asimilación con frecuencia de 1 minutos de observaciones sintéticas en un caso de precipitación intensa, encontrando que la asimilación de observaciones de alta frecuencia y espacialmente espacialmente densas conducen a una mejor representación de la circulación de mesoescala aunque el número de observaciones proporcionadas por las estaciones de superficie es mucho menor que el proporcionado por los radares meteorológicos. Gasperoni, Wang, Brewster, & Carr (2018) realizó un estudio de caso para evaluar el impacto de la asimilación de las observaciones producidas por estaciones meteorológicas privadas que no se incorporan a los análisis operativos globales. Encontró un efecto positivo al asimilar estas observaciones sobre el inicio de la convección húmeda profunda a lo largo de una línea seca. Este resultado es especialmente importante para regiones con pocos datos, como el sur de Sudamérica, donde las redes operativas no son lo suficientemente densas como para captar los detalles de la mesoescala. En ese sentido, Dillon et al. (2021) intentó utilizar por primera vez observaciones de estaciones meteorológicas automáticas de redes privadas en el sur de Sudamérica, sin embargo, la contribución específica de este tipo de observaciones sobre esta región, no ha sido investigada hasta ahora.

Se ha investigado el impacto de otros tipos de observaciones de resolución espacial

y temporal relativamente alta, como observaciones de satélites, en el contexto de la asimilación de datos de mesoescala. Estas observaciones incluyen radianzas y productos derivados, Wu, Liu, Majumdar, Velden, & Anderson (2014), Cherubini, Businger, Velden, & Ogasawara (2006) y Sawada et al. (2019) observaron un impacto positivo de la asimilación de viento derivado de información satelital de alta frecuencia en un estudio de caso de un ciclón tropical utilizando un sistema de asimilación de datos basado en ensambles. Por otro lado, Gao, Huang, Jacobs, & Wang (2015) encontró un impacto positivo en la asimilación de viento estimado a partir de las observaciones de satélites geoestacionarios.

0.3.1 Asimilación de radianzas de satélites

Uno de los objetivos y aporte original de este trabajo es la asimilación de radianzas para aplicaciones de mesoescala, por lo que en esta sección se resumirá los avances en la asimilación de estas observaciones a nivel global y regional.

Los primeros satélites de órbita polar en proveer información meteorológica fueron desarrollados en las décadas de los 60 y 70. Incluían sensores infrarrojos y de microondas para monitorear la temperatura y humedad. Hacia finales de la década de los 70, Estados Unidos, Europa y Japón ya habían lanzando los primeros satélites geoestacionarios. Pocos años después este tipo de observaciones se incorporaban al Sistema de Observación Global (Global Observing System en inglés).

El primer conjunto de satélites compuesto por los sensores HIRS, MSU y SSU (sistema TOVS) podían cubrir el globo completo cada 12 hs. Si bien cada uno de estos sensores generaba información complementaria en la tropósfera y baja estratósfera, la resolución horizontal y vertical era limitada. En particular HIRS, un sensor infrarrojo tiene una resolución horizontal de 40 km, mientras que MSU y SSU, sensores sensibles en las microondas, tiene una resolución de 160 y 200 km respectivamente. En la vertical, la función de peso de los distintos canales ronda entre los 5 y 10 km y aún en los casos donde los canales se solapan, la resolución apenas alcanza los 3 km.

Las primeras pruebas de asimilación de observaciones de satélites fueron desarrolladas principalmente en Australia, motivadas particularmente por la escasez de observaciones en el hemisferio sur. Kelly, Mills, & Smith (1978) mostró una importante mejora en pronósticos a 24 horas de altura geopotencial entre 1000 y 200 hPa cuando se asimilaba de manera continua perfiles de temperatura derivados del satélite Nimbus-6, conocidos también como retrievals. A nivel global Ohring (1979) resume los avances de la década indicando los impactos son positivos aunque pequeños y que

la mayor mejora se observa en los pronósticos en el hemisferio sur. Al mismo tiempo Ohring (1979) señala algunos de los posibles problemas asociados, por ejemplo la baja resolución vertical de los perfiles de las distintas variables y problemas en la generación de los mismos.

A principios de los 80 los centros de pronóstico mundiales continuaron estudiando la posibilidad de asimilar observaciones satelitales obteniendo resultados similares y tomando una mejora en la calidad de los perfiles de temperatura generados (Eyre, English, & Forsythe, 2020). En particular el ECMWF Seminar on Data Assimilation Systems and Observing System Experiments concluye que la asimilación de estas observaciones cumple un rol importante en el análisis de sistemas meteorológicos de larga escala en latitudes medias y altas, y en particular en el hemisferio sur. Sin embargo, hacia finales de los 80, los modelos de pronóstico habían mejorado sustancialmente haciendo que el potencial impacto de observaciones erróneas u observaciones asimiladas de manera incorrecta degradaran sustancialmente el pronóstico particularmente en el hemisferio norte. Andersson et al. (1991) mostró que los incrementos en el análisis presentaba patrones con importante sesgo cuando se asimilaba retrievals de TOVS.

Eyre et al. (2020) explica que la principal razón por la que los resultados obtenidos no fuera bueno era que se trataba a los retrievals como “sondeos de baja calidad” sin tener en cuenta las características particulares de las observaciones de satélite.

En la década de los 90, luego de que los centros de asimilación comenzaran a utilizar técnicas avanzadas de asimilación de datos como 3D-Var, se dieron las condiciones necesarias para asimilar radianzas de satélites de manera directa. Sin embargo, la correcta asimilación de estas observaciones depende de 3 factores, que las observaciones no tengan sesgo, que sus errores tengan una distribución Gaussiana y que el problema no es afectado fuertemente por procesos no lineales (Eyre et al., 2022). Para asegurar estas condiciones fue necesario el desarrollo de técnicas de detección de nubes que permitan filtrar las regiones afectadas por nubosidad, principalmente para observaciones de sensores infrarrojos. Otro importante avance fue el desarrollo de modelos de transferencia radiativa que pudieran transformar el campo preliminar en radianzas comparables con las observaciones en tiempos razonables para ser usados de manera operativa. Finalmente, el desarrollo de métodos de corrección del bias de radianzas aplicados directamente en el proceso de asimilación fue determinante para la asimilación directa de este tipo de observaciones.

Justo al desarrollo de la asimilación de radianzas, también continuó el desarrollo de nuevos sensores, como la serie AMSU-A y AMSU-B y el sistema ATOVS (Advance TOVS) que cuenta con mayores canales y por lo tanto una mayor resolución vertical.

Posteriormente el desarrollo de los sensores multiespectrales como IASI y AIRS permitieron obtener información con mayor resolución vertical al contar con más de 3000 canales en la región infrarroja del espectro electromagnético.

Una parte importante del desarrollo la asimilación de datos en los últimos 20 años tiene que ver con el desarrollo de metodologías que tengan en cuenta la influencia de la superficie y la interacción entre las nubes y la energía electromagnética para los distintos canales infrarrojo y microondas. Inicialmente solo se asimilaron observaciones sobre agua y durante cielos despejados. Sin embargo mejoras en los modelos de transferencia radiativa respecto del tratamiento de los distintos tipos de superficie y la representación y tratamiendo de las nubes permiten en la actualidad incorporar observaciones que usualmente no podrían asimilarse.

Mientras que la asimilación directa de radianzas en modelos globales está establecida y estudiada [], la las aplicaciones en modelos regionales, sin embargo, sigue siendo un desafío debido a la escasa cobertura de las observaciones debido a la orbita de los satélites, la corrección del sesgo y el tope de la atmósfera bajos usados en modelos regionales. Bao et al. (2015) estudió el impacto de la asimilación de datos de radiancia de microondas e infrarrojo en el pronóstico de temperatura y humedad en el oeste de EE.UU. y encontró una reducción del sesgo de la temperatura en niveles bajos y medios como resultado de las observaciones de microondas, pero un efecto opuesto cuando se asimilaban radianzas en el infrarrojo. Más recientemente, Zhu et al. (2019) estudió el impacto de la asimilación frecuente de radiancias de satélites para un sistema regional y mostró una mejora para todas las variables, en particular para la humedad relativa en los niveles superiores. Wang & Randriamampianina (2021) estudiaron el impacto de la asimilación de radiancias en el Reanálisis Regional Europeo Copernicus de alta resolución e informaron de que las observaciones de radiancia de satélite tuvieron un impacto neutro en los análisis de la altura geopotencial en la tropósfera baja, mientras que el impacto fue ligeramente negativo para la tropósfera superior y estratosfera. También observaron resultados similares para pronósticos a 3 hs inicializados a partir del análisis, pero un impacto positivo en las previsiones de mediano plazo (12 y 24 hs). Teniendo en cuenta los variados resultados, es necesario continuar estudiando la utilidad de asimilar las observaciones de radiancia en un sistema de asimilación de datos de área limitada sobre tierra. El estudio de la asimilación de radianzas a nivel regional cobra aún mayor importancia en Sudamérica ya que no se conocen estudios realizados previamente.

0.4 Asimilación de datos en Sudamérica

La historia de la asimilación de datos en Sudamérica y en particular en Argentina es relativamente corta. A principios de la década del 90 Vera (1992) en su tesis doctoral desarrolló un Sistema de Asimilación de Datos Intermitente que utilizaba la interpolación optima en un modelo cuasigeostrófico en la región sur de Sudamérica. Algunos años después, en 1997, el Servicio Meteorológico Nacional se implementó un análisis utilizando el método de Cressman en un modelo de 10 niveles verticales (García Skabar, 1997).

Por otro lado el Centro de Pronóstico del Tiempo y Estudios Climáticos (CPTEC) de Brazil desarrollo un sistema de asimilación de datos global que utiliza el sistema Gridpoint Statistical Interpolation (GSI) en conjunto con su modelo global BAM y posteriormente aplicaciones regionales utilizando el modelo WRF en conjunto con el sistema de asimilación GSI. En particular, gustavo Goncalves de Goncalves et al. (2015) mostró experimentos realizados en el CPTEC usando el sistema de asimilación de datos regional para simulaciones de 12, 0 y 3 kilometros durante un mes. Ferreira, Herdies, Vendrasco, Beneti, & Biscaro (2017), Baucé Machado et al. (2017), Toshio Inouye et al. (2017), Vendrasco, Machado, Araujo, Ribaud, & Ferreira (2020), Ferreira et al. (2020) tambien mostraron resultados positivos al aplicar asimilación de datos en aplicaciones regionales sobre Brasil con resoluciones de entre 1 y 10 km.

En los últimos años, se documentaron importantes avances asociados a asimilación de datos en Argentina. Por ejemplo Marcos Saucedo realizó un estudio teórico de asimilación de datos utilizando LETKF acomplado al modelo WRF donde mostró con experimentos idealizados mejoras en la calidad del análisis aún cuando se asimilaban pocas observaciones. Posteriormente Maria Eugenia Dillon avanzó en su tesis de doctorado en el desarrollo de un sistema de asimilación de datos reales y concluyó que la implementación de un ensamble multifísica que considere los posibles errores del modelo y la inclusión de retrievals de temperatura y humedad en la asimilación tienen un impacto positivo en los análisis y pronósticos. Más recientemente, el Servicio Meteorológico Nacional (SMN) en conjunto con el Centro de Investigaciones del Mar y la Atmósfera desarrollaron y probaron el sistema de asimilación de actualización rápida LETKF-WRF de manera operativa durante la campaña de investigación RELAMPAGO (Nesbitt et al., 2021). El sistema incorporó observaciones convencionales, retrievals de satélites multiespectrales y viento derivado de observaciones satelitales y observaciones de radar de manera horaria y generó pronósticos a 36 hs cada 3 hs. Dillon et al. (2021) mostraron que el pronóstico inicializado a partir de los análisis muestra un rendimiento

general similar al de los pronósticos inicializados a partir del sistema GFS, e incluso un impacto positivo en algunos casos. Actualmente el SMN está probando un sistema de asimilación similar al implementado en Dillon et al. (2021) para utilizarlo en la generación de pronósticos de manera operativa.

0.5 Motivación y objetivos

En base a los importantes avances en la asimilación de datos en general y en las aplicaciones regionales en Argentina y Sudamérica, el objetivo principal de este trabajo es contribuir a la cuantificación y comparación del impacto de las estaciones meteorológicas de superficie de alta resolución, las observaciones de viento derivadas de satélite y las radiancias satelitales en cielo claro, en un sistema de DA de mesoescala, frecuentemente actualizado y basado en ensambles. En particular, este trabajo se centrará en el potencial impacto de la asimilación en el contexto de los eventos de sistemas convectivos de mesoescala (SCM) debido a la importancia que cobran este tipo de eventos en la región.

En particular, este trabajo busca investigar el impacto de distintas fuentes de datos en una región donde la red de observación convencional es bastante escasa y donde las contribuciones potenciales de sistemas de observación como redes de estaciones automáticas y observaciones de satélite son mayores. Para alcanzar este objetivo, se realizaron distintos experimentos de asimilación de datos aplicados a un estudio de caso de un SCM que se desarrolló sobre el sur de Sudamérica durante el 22 y 23 de noviembre de 2018 durante el período de observación intensa de la campaña de campo RELAMPAGO.

Chapter 1

R Markdown Basics

Here is a brief introduction into using *R Markdown*. *Markdown* is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. *R Markdown* provides the flexibility of *Markdown* with the implementation of **R** input and output. For more details on using *R Markdown* see <https://rmarkdown.rstudio.com>.

Be careful with your spacing in *Markdown* documents. While whitespace largely is ignored, it does at times give *Markdown* signals as to how to proceed. As a habit, try to keep everything left aligned whenever possible, especially as you type a new paragraph. In other words, there is no need to indent basic text in the Rmd document (in fact, it might cause your text to do funny things if you do).

1.1 Lists

It's easy to create a list. It can be unordered like

- Item 1
- Item 2

or it can be ordered like

1. Item 1
2. Item 2

Notice that I intentionally mislabeled Item 2 as number 4. *Markdown* automatically figures this out! You can put any numbers in the list and it will create the list. Check it out below.

To create a sublist, just indent the values a bit (at least four spaces or a tab). (Here's one case where indentation is key!)

1. Item 1
2. Item 2
3. Item 3
 - Item 3a
 - Item 3b

1.2 Line breaks

Make sure to add white space between lines if you'd like to start a new paragraph. Look at what happens below in the outputted document if you don't:

Here is the first sentence. Here is another sentence. Here is the last sentence to end the paragraph. This should be a new paragraph.

Now for the correct way:

Here is the first sentence. Here is another sentence. Here is the last sentence to end the paragraph.

This should be a new paragraph.

1.3 R chunks

When you click the **Knit** button above a document will be generated that includes both content as well as the output of any embedded **R** code chunks within the document. You can embed an **R** code chunk like this (`cars` is a built-in **R** dataset):

	speed		dist
Min.	: 4.0	Min.	: 2.00
1st Qu.	:12.0	1st Qu.	: 26.00
Median	:15.0	Median	: 36.00
Mean	:15.4	Mean	: 42.98
3rd Qu.	:19.0	3rd Qu.	: 56.00
Max.	:25.0	Max.	:120.00

1.4 Inline code

If you'd like to put the results of your analysis directly into your discussion, add inline code like this:

The `cos` of 2π is 1.

Another example would be the direct calculation of the standard deviation:

The standard deviation of `speed` in `cars` is 5.2876444.

One last neat feature is the use of the `ifelse` conditional statement which can be used to output text depending on the result of an **R** calculation:

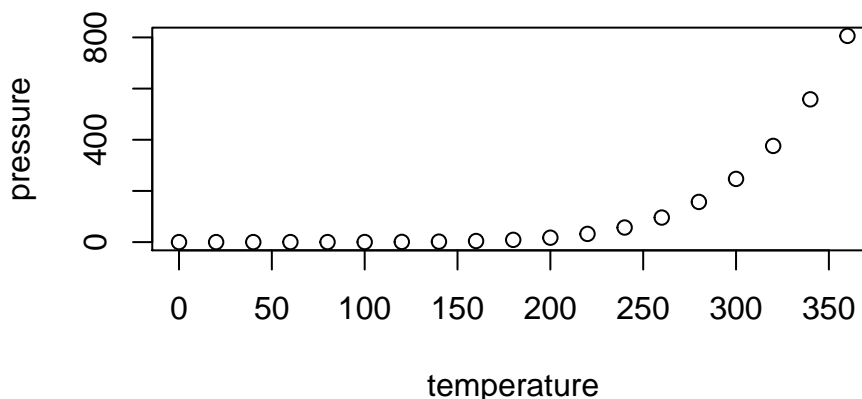
The standard deviation is less than 6.

Note the use of `>` here, which signifies a quotation environment that will be indented.

As you see with `2π` above, mathematics can be added by surrounding the mathematical text with dollar signs. More examples of this are in Mathematics and Science if you uncomment the code in Math.

1.5 Including plots

You can also embed plots. For example, here is a way to use the base **R** graphics package to produce a plot using the built-in `pressure` dataset:



Note that the `echo=FALSE` parameter was added to the code chunk to prevent printing of the **R** code that generated the plot. There are plenty of other ways to add chunk options (like `fig.height` and `fig.width` in the chunk above). More information is available at <https://yihui.org/knitr/options/>.

Another useful chunk option is the setting of `cache=TRUE` as you see here. If document rendering becomes time consuming due to long computations or plots that are expensive to generate you can use knitr caching to improve performance. Later in this file, you'll see a way to reference plots created in **R** or external figures.

1.6 Loading and exploring data

Included in this template is a file called `flights.csv`. This file includes a subset of the larger dataset of information about all flights that departed from Seattle and Portland in 2014. More information about this dataset and its **R** package is available at <https://github.com/ismayc/pnwflights14>. This subset includes only Portland flights and only rows that were complete with no missing values. Merges were also done with the `airports` and `airlines` data sets in the `pnwflights14` package to get more descriptive airport and airline names.

We can load in this data set using the following commands:

The data is now stored in the data frame called `flights` in **R**. To get a better feel for the variables included in this dataset we can use a variety of functions. Here we can see the dimensions (rows by columns) and also the names of the columns.

```
[1] 12649    16
```

```
[1] "month"      "day"         "dep_time"    "dep_delay"
[5] "arr_time"   "arr_delay"   "carrier"     "tailnum"
[9] "flight"     "dest"        "air_time"    "distance"
[13] "hour"       "minute"      "carrier_name" "dest_name"
```

Another good idea is to take a look at the dataset in table form. With this dataset having more than 20,000 rows, we won't explicitly show the results of the command here. I recommend you enter the command into the Console *after* you have run the **R** chunks above to load the data into **R**.

While not required, it is highly recommended you use the `dplyr` package to manipulate and summarize your data set as needed. It uses a syntax that is easy to understand using chaining operations. Below I've created a few examples of using `dplyr` to get information about the Portland flights in 2014. You will also see the use of the `ggplot2` package, which produces beautiful, high-quality academic visuals.

We begin by checking to ensure that needed packages are installed and then we load them into our current working environment:

The example we show here does the following:

- Selects only the `carrier_name` and `arr_delay` from the `flights` dataset and then assigns this subset to a new variable called `flights2`.
- Using `flights2`, we determine the largest arrival delay for each of the carriers.

A useful function in the `knitr` package for making nice tables in *R Markdown* is called `kable`. It is much easier to use than manually entering values into a table by copying and pasting values into Excel or LaTeX. This again goes to show how nice reproducible documents can be! (Note the use of `results="asis"`, which will produce the table instead of the code to create the table.) The `caption.short` argument is used to include a shorter title to appear in the List of Tables.

Table 1.1: Maximum Delays by Airline

Airline	Max Arrival Delay
Alaska Airlines Inc.	338
American Airlines Inc.	1539
Delta Air Lines Inc.	371
Frontier Airlines Inc.	166
Hawaiian Airlines Inc.	116
JetBlue Airways	256
SkyWest Airlines Inc.	321
Southwest Airlines Co.	315
United Air Lines Inc.	319
US Airways Inc.	347
Virgin America	366

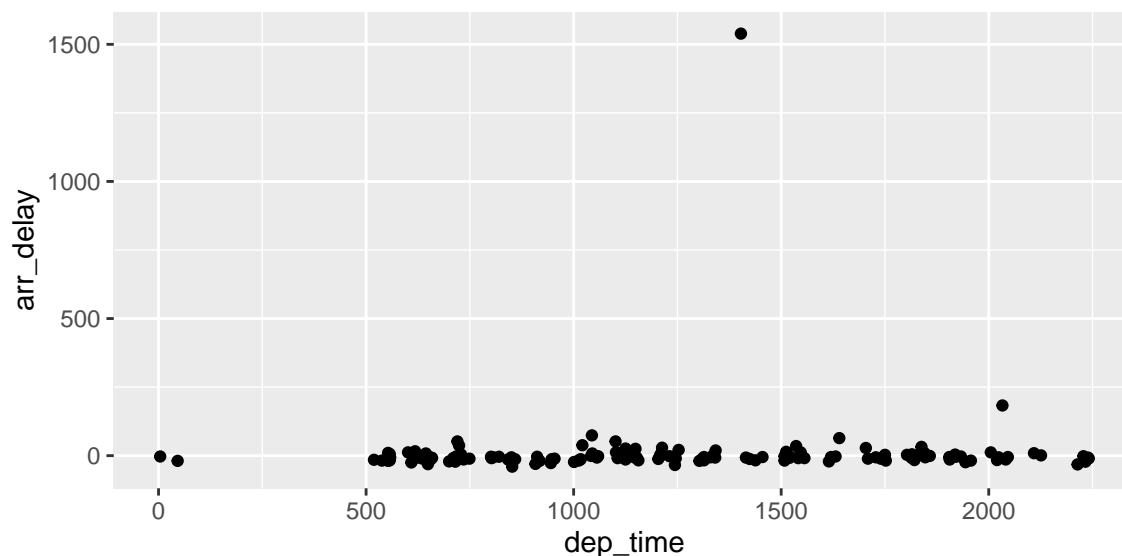
The last two options make the table a little easier-to-read.

We can further look into the properties of the largest value here for American Airlines Inc. To do so, we can isolate the row corresponding to the arrival delay of 1539 minutes for American in our original `flights` dataset.

```
dep_time dep_delay arr_time tailnum flight dest air_time distance
1      1403      1553     1934  N595AA  1568  DFW       182      1616
```

We see that the flight occurred on March 3rd and departed a little after 2 PM on its way to Dallas/Fort Worth. Lastly, we show how we can visualize the arrival delay of

all departing flights from Portland on March 3rd against time of departure.



1.7 Additional resources

- *Markdown Cheatsheet* - <https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>
- *R Markdown*
 - Reference Guide - <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>
 - Cheatsheet - <https://github.com/rstudio/cheatsheets/raw/master/rmarkdown-2.0.pdf>
- *RStudio IDE*
 - Cheatsheet - <https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf>
 - Official website - <https://rstudio.com/products/rstudio/>
- Introduction to dplyr - <https://cran.rstudio.com/web/packages/dplyr/vignettes/dplyr.html>
- ggplot2
 - Documentation - <https://ggplot2.tidyverse.org/>

-
- Cheatsheet - <https://github.com/rstudio/cheatsheets/raw/master/data-visualization-2.1.pdf>

Chapter 2

Mathematics and Science

2.1 Math

T_EX is the best way to typeset mathematics. Donald Knuth designed T_EX when he got frustrated at how long it was taking the typesetters to finish his book, which contained a lot of mathematics. One nice feature of *R Markdown* is its ability to read LaTeX code directly.

If you are doing a thesis that will involve lots of math, you will want to read the following section which has been commented out. If you're not going to use math, skip over or delete this next commented section.

2.2 Chemistry 101: Symbols

Chemical formulas will look best if they are not italicized. Get around math mode's automatic italicizing in LaTeX by using the argument `$\mathrm{formula here}$` , with your formula inside the curly brackets. (Notice the use of the backticks here which enclose text that acts as code.)

So, Fe₂²⁺Cr₂O₄ is written `$\mathrm{Fe_2^{2+}Cr_2O_4}$` .

Exponent or Superscript: O⁻

Subscript: CH₄

To stack numbers or letters as in Fe₂²⁺, the subscript is defined first, and then the superscript is defined.

Bullet: CuCl • 7H₂O

Delta: Δ

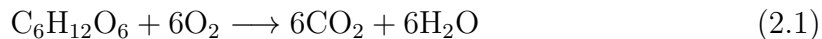
Reaction Arrows: \longrightarrow or $\xrightarrow{\text{solution}}$

Resonance Arrows: \leftrightarrow

Reversible Reaction Arrows: \rightleftharpoons

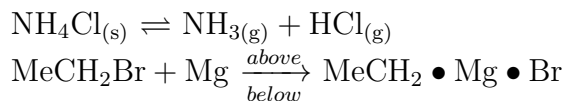
2.2.1 Typesetting reactions

You may wish to put your reaction in an equation environment, which means that LaTeX will place the reaction where it fits and will number the equations for you.



We can reference this combustion of glucose reaction via Equation (2.1).

2.2.2 Other examples of reactions



2.3 Physics

Many of the symbols you will need can be found on the math page <https://web.reed.edu/cis/help/latex/math.html> and the Comprehensive LaTeX Symbol Guide (<https://mirror.utexas.edu/ctan/info/symbols/comprehensive/symbols-letter.pdf>).

2.4 Biology

You will probably find the resources at <https://www.lecb.ncifcrf.gov/~toms/latex.html> helpful, particularly the links to bst files for various journals. You may also be interested in TeXShade for nucleotide typesetting (<https://homepages.uni-tuebingen.de/beitz/txe.html>). Be sure to read the proceeding chapter on graphics and tables.

Chapter 3

Graphics, References, and Labels

3.1 Figures

If your thesis has a lot of figures, *R Markdown* might behave better for you than that other word processor. One perk is that it will automatically number the figures accordingly in each chapter. You'll also be able to create a label for each figure, add a caption, and then reference the figure in a way similar to what we saw with tables earlier. If you label your figures, you can move the figures around and *R Markdown* will automatically adjust the numbering for you. No need for you to remember! So that you don't have to get too far into LaTeX to do this, a couple **R** functions have been created for you to assist. You'll see their use below.

In the **R** chunk below, we will load in a picture stored as `reed.jpg` in our main directory. We then give it the caption of "Reed logo", the label of "reedlogo", and specify that this is a figure. Make note of the different **R** chunk options that are given in the R Markdown file (not shown in the knitted document).



Figure 3.1: Reed logo

Here is a reference to the Reed logo: Figure 3.1. Note the use of the `fig:` code here. By naming the **R** chunk that contains the figure, we can then reference that figure later as done in the first sentence here. We can also specify the caption for the

figure via the R chunk option `fig.cap`.

Below we will investigate how to save the output of an **R** plot and label it in a way similar to that done above. Recall the `flights` dataset from Chapter 1. (Note that we've shown a different way to reference a section or chapter here.) We will next explore a bar graph with the mean flight departure delays by airline from Portland for 2014.

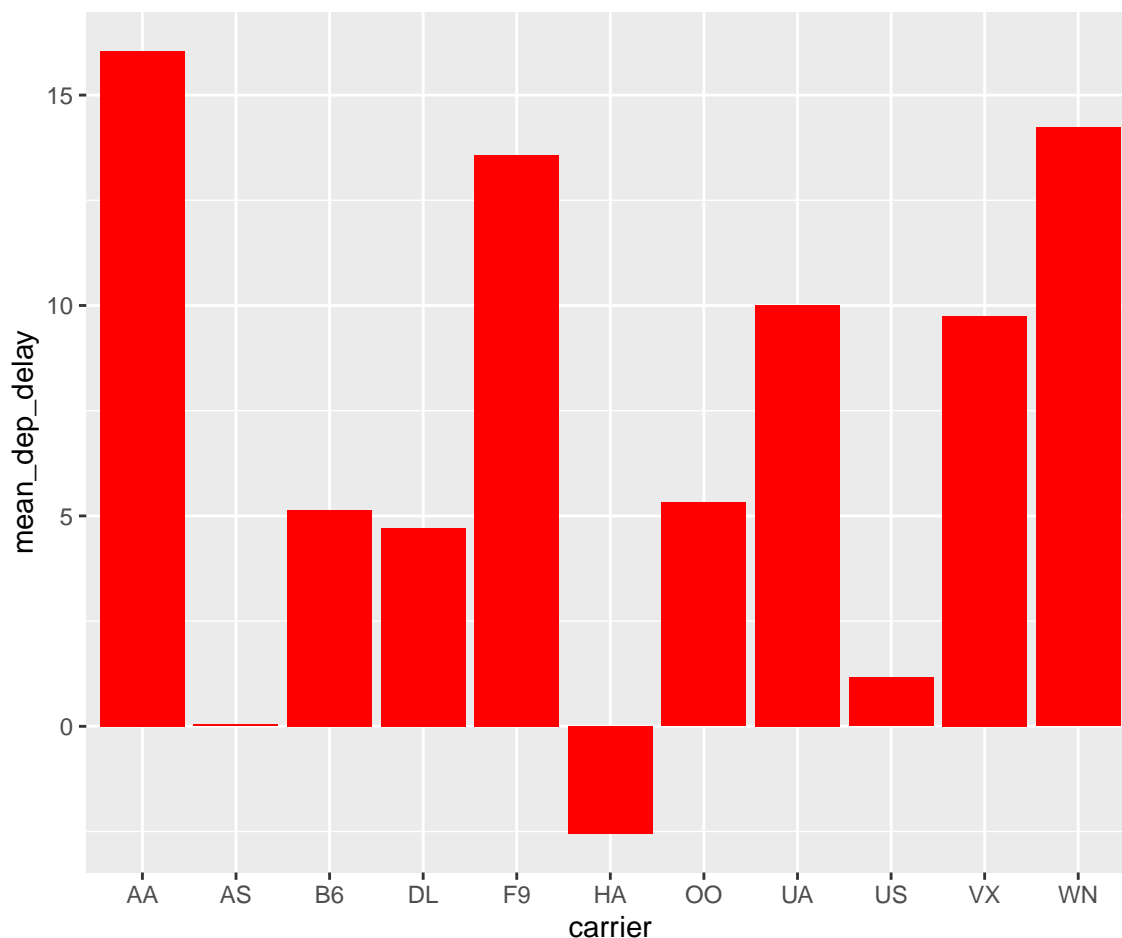


Figure 3.2: Mean Delays by Airline

Here is a reference to this image: Figure 3.2.

A table linking these carrier codes to airline names is available at <https://github.com/ismayc/pnwflights14/blob/master/data/airlines.csv>.

Next, we will explore the use of the `out.extra` chunk option, which can be used to shrink or expand an image loaded from a file by specifying "`scale=` ". Here we use the mathematical graph stored in the “subdivision.pdf” file.

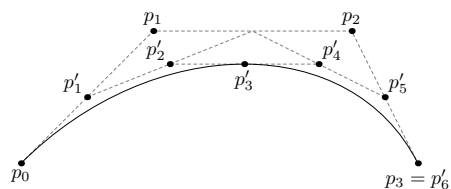


Figure 3.3: Subdiv. graph

Here is a reference to this image: Figure 3.3. Note that `echo=FALSE` is specified so that the **R** code is hidden in the document.

More Figure Stuff

Lastly, we will explore how to rotate and enlarge figures using the `out.extra` chunk option. (Currently this only works in the PDF version of the book.)

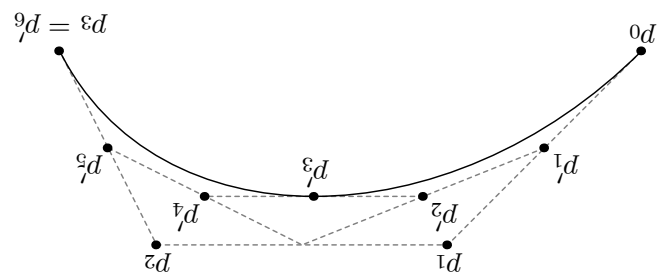


Figure 3.4: A Larger Figure, Flipped Upside Down

As another example, here is a reference: Figure 3.4.

3.2 Footnotes and Endnotes

You might want to footnote something.¹ The footnote will be in a smaller font and placed appropriately. Endnotes work in much the same way. More information can be found about both on the CUS site or feel free to reach out to `data@reed.edu`.

3.3 Bibliographies

Of course you will need to cite things, and you will probably accumulate an armful of sources. There are a variety of tools available for creating a bibliography database

¹footnote text

(stored with the .bib extension). In addition to BibTeX suggested below, you may want to consider using the free and easy-to-use tool called Zotero. The Reed librarians have created Zotero documentation at <https://libguides.reed.edu/citation/zotero>. In addition, a tutorial is available from Middlebury College at <https://sites.middlebury.edu/zoteromiddlebury/>.

R Markdown uses *pandoc* (<https://pandoc.org/>) to build its bibliographies. One nice caveat of this is that you won't have to do a second compile to load in references as standard LaTeX requires. To cite references in your thesis (after creating your bibliography database), place the reference name inside square brackets and precede it by the "at" symbol. For example, here's a reference to a book about worrying: (???). This Molina1994 entry appears in a file called `thesis.bib` in the `bib` folder. This bibliography database file was created by a program called BibTeX. You can call this file something else if you like (look at the YAML header in the main .Rmd file) and, by default, is to placed in the `bib` folder.

For more information about BibTeX and bibliographies, see our CUS site (<https://web.reed.edu/cis/help/latex/index.html>)². There are three pages on this topic: *bibtex* (which talks about using BibTeX, at <https://web.reed.edu/cis/help/latex/bibtex.html>), *bibtexstyles* (about how to find and use the bibliography style that best suits your needs, at <https://web.reed.edu/cis/help/latex/bibtexstyles.html>) and *bibman* (which covers how to make and maintain a bibliography by hand, without BibTeX, at <https://web.reed.edu/cis/help/latex/bibman.html>). The last page will not be useful unless you have only a few sources.

If you look at the YAML header at the top of the main .Rmd file you can see that we can specify the style of the bibliography by referencing the appropriate csl file. You can download a variety of different style files at <https://www.zotero.org/styles>. Make sure to download the file into the csl folder.

Tips for Bibliographies

- Like with thesis formatting, the sooner you start compiling your bibliography for something as large as thesis, the better. Typing in source after source is mind-numbing enough; do you really want to do it for hours on end in late April? Think of it as procrastination.
- The cite key (a citation's label) needs to be unique from the other entries.

²(???)

- When you have more than one author or editor, you need to separate each author’s name by the word “and” e.g. `Author = {Noble, Sam and Youngberg, Jessica},.`
- Bibliographies made using BibTeX (whether manually or using a manager) accept LaTeX markup, so you can italicize and add symbols as necessary.
- To force capitalization in an article title or where all lowercase is generally used, bracket the capital letter in curly braces.
- You can add a Reed Thesis citation³ option. The best way to do this is to use the `phdthesis` type of citation, and use the optional “type” field to enter “Reed thesis” or “Undergraduate thesis.”

3.4 Anything else?

If you’d like to see examples of other things in this template, please contact the Data @ Reed team (email data@reed.edu) with your suggestions. We love to see people using *R Markdown* for their theses, and are happy to help.

³(???)

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readability and/or setup.

In the main Rmd file

```
# This chunk ensures that the thesishdown package is  
# installed and loaded. This thesishdown package includes  
# the template files for the thesis.  
if (!require(remotes)) {  
  if (params$`Install needed packages for {thesishdown}`) {  
    install.packages("remotes", repos = "https://cran.rstudio.com")  
  } else {  
    stop(  
      paste('You need to run install.packages("remotes")',  
            "first in the Console.")  
    )  
  }  
}  
  
if (!require(thesishdown)) {  
  if (params$`Install needed packages for {thesishdown}`) {  
    remotes::install_github("ismayc/thesishdown")  
  } else {  
    stop(  
      paste(  
        "You need to run",
```

```

      'remotes::install_github("ismayc/thesisdown")',
      "first in the Console."
    )
  )
}
}
library(thesisdown)
# Set how wide the R output will go
options(width = 70)

```

In Chapter 3:

```

# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis and also two functions
# used for labeling and referencing
if (!require(remotes)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("remotes", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("remotes")',
        "first in the Console."
      )
    )
  }
}

if (!require(dplyr)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("dplyr", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("dplyr")',
        "first in the Console."
      )
    )
  }
}

```

```

    )
  )
}
}
if (!require(ggplot2)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("ggplot2", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("ggplot2")',
        "first in the Console."
      )
    )
  }
}
if (!require(bookdown)) {
  if (params$`Install needed packages for {thesisdown}`) {
    install.packages("bookdown", repos = "https://cran.rstudio.com")
  } else {
    stop(
      paste(
        'You need to run install.packages("bookdown")',
        "first in the Console."
      )
    )
  }
}
if (!require(thesisdown)) {
  if (params$`Install needed packages for {thesisdown}`) {
    remotes::install_github("ismayc/thesisdown")
  } else {
    stop(
      paste(
        "You need to run",
        'remotes::install_github("ismayc/thesisdown")',

```

```
      "first in the Console."
    )
  )
}
}
library(thesisdown)
library(dplyr)
library(ggplot2)
library(knitr)
flights <- read.csv("data/flights.csv", stringsAsFactors = FALSE)
```

Appendix B

The Second Appendix, for Fun

References

- Aksoy, A., Dowell, D. C., & Snyder, C. (2010). A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part II: Short-Range Ensemble Forecasts. *Monthly Weather Review*, 138(4), 1273–1292. <http://doi.org/10.1175/2009MWR3086.1>
- Andersson, E., Hollingsworth, A., Kelly, G., Lönnberg, P., Pailleux, J., & Zhang, Z. (1991). Global Observing System Experiments on Operational Statistical Retrievals of Satellite Sounding Data. *Monthly Weather Review*, 119(8), 1851–1865. [http://doi.org/10.1175/1520-0493\(1991\)119%3C1851:GOSE00%3E2.0.CO;2](http://doi.org/10.1175/1520-0493(1991)119%3C1851:GOSE00%3E2.0.CO;2)
- Bae, J.-H., & Min, K.-H. (2022). Forecast Characteristics of Radar Data Assimilation Based on the Scales of Precipitation Systems. *Remote Sensing*, 14(3, 3), 605. <http://doi.org/10.3390/rs14030605>
- Bao, Y., Xu, J., Powell Jr., A. M., Shao, M., Min, J., & Pan, Y. (2015). Impacts of AMSU-A, MHS and IASI data assimilation on temperature and humidity forecasts with GSI-WRF over the western United States. *Atmospheric Measurement Techniques*, 8(10), 4231–4242. <http://doi.org/10.5194/amt-8-4231-2015>
- Bauce Machado, V., gustavo de goncalves, luis, Vendrasco, E., Sinhori, N., Herdies, D., Sapucci, L., ... Biscaro, T. (2017). Investigating the impacts of convective scale hazardous weather events in Santa Catarina State through the CPTEC/INPE local data assimilation system. In. Presented at the Seventh International WMO Symposium on Data Assimilation. Retrieved from https://www.researchgate.net/publication/320306290_Investigating_the_impacts_of_convective_scale_hazardous_weather_events_in_Santa_Catarina_State_through_the_CPTecINPE_local_data_assimilation_system
- Chang, W., Jacques, D., Fillion, L., & Baek, S.-J. (2017). Assimilation of Hourly Surface Observations with the Canadian High-Resolution Ensemble Kalman Filter. *Atmosphere-Ocean*, 55(4-5), 247–263. <http://doi.org/10.1080/07055900.2017.1384361>
- Chen, X., Zhao, K., Sun, J., Zhou, B., & Lee, W.-C. (2016). Assimilating surface

- observations in a four-dimensional variational Doppler radar data assimilation system to improve the analysis and forecast of a squall line case. *Advances in Atmospheric Sciences*, 33(10), 1106–1119. <http://doi.org/10.1007/s00376-016-5290-0>
- Cherubini, T., Businger, S., Velden, C., & Ogasawara, R. (2006). The Impact of Satellite-Derived Atmospheric Motion Vectors on Mesoscale Forecasts over Hawaii. *Monthly Weather Review*, 134(7), 2009–2020. <http://doi.org/10.1175/MWR3163.1>
- Clark, A. J., Gallus, W. A., Xue, M., & Kong, F. (2009). A Comparison of Precipitation Forecast Skill between Small Convection-Allowing and Large Convection-Parameterizing Ensembles. *Weather and Forecasting*, 24(4), 1121–1140. <http://doi.org/10.1175/2009WAF2222222.1>
- Dillon, M. E., Maldonado, P., Corrales, P., Skabar, Y. G., Ruiz, J., Sacco, M., ... Miyoshi, T. (2021). A rapid refresh ensemble based data assimilation and forecast system for the RELAMPAGO field campaign. *Atmospheric Research*, 105858. <http://doi.org/10.1016/j.atmosres.2021.105858>
- Eyre, J. R., Bell, W., Cotton, J., English, S. J., Forsythe, M., Healy, S. B., & Pavein, E. G. (2022). Assimilation of satellite data in numerical weather prediction. Part II: Recent years. *Quarterly Journal of the Royal Meteorological Society*, 148(743), 521–556. <http://doi.org/10.1002/qj.4228>
- Eyre, J. R., English, S. J., & Forsythe, M. (2020). Assimilation of satellite data in numerical weather prediction. Part I: The early years. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 49–68. <http://doi.org/10.1002/qj.3654>
- Ferreira, R. C., Alves Júnior, M. P., Vendasco, Éder P., Aravéquia, J. A., Nolasco Junior, L. R., Biscaro, T. S., ... Biscaro, T. S. (2020). The Impact of Microphysics Parameterization on Precipitation Forecast Using Radar Data Assimilation. *Revista Brasileira de Meteorologia*, 35(1), 123–134. <http://doi.org/10.1590/0102-7786351005>
- Ferreira, R. C., Herdies, D. L., Vendasco, É. P., Beneti, C. A. A., & Biscaro, T. S. (2017). Impacto da Assimilação de Dados de Radar em Sistemas Convectivos de Mesoescala: Um Estudo de Caso. *Revista Brasileira de Meteorologia*, 32(3), 447–458. <http://doi.org/10.1590/0102-77863230011>
- Gao, F., Huang, X.-Y., Jacobs, N. A., & Wang, H. (2015). Assimilation of wind speed and direction observations: Results from real observation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1), 27132. <http://doi.org/10.3402/tellusa.v67.27132>
- García Skabar, Y. (1997). *Análisis objetivo regional para inicializar un modelo de diez niveles en forma operativa. Tesis de licenciatura en ciencias de la atmósfera.* Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales.

- Gasperoni, N. A., Wang, X., Brewster, K. A., & Carr, F. H. (2018). Assessing Impacts of the High-Frequency Assimilation of Surface Observations for the Forecast of Convection Initiation on 3 April 2014 within the Dallas–Fort Worth Test Bed. *Monthly Weather Review*, 146(11), 3845–3872. <http://doi.org/10.1175/MWR-D-18-0177.1>
- gustavo Goncalves de Goncalves, L., Sapucci, L., Vendrasco, E., de Mattos, J. G., Ferreira, C., Khamis, E., & Cruz, N. (2015). A rapid update data assimilation cycle over South America using 3DVar and EnKF. In *The 20th International TOVS Study Conference (ITSC-20)*. Lake Geneva, Wisconsin, USA: The 20th International TOVS Study Conference (ITSC-20).
- Ha, S.-Y., & Snyder, C. (2014). Influence of Surface Observations in Mesoscale Data Assimilation Using an Ensemble Kalman Filter. *Monthly Weather Review*, 142(4), 1489–1508. <http://doi.org/10.1175/MWR-D-13-00108.1>
- Kalnay, E. (2002, November 6). Atmospheric Modeling, Data Assimilation and Predictability. <http://doi.org/10.1017/CB09780511802270>
- Kelly, G. a. M., Mills, G. A., & Smith, W. L. (1978). Impact of Nimbus-6 Temperature Soundings on Australian Region Forecasts. *Bulletin of the American Meteorological Society*, 59(4), 393–406. <http://doi.org/10.1175/1520-0477-59.4.393>
- Maejima, Y., Miyoshi, T., Kunii, M., Seko, H., & Sato, K. (2019). Impact of Dense and Frequent Surface Observations on 1-Minute-Update Severe Rainstorm Prediction: A Simulation Study. *Journal of the Meteorological Society of Japan. Ser. II*, 97(1), 253–273. <http://doi.org/10.2151/jmsj.2019-014>
- Nesbitt, S. W., Salio, P. V., Ávila, E., Bitzer, P., Carey, L., Chandrasekar, V., ... Grover, M. A. (2021). A storm safari in Subtropical South America: Proyecto RELAMPAGO. *Bulletin of the American Meteorological Society*, -1(aop), 1–64. <http://doi.org/10.1175/BAMS-D-20-0029.1>
- Ohring, G. (1979). Impact of Satellite Temperature Sounding Data on Weather Forecasts. *Bulletin of the American Meteorological Society*, 60(10), 1142–1147. [http://doi.org/10.1175/1520-0477\(1979\)060%3C1142:I0STSD%3E2.0.CO;2](http://doi.org/10.1175/1520-0477(1979)060%3C1142:I0STSD%3E2.0.CO;2)
- Sawada, M., Ma, Z., Mehra, A., Tallapragada, V., Oyama, R., & Shimoji, K. (2019). Impacts of Assimilating High-Resolution Atmospheric Motion Vectors Derived from Himawari-8 on Tropical Cyclone Forecast in HWRF. *Monthly Weather Review*, 147(10), 3721–3740. <http://doi.org/10.1175/MWR-D-18-0261.1>
- Sobash, R. A., & Stensrud, D. J. (2015). Assimilating Surface Mesonet Observations with the EnKF to Improve Ensemble Forecasts of Convection Initiation on 29 May 2012. *Monthly Weather Review*, 143(9), 3700–3725. <http://doi.org/10.1175/MWR-D-14-00126.1>

- Stensrud, D. J., Wicker, L. J., Xue, M., Dawson, D. T., Yussouf, N., Wheatley, D. M., ... Brewster, K. A. (2013). Progress and challenges with Warn-on-Forecast. *Atmospheric Research*, 123, 2–16. <http://doi.org/10.1016/j.atmosres.2012.04.004>
- Toshio Inouye, R., Calvetti, L., Gonçalves, J., Maske, B., Neundorff, R., Beneti, C., ... gustavo de goncalves, luis. (2017). Impact of radar data assimilation on a severe storm study in brazil. In. Presented at the 97th American Meteorological Meeting Annual Meeting, Seattle, WA. Retrieved from https://www.researchgate.net/publication/313040194_Impact_of_Radar_Data_Assimilation_on_a_Severe_Storm_Study_in_Brazil/link/5bf7d30b299bf1a0202cbcc0/download
- Vendrasco, E. P., Machado, L. A. T., Araujo, C. S., Ribaud, J.-F., & Ferreira, R. C. (2020). Potential use of the GLM for nowcasting and data assimilation. *Atmospheric Research*, 242, 105019. <http://doi.org/10.1016/j.atmosres.2020.105019>
- Vera, C. S. (1992). *Un sistema de asimilación de datos para la región extratropical de Sudamérica* (Tesis Doctoral). Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Retrieved from https://bibliotecadigital.exactas.uba.ar/collection/tesis/document/tesis_n2498_Vera
- Wang, Z. Q., & Randriamampianina, R. (2021). The Impact of Assimilating Satellite Radiance Observations in the Copernicus European Regional Reanalysis (CERRA). *Remote Sensing*, 13(3, 3), 426. <http://doi.org/10.3390/rs13030426>
- Wheatley, D. M., & Stensrud, D. J. (2010). The Impact of Assimilating Surface Pressure Observations on Severe Weather Events in a WRF Mesoscale Ensemble System. *Monthly Weather Review*, 138(5), 1673–1694. <http://doi.org/10.1175/2009MWR3042.1>
- Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., & Toth, Z. (2008). Ensemble Data Assimilation with the NCEP Global Forecast System. *Monthly Weather Review*, 136(2), 463–482. <http://doi.org/10.1175/2007MWR2018.1>
- Wu, T.-C., Liu, H., Majumdar, S. J., Velden, C. S., & Anderson, J. L. (2014). Influence of Assimilating Satellite-Derived Atmospheric Motion Vector Observations on Numerical Analyses and Forecasts of Tropical Cyclone Track and Intensity. *Monthly Weather Review*, 142(1), 49–71. <http://doi.org/10.1175/MWR-D-13-00023.1>
- Zhu, K., Xue, M., Pan, Y., Hu, M., Benjamin, S. G., Weygandt, S. S., & Lin, H. (2019). The Impact of Satellite Radiance Data Assimilation within a Frequently Updated Regional Forecast System Using a GSI-based Ensemble Kalman Filter. *Advances in Atmospheric Sciences*, 36(12), 1308–1326. <http://doi.org/10.1007/s00376-019-9011-3>