



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ  
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра метеорологических прогнозов

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**  
(магистерская диссертация)

На тему Уточнение результатов гидродинамического прогноза с  
использованием концепции Model Output Statistics (MOS)

Исполнитель Воробьева Ольга Владимировна  
(фамилия, имя, отчество)

Руководитель кандидат физико-математических наук  
(ученая степень, ученое звание)

Анискина Ольга Георгиевна  
(фамилия, имя, отчество)

«К защите допускаю»  
Заведующий кафедрой

(подпись)

доктор физико-математических наук, доцент  
(ученая степень, ученое звание)

Дробжева Яна Викторовна  
(фамилия, имя, отчество)

5 мая 2018 г.

Санкт-Петербург  
2018

## СОДЕРЖАНИЕ

Введение	4
1 Статические методы коррекции гидродинамических прогнозов погоды	7
1.1 Регрессионные техники коррекции результатов гидродинамического прогноза	7
1.1.1 Одномерная линейная регрессия	8
1.1.2 Множественная линейная регрессия	9
1.2 Построение регрессионных моделей	10
1.2.1 Определение связи между предиктантом и предиктором	10
1.2.2 Выбор предикторов	11
1.3 Техника Model Output Statistics (MOS)	15
2 Гидродинамическая мезомасштабная модель WRF	17
2.1 Модель WRF и ее применение	18
2.2 Характеристики используемой WRF	21
3 Численные эксперименты и анализ результатов	25
3.1 Данные, используемые в исследовании, и их предварительная обработка	25
3.1.1 Предварительная обработка результатов моделирования мезомасштабной гидродинамической модели WRF	26
3.1.2 Предварительная обработка данных синоптических сообщений	29
3.2 Выбор модели для определения лучших предикторов	31
3.2.1 Описание, использованных программных моделей, и алгоритмов вычислений	32
3.2.2 Оценка потенциальных предикторов корреляционным анализом	38
3.2.3 Выбор статистической модели	40
3.3 Расчет ошибок регрессионных моделей	42
4 Анализ результатов	48

4.1 Анализ полученных результатов	48
4.2 Применение	52
Заключение	53
Список использованных источников	55
Приложение А	57
Приложение Б	58
Приложение В	59

## ВВЕДЕНИЕ

Последнее время для прогнозирования погоды широко используются численные гидродинамические модели. Самая распространённая из них модель WRF (Weather Research and Forecasting Model). Но, как известно, моделирование не способно предсказывать с необходимой точностью. Одним из вариантов улучшения результатов прогноза является пост-процессинг, который содержит в себе несколько методов. В данной работе была рассмотрена техника MOS (Model Output Statistics). Эта техника подразумевает под собой использование множественной линейной регрессии, где в левой части уравнения находится прогнозируемая величина (предиктант), а в правой результаты моделирования (предикторы) метеорологических величин.

Актуальность этой работы заключается в использовании техники пост-процессинга MOS в качестве варианта улучшения результатов прогнозирования численной гидродинамической модели WRF.

Цель работы: уточнить результаты гидродинамического прогноза с использованием концепции Model Output Statistics (MOS).

Перед началом работы, были поставлены следующие задачи:

- 1) Научиться обрабатывать большое количество данных синоптических сообщений с метеорологических станций с помощью высокоуровневого языка программирования Python.
- 2) Научиться обрабатывать большое количество данных результатов прогнозирования модели WRF с помощью высокоуровневого языка программирования Python.
- 3) Изучить научную литературу по технике пост-процессинга MOS.

4) Разработать программный код с применением специальных модулей для расчета всех необходимых параметров техники MOS с помощью высокоуровневого языка программирования Python.

5) Произвести исследование возможностей улучшения результатов прогноза WRF для температуры воздуха на уровне 2 метра для каждой метеорологической станции, используя технику MOS.

6) Произвести верификацию результатов применения техники MOS с результатами прогнозирования численной модели WRF.

7) Разработать технологию внедрения техники MOS в оперативную практику.

Магистерская диссертация состоит из введения, четырёх глав и заключения.

В первой главе приводится обоснование использования статистической обработки результатов моделирования и приводятся основные положения техники MOS.

Во второй главе описываются характеристики используемой численной модели прогнозирования WRF.

В третьей главе описывается алгоритм уточнения результатов гидродинамического прогноза температуры воздуха на уровне 2 метра.

В четвертой главе анализируются полученные результаты и подводятся итоги работы.

В качестве литературных источников для теоретической составляющей техники MOS и численной гидродинамической модели WRF, были использованы преимущественно научные труды последнего десятилетия следующих авторов: Armstrong, J. Scott, Jordan G. Powers, Joseph B. Klemp и др. Не без внимания остались научные работы прошлого века, в которых зарождались исследования данной техники, например, автора Harry R. Glahn. А так же исследования в области прогностического моделирования наших

соотечественников. Всего список использованных источников содержит 11 наименований.

## 1 Статические методы коррекции гидродинамических прогнозов погоды

Все современные прогнозы не лишены ошибок. Это напрямую относится и к гидродинамическим прогнозам погоды. Ошибки связаны с неточностями в начальных данных, с упрощением уравнений гидродинамики атмосферы, с неточными (приближенными, вычислительными, численными) методами решения уравнений, неточными методами описания физических процессов (параметризациями), которые не могут быть описаны на сетке точек или при спектральном представлении.

Мировое научное сообщество старается уменьшить все перечисленные ошибки. Наряду с этим сегодня разрабатываются методы коррекции результатов гидродинамического прогноза. Эти методы могут быть разделены на два типа: регрессионные и техника MOS [1].

### 1.1 Регрессионные техники коррекции результатов гидродинамического прогноза

В статистическом моделировании регрессионный анализ представляет собой набор статистических процессов для оценки взаимосвязи между переменными. Он включает в себя множество методов моделирования и анализа нескольких переменных, когда основное внимание уделяется взаимосвязи между зависимой переменной и одной или несколькими независимыми переменными (или предикторами). Более конкретно, регрессионный анализ помогает понять, как изменяется типичное значение зависимой переменной при изменении любой из независимых переменных, в то время как другие независимые переменные удерживаются фиксированными.

Регрессионный анализ широко используется для прогнозирования, где его использование имеет много общих черт машинным обучением. Регрессионный анализ также используется для понимания того, какие из независимых переменных связаны с зависимой переменной, и для изучения форм этих отношений. В ограниченных обстоятельствах регрессионный анализ может быть использован для определения причинно-следственных связей между независимыми и зависимыми переменными. Однако, это может привести к иллюзиям или ложным выводам, поэтому рекомендуется соблюдать осторожность, например, чётко понимать, что корреляция не доказывает причинно-следственную связь.

Разработаны многие методы проведения регрессионного анализа. Известные методы, такие как линейной регрессии и привычная регрессия, построенная по методу наименьших квадратов, являются параметрическими, в том смысле, что регрессионная функция регрессии определена в терминах конечного числа неизвестных параметров, которые оцениваются по данным. Непараметрическая регрессия относится к методам, которые позволяют регрессионной функции принадлежать к заданному набору функций, который может быть бесконечномерным [2].

### 1.1.1 Одномерная линейная регрессия

Линейная регрессия это очень простой метод прогноза переменной  $Y$  на основе одного предиктора  $X$ . В основе линейной регрессии лежит предположение о том, что существует линейная зависимость между  $X$  и  $Y$ . Математически эту линейную зависимость можно записать следующим образом

$$Y \approx \beta_0 + \beta_1 X_1, \quad (1.1)$$

где

$Y$  — предиктант,

$X$  — потенциальный предиктор,

$\beta_j$  — коэффициент, который определяет связь между предиктором и предиктантом,  $j=0;1$ .

В уравнении (1.1)  $\beta_0$  и  $\beta_1$  представляют собой две неизвестные константы, которые характеризуют точку пересечения регрессионной прямой с осью ординат и тангенс угла наклона линейной модели, соответственно. Параметры  $\beta_0$  и  $\beta_1$  называются коэффициентами регрессионной модели [3].

### 1.1.2 Множественная линейная регрессия

Простая линейная регрессия - полезный подход для прогнозирования на основе одного предиктора. Однако на практике часто этого бывает недостаточно, в этом случае используют несколько предикторов – например, имеется три потенциальных предиктора, вместо одного. Одним из вариантов является разработка трёх отдельных моделей одномерной линейной регрессии, каждая из которых использует различные параметры в качестве предиктора. Однако разработка отдельной модели простой линейной регрессии для каждого предиктора не является полностью удовлетворительным подходом. Прежде всего, неясно, как сделать единый прогноз с учётом всех трёх потенциальных предикторов, поскольку каждый из них связан с предиктантом отдельным уравнением регрессии. Во-вторых, каждое из трёх уравнений модели регрессии игнорирует два других, при формировании оценок коэффициентов регрессии. Вместо того чтобы подгонять каждую отдельную простую модель линейной регрессии для отдельного предиктора, лучший подход состоит в том, чтобы расширить простую модель линейной регрессии так, чтобы она могла непосредственно учитывать несколько предикторов одновременно. Можно сделать это, предоставив каждому предиктору отдельный модельный коэффициент. Предположим, что есть  $p$  различных

предикторов. Тогда модель множественной линейной регрессии имеет следующий вид

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p, \quad (1.2)$$

где

$X_j$  —  $j$ -й предиктор,

$\beta_j$  — модельный коэффициент, который определяет связь между  $j$ -й предиктором и предиктантом,  $j=0, \dots, p$  [3].

## 1.2 Построение регрессионных моделей

### 1.2.1 Определение связи между предиктантом и предиктором

В простой линейной регрессии, чтобы определить, существует ли связь между предиктантом и предиктором можно просто проверить – не равен ли коэффициент наклона прямой нулю, т.е.  $\beta_1 = 0$ . В модели множественной регрессии на основе  $p$ -предикторов необходимо проверить – все ли коэффициенты регрессии равны нулю, т.е.  $\beta_1 = \beta_2 = \dots = \beta_p = 0$ . Для ответа на этот вопрос в обоих случаях проверяется нулевая гипотеза

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

и её альтернатива:

$H_a$ : по крайней мере, один коэффициент  $\beta_j$  не равен нулю.

Проверка выполняется вычислением  $F$ -статистики,

$$F = \frac{\frac{(TSS-RSS)}{P}}{\frac{RSS}{(n-p-1)}}, \quad (1.3)$$

где

$F$  — критерий Фишера,

$RSS = \sum (y_i - \hat{y}_i)^2$  — сумма квадратов остатков,

$TSS = \sum (y_i - \bar{y})^2$  — дисперсия, где:

$\bar{y}$  — среднее значение переменных ряда,

$\hat{y}_i$  — прогнозируемая величина,

$y_i$  — наблюдаемая величина, предиктор,

$n$  — длина ряда наблюдений,

$p$  — длина ряда потенциального предиктора.

Следовательно, когда нет никакой зависимости между предиктором и предиктантом,  $F$ -статистика примет значение близкое к 1. С другой стороны, если гипотеза  $H_a$  верна, то  $F$  будет больше 1. Другими словами, большое значение  $F$ -статистики предполагает, что, по крайней мере, один из потенциальных предикторов связан с предиктантом. Однако, встаёт вопрос: насколько велика должна быть  $F$ -статистика, чтобы можно было отклонить нулевую гипотезу  $H_0$  и сделать вывод о наличии связи предиктора и предиктанта? Оказывается, что ответ зависит от значения длины анализируемой выборки  $n$ . Когда  $n$  велико, критерий Фишера ( $F$ -статистика) немного больше 1, может быть доказательством против нулевой гипотезы  $H_0$ . Если анализируемый ряд короткий ( $n$  небольшое), напротив, только большое значения критерия Фишера ( $F$ -статистика) позволит отклонить нулевую гипотезу  $H_0$ .

### 1.2.2 Выбор предикторов

Первым шагом построения модели множественной регрессии является вычисление  $F$ -статистики. Если делается вывод, что, по крайней мере, один из потенциальных предикторов связан с предиктантом, то необходимо выяснить какой именно. Возможно, что все потенциальные предикторы связаны с предиктантом, но чаще бывает, что предиктант связан только с некоторым подмножеством предикторов. Задача определения того, какие предикторы

связаны с предиктантом, чтобы создать одну модель на основе только этих предикторов, называется выбором переменных. В идеале необходимо выполнить выбор переменных, опробовав множество разных моделей, каждая из которых содержит разное подмножество предикторов. Например, если есть два потенциальных предиктора  $X_1$  и  $X_2$  ( $p=2$ ), то можно рассмотреть четыре модели:

- 1) модель, не содержащую ни одну из переменных  $X$ ,
- 2) модель, содержащая только  $X_1$ ,
- 3) модель, содержащая только  $X_2$ ,
- 4) модель, содержащая обе переменных и  $X_1$  и  $X_2$ .

Затем можно выбрать лучшую модель из всех рассмотренных моделей.

Как определить, какая модель лучше? Для оценки качества модели можно использовать различные статистические данные. К ним относятся

- 1) информационный критерий Akaike (AIC) [4],
- 2) байесовский информационный критерий (BIC) [5],
- 3) коэффициент детерминации  $R^2$  [6].

К сожалению, в общем случае существует  $2^p$  модели, которые содержат подмножества потенциальных предикторов. Это означает, что даже при умеренном количестве первоначальных предикторов, опробовать все возможные подмножества предикторов невозможно. Например, как было показано выше, если существует всего лишь два потенциальных предиктора, то возможно рассмотреть все 4 потенциальные модели. Но если количество предикторов увеличить до 30 ( $p=30$ ), то необходимо рассмотреть  $2^{30} = 1\,073\,741\,824$  моделей! Это невозможно. Поэтому, если потенциальных предикторов много, то необходим автоматизированный и эффективный подход к выбору меньшего набора моделей для анализа их качества. Для решения этой задачи существуют три классических подхода:

- 1) Подход прямого выбора. Начинают с нулевой модели - модели, в которой присутствует только значение точки пересечения с координатной

осью и не используют потенциальных предикторов. Затем определяют  $p$  простых линейных регрессий и добавляют к нулевой модели переменную, которая приводит к самой низкой сумме квадратов остатков  $RSS$ . Затем добавляют к этой модели переменную, которая приводит опять к самому низкому  $RSS$  для новой модели с двумя переменными. Это добавление новых предикторов продолжается до тех пор, пока не будет выполнено некоторое условие для завершения подбора. Это шаговая прямая регрессия.

2) Подход обратного выбора. Начинают создание модели, включая в неё сразу все потенциальные предикторы, и удаляют из модели предиктор, который является наименее статистически значимой. Образуется новый набор потенциальных предикторов, среди которых снова производится выбор наименее статистически значимого предиктора, который удаляется. Эта процедура продолжается до тех пор, пока не будет достигнуто некое условие для остановки. Например, можно остановиться, когда все оставшиеся предикторы имеют статистическую значимость ниже какого-то порогового значения какого-либо оценочного коэффициента. Это шаговая обратная регрессия.

3) Подход смешанного выбора. Это комбинация первого и второго подходов. Начинают, как при прямом выборе, с отсутствия предикторов в модели и добавляют предиктор, который наилучшим образом подходит, то есть лучше аппроксимирует предиктант. Далее продолжают добавлять предикторы поочерёдно. Качество аппроксимации предиктанта может стать лучше по мере добавления новых предикторов к модели, а может ухудшаться. Следовательно, если в какой-либо точке алгоритма качество аппроксимации ухудшается, то этот предиктор удаляется из модели. Продолжают выполнять эти шаги вперед и назад, пока все переменные в модели не будут иметь достаточно низкое значение оценки уровня значимости.

Подход обратного выбора не может быть использован, если количество потенциальных предикторов больше, чем длина рядов, в то время как подход

прямого выбора всегда можно использовать. Подход прямого выбора может включать в себя переменные, которые позже станут излишними. Подход смешанного выбора может исправить это.

### *Выбор подходящей модели*

Двумя наиболее распространёнными численными мерами пригодности модели являются  $RSE$  и  $R^2$ , где  $RSE$  – это стандартная ошибка остатков,  $R^2$  – это коэффициент детерминации, которые объясняют долю описанной моделью дисперсии. Эти величины вычисляются следующим образом

$$RSE = \frac{\sqrt{RSS}}{(n - 2)}, \quad (1.4)$$

$$R^2 = 1 - \frac{RSS}{TSS}, \quad (1.5)$$

Коэффициент детерминации  $R^2$  является квадратом коэффициента корреляции между предиктантом и предиктором. При использовании модели множественной линейной регрессии  $R^2$  равен квадрату коэффициента корреляции между предиктантом и результатом использования прогностической модели множественной регрессии. Одно из свойств линейной модели состоит в том, что она максимизирует эту корреляцию между всеми возможными линейными моделями.

Значение  $R^2$ , близкое к 1, указывает, что модель объясняет большую часть дисперсии предиктанта. Оказывается, что  $R^2$  всегда будет увеличиваться, когда к модели добавляются больше переменных, даже если эти переменные слабо связаны с предиктантом. Это связано с тем, что добавление новой переменной в уравнение всегда позволяет более точно подобрать данные для обучения модели. Таким образом, статистика  $R^2$ , которая также вычисляется по данным обучения, должна увеличиваться.

В дополнение к рассмотренным статистическим параметрам RSE и  $R^2$  можно так же отметить, что полезно представить данные графическим способом. Графики могут выявить проблемы с моделью, которые не видны из числовой статистики [3].

### 1.3 Техника Model Output Statistics (MOS)

Другой метод уточнения результатов гидродинамического прогноза, который называют MOS (статистическая модель результатов моделирования) состоит в определении статистической связи между предиктантом и переменными, которые также являются результатом гидродинамического прогноза.

MOS техника – это определение связанной с погодой статистики результатов гидродинамического моделирования. Например, необходимо знать: сколько процентов времени идет дождь, когда модель предсказывает относительную влажность 80% или какая будет лучшая оценка ветра на высоте 10 м в аэропорту, если модель предсказывает геострофический ветер на высоте 1000 гПа в этой точке [7].

MOS техника подразумевает сопоставление наблюдаемых погодных элементов (предиктанты) с соответствующими переменными (предикторы) на основе статистического подхода. Предикторы могут быть получены из гидродинамических моделей прогноза погоды, предыдущих метеорологических наблюдений или геоклиматической информации. В настоящее время MOS техника использует множественную линейную регрессию, как статистический метод. Такой подход является математически простым, но достаточно мощным. Для улучшения его качества используют достаточно продолжительные ряды наблюдений и прогноза. Как и у всех методов технике MOS присущи, как достоинства, так и недостатки.

#### Преимущества:

- 1) Признание существования модельной предсказуемости.
- 2) Удаление систематических ошибок модели.
- 3) Оптимальный выбор предикторов.
- 4) Повышение оправдываемости гидродинамических прогнозов.
- 5) Широкий спектр метеорологических величин и явлений, участвующих в анализе

#### Недостатки:

- 1) Короткие выборки не позволяют получить статистически значимые результаты корректировки прогнозов.
- 2) Любое изменение в гидродинамической модели требует новых разработок MOS.
- 3) Для успешного MOS требуется достаточное количество результатов наблюдений и из высокое качество (малые ошибки наблюдений) [8].

## 2 Гидродинамическая мезомасштабная модель WRF

В данном исследовании корректировались результаты моделирования атмосферных процессов гидродинамической мезомасштабной моделью WRF.

WRF (Weather Research and Forecasting - WRF) - модель для метеорологических исследований и прогноза – представляет собой атмосферную модель, разработанную, как следует из названия, как для исследования, так и для прогноза погоды (ЧПП). Хотя она официально поддерживается Национальным центром атмосферных исследований (NCAR), WRF стала настоящей моделью мирового метеорологического сообщества и существует уже продолжительное время благодаря интересам и вкладу мировой пользовательской базы исследователей. Модель WRF может предоставлять специальные возможности для очень широкого спектра исследований геосистемы – химия атмосферы, гидрология, пожары в дикой природе, ураганы и региональный климат. Структура программного обеспечения WRF поддерживает эффективные параллельные вычисления на широком диапазоне вычислительных платформ (Skamarock et al., 2008).

Со времени своего первого публичного выпуска в 2000 году WRF стала, вероятно, самой используемой в мире моделью атмосферы. Об этом свидетельствуют число её зарегистрированных пользователей и публикаций. Например, совокупное количество заявок на вступление в число зарегистрированных пользователей WRF на 2013 год составляло более 36 000 человек из 162 стран. В каталоге публикаций, относящихся к WRF, говорится о влиянии модели на науку. Количество опубликованных журнальных публикаций с участием WRF на 2016 год составляло более 3500, а среднегодовой показатель за 2011-2015 годы - 510 публикаций в год. Количество уникальных учреждений в рецензируемых публикациях WRF составляет более 1340, а число уникальных авторов превышает 11700 человек.

К 2016 году число ссылок на документы WRF составляет более 26 500, в среднем более 10 цитирований за публикацию. Хотя модель WRF уже достаточно сформирована, она продолжает развиваться. Система активно применяется в новых направлениях исследований, используется в прогнозах в реальном времени и широко представлена на рынке геофизических услуг[9].

## 2.1 Модель WRF и ее применение

WRF осуществляет моделирование атмосферы. Процесс состоит из двух этапов: сначала необходимо настроить домен(ы) модели, ввести входные данные и подготовить начальные условия, а затем - запустить модель прогноза.

Компоненты прогнозируемой модели работают в рамках программного обеспечения WRF, которое обрабатывает операции ввода-вывода и параллельных вычислений. Программный код WRF написан в основном на языке программирования Fortran. Модель может собираться несколькими компиляторами и работает преимущественно на платформах с UNIX-подобными операционными системами от ноутбуков до суперкомпьютеров. Архитектура WRF позволила переносить ее практически на все типы платформ на 500 лучших суперкомпьютерах мира.

Моделирование WRF начинается с WRF Preprocessing System (WPS). Ряд утилит WPS сначала извлекает географическую информацию (например, характеристики рельефа и подстилающей поверхности) для создания пользовательских доменов модели. Затем принимают, переформатирует и интерполирует требуемые данные (например, глобальный анализ или прогноз материнской модели) в пользовательские домены. Наконец, начальные поля интерполируются на вертикальные уровни модели и генерируются боковые граничные условия. После этого модель WRF готова к модельным расчётам.

Это осуществляется компонентой прогноза, которая содержит динамическое ядро и пакеты параметризаций физических процессов

(например, микрофизика облаков, коротковолновое и длинноволновое излучения, процессы в планетарном пограничном слое).

По своей конструкции в качестве инструмента исследования WRF также может быть сконфигурирована для проведения исследований идеализированных ситуаций. Эта возможность позволяет пользователям изучать процессы в упрощенной настройке (например, изучать отдельные процессы или идеализированную топографию) путем изменения параметров и начальных условий при использовании описания избранных физических процессов.

В настоящее время модель WRF предоставляет на выбор 12 идеализированных сценариев, включая бароклинные волны, конвективные процессы, поток над разнообразной орографией, большие вихри и тропические циклоны. Кроме того, пользователи могут легко построить другие идеализированные конфигурации.

WRF также может быть запущена как глобальная модель на широтно-долготной сетке. Несмотря на то, что глобальная WRF была первоначально построена для изучения атмосфер других планет, она также стала использоваться для моделирования земных процессов, химии атмосферы и климатических исследований.

Как в реальных данных, так и в идеализированных конфигурациях WRF широко используется для исследований. Поскольку WRF является в основном мезомасштабной моделью, исследования WRF на протяжении многих лет охватывают разнообразные темы в мезомасштабной метеорологии: синоптические и мезомасштабные процессы, связанные с внетропическими циклонами, фронтами и струйными течениями; мезомасштабные погодные процессы и явления; организованная конвекция; ураганы. В последние годы использование WRF для региональных исследований климата увеличилось и в этом случае WRF описывает процессы мелкомасштабной атмосферы и земной

поверхности лучше, чем глобальные модели традиционно используемые для климатических прогнозов.

Система ассимиляции модели WRF (WRF Data Assimilation System – WRFDA) является основной системой подготовки начальных данных для WRF. Она реализует на выбор пользователя трехмерный и четырехмерный вариационные (3DVAR, 4DVAR) подходы, а также гибридный вариационно-ансамблевый подход – ансамблевый фильтр Калмана (ETKF). Эти методы позволяют ассимилировать широкий спектр результатов прямых и косвенных наблюдения, от традиционных данных на поверхности и данных зондирования до спутниковых измерений.

Предназначенная для гидродинамических прогнозов WRF используется оперативно в государственных прогностических центрах по всему миру, а также частными компаниями и лицами. Конфигурируемость доменов с высоким разрешением, множество возможных входных данных и вычислительная гибкость (особенно в условиях ограниченного ресурса), а также возможность использовать модельные достижения от мирового исследовательского сообщества сделали ее особенно привлекательной для прогнозирования в реальном времени.

NCEP использует WRF для поддержки Национальной метеорологической службы в ряде систем: WRF (ARW) запускается в системах быстрого обновления (RAP) и быстрого обновления высокого разрешения (HRRR), разработанных NOAA ESRL и NCEP. Предоставляя преимущества ежечасного циклического обновления, системы RAP и HRRR оснащены сетками 13 и 3 км по всей Северной Америке и Соединенным Штатам (CONUS). Приложения HRRR и RAP включаются в системы, решающие задачи энергетики, гидрологии, предупреждения об опасных явлениях погоды, авиационной безопасности и контроля качества воздуха, а также и частично связаны с несколькими специализированными системами WRF [9].

WRF используется в университетах по всему миру в качестве средства для обучения и исследований в дипломных работах. Основываясь на опросе профессоров в институтах-членах Университетской корпорации по исследованию атмосферы (UCAR), способы использования этой системы включают в себя инструкцию по ЧПП, численные методы и атмосферные процессы; исследовательские проекты; прогнозы погоды в режиме реального времени и модель для лекций и учебных материалов. Для исследований, как выпускники, так и студенты могут с легкостью применять WRF для изучения спектра проблем исследования атмосферы, используя свои идеализированные, реальные данные или специальные возможности. Кроме того, они могут быстро начать заниматься научными вопросами без необходимости изобретать и строить сложную модель в качестве инструмента исследования. WRF может быть легко установлена и запущена на обычном компьютерном оборудовании (то есть, на рабочих станциях, кластерах и ноутбуках), что является ещё одним фактором использования ее в образовании [9].

## 2.2 Характеристики используемой WRF

В данном исследовании использовались результаты моделирования атмосферных процессов на территории Словакии с использованием следующей конфигурации модели WRF-ARW:

- 1) Модель реализовывалась с использованием телескопизации. Материнский домен имеет шаг по горизонтали 50 км, два вложенных 16,6 км и 5,5 км, соответственно.
- 2) Используется ламбертова проекция декартовой системе координат.
- 3) Шаг по времени на материнском домене 300 сек, на вложенных 100 и 33 секунды.
- 4) Начальные данные получают из реанализа NCEP/NCAR с разрешением 0.5 градусов. Граничные условия обновляются каждые 3 часа.

Анализируются результаты на самой мелкой сетке, которая охватывает всю территорию Словакии (Приложение А).

Анализируются результаты прогноза за период с января 2012 года по сентябрь 2013 года в следующих городах: Братислава, Дудинце, Жилина, Хричов, Каменица на Цирохоу, Койшовская Хола, Кошице, Ломницки-Штит, Лученец, Льесек, Милхостов, Попрад-Гановце, Попрад-Татры, Прьевидза, Пьештяни, Стропков, Тисинец, Телгарт, Хопок, Штрбске Плесо. Месторасположение всех станций представлены на рисунке в Приложении А и обозначены синими точками с номерами (с левого края перечислены станции с их высотой над уровнем моря).

В исследовании используются результаты измерения температуры воздуха на высоте 2 м, на указанных выше метеорологических станциях. Дискретность измерений и прогнозов 1 час. Итого в анализе участвует более 20000 измерений для каждой станции.

В анализе участвовали модельные данные о следующих метеорологических величинах:

- геопотенциальная высота,
- температура воздуха,
- температура точки росы,
- относительная влажность,
- компоненты скорости ветра.
- на следующих изобарических поверхностях: 500, 700, 850, 925, 950, 975, 1000 гПа и на уровне 2 м над поверхностью земли.

Дополнительно по прогностическим данным были рассчитаны

– температура точки росы на всех уровнях,

– индекс Боудена (индекс для оценки вероятности грозы),

$$BI = 0.1(H_{700} - H_{1000}) - T_{700} - 200,$$

где

$H_{700}$ - геопотенциальная высота на уровне 700 гПа,

$H_{1000}$ - геопотенциальная высота на уровне 1000 гПа,

$T_{700}$ - температура воздуха в градусах Цельсий на уровне 700 гПа.

– индекс статической устойчивости (Vertical Totals),

$$VT = T_{850} - T_{500},$$

где

$T_{850}$ - температура воздуха в градусах Цельсий на уровне 850 гПа,

$T_{500}$ - температура воздуха в градусах Цельсий на уровне 500 гПа.

– индекс Cross Totals,

$$CT = T_{d850} - T_{500},$$

где

$T_{d850}$ - температура точки росы в градусах Цельсий на уровне 850 гПа,

$T_{500}$ - температура воздуха в градусах Цельсий на уровне 500 гПа;

– разница между значениями метеорологических величин на вертикальных уровнях: 500 и 700, 500 и 850, 700 и 850, 500 и 925, 700 и 925, 850 и 925, 500 и 950, 700 и 950, 850 и 950, 925 и 950, 500 и 975, 700 и 975, 850 и 975, 950 и 975, 500 и 1000, 700 и 1000, 850 и 1000, 925 и 1000, 950 и 1000, 975 и 1000 гПа;

а так же для геопотенциальной высоты: 500 гПа и уровень 1 м над землей, 700 гПа и уровень 1 м над землей, 850 гПа и уровень 1 м над землей, 925 гПа и уровень 1 м над землей, 950 гПа и уровень 1 м над землей, 975 гПа и уровень 1 м над землей, 1000 гПа и уровень 1 м над землей;

а так же для температуры воздуха и относительной влажности воздуха: 500 гПа и уровень 2 м над землей, 700 гПа и уровень 2 м над землей, 850 гПа и уровень 2 м над землей, 925 гПа и уровень 2 м над землей, 950 гПа и уровень 2 м над землей, 975 гПа и уровень 2 м над землей, 1000 гПа и уровень 2 м над землей;

а так же для компонент ветра U и C: 500 гПа и уровень 10 м над землей, 700 гПа и уровень 10 м над землей, 850 гПа и уровень 10 м над землей, 925 гПа и уровень 10 м над землей, 950 гПа и уровень 10 м над землей, 975 гПа и уровень 10 м над землей, 1000 гПа и уровень 10 м над землей.

Итого из результатов моделирования было получено 168 временных рядов, в каждом из которых более 20000 строк с данными.

Массивы предикторов и предиктанта рассматривались, как временные ряды, по которым оценивалась значимость предиктора и качество построенной регрессионной модели [10].

### 3 Численные эксперименты и анализ результатов

#### 3.1 Данные, используемые в исследовании, и их предварительная обработка

Как уже говорилось выше, целью исследования была разработка алгоритма уточнения гидродинамического прогноза температуры воздуха на уровне 2м в 18 городах Словакии.

Для коррекции прогностической приземной температуры использовались прогностические значения метеорологических величин (геопотенциальная высота, влажность, температура воздуха) на изобарических поверхностях от 1000 до 500 гПа.

Верификации проводилась по результатам наблюдений за температурой воздуха на 18 метеорологических станциях, представленных на рисунке в Приложении А.

Всего в исследовании анализировалось 3024 временных ряда метеорологических величин, содержащих значения с дискретность один час. Длина каждого ряда в среднем составляет 20000 значений, что позволяет получить статистически значимые результаты.

В исследовании использовался язык программирования Python, который технической точки зрения представляет собой объектно-ориентированный, высокоуровневый язык программирования со встроенной динамической семантикой и удобен, в первую очередь, для разработки веб-приложений. Python чрезвычайно привлекателен в области Rapid Application Development, поскольку предлагает динамическую типизацию и возможности динамического привязки. Язык программирования Python относительно прост, поэтому его легко освоить, поскольку для него требуется уникальный синтаксис, который фокусируется на удобочитаемости. Разработчики могут читать и переводить код Python намного проще, чем другие языки. В свою

очередь, это снижает затраты на обслуживание и разработку программ, поскольку позволяет командам работать совместно без существенных языковых барьеров и препятствий. Кроме того, Python поддерживает использование модулей и пакетов, что означает, что программы могут быть разработаны в модульном стиле, и код может быть повторно использован во множестве проектов. После разработки необходимого модуля, его можно масштабировать для использования в других проектах, легко импортировать или экспортировать [11].

Python - это язык программирования общего назначения, что говорит о том, что его можно использовать практически для всего. Большинство веб-приложений и платформ полагаются на Python, включая поисковую систему Google, YouTube и веб-ориентированную транзакционную систему Нью-Йоркской фондовой биржи. NASA использует именно Python для программирования оборудования и управления космической техники. Python также может использоваться для обработки текста, отображения номеров или изображений, решения научных уравнений и сохранения данных [1].

Всё вышесказанное является обоснованием выбора именно языка программирования Python в данном исследовательском проекте.

### 3.1.1 Предварительная обработка результатов моделирования мезомасштабной гидродинамической модели WRF

Работа началась с подготовки данных к обработке. В одной общей директории находились папки, с результатами моделирования за один срок. Названиями папок служат дата и время начала прогноза. Организация директории с результатами моделирования представлена на рисунке 3.1.



Рисунок 3.1 – Отображение директорий с данными результатов моделирования

В каждой директории с результатами прогноза за один срок находятся файлы с обозначением номера и названия станции. Вид одной из таких директорий приведён на рисунке 3.2.

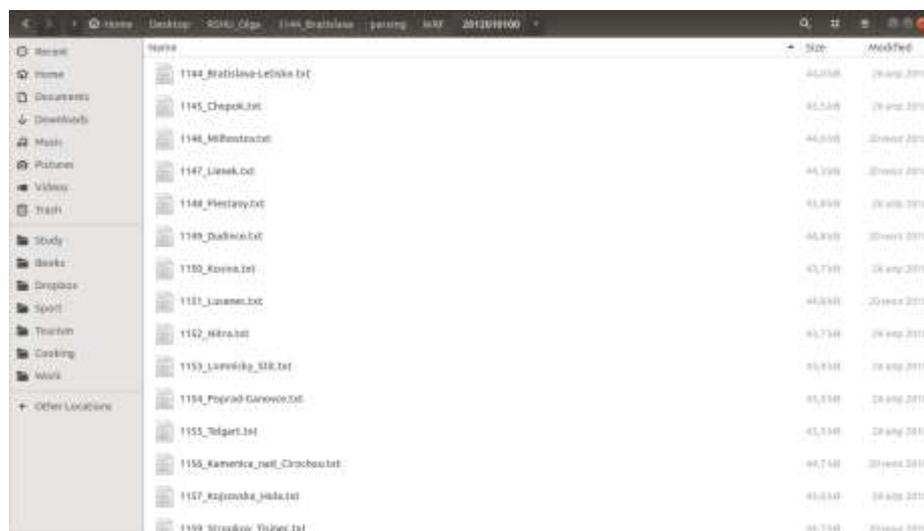


Рисунок 3.2 – Отображение файлов с результатами прогноза в один срок прогноза

В первоначальном виде данные результатов моделирования WRF имеют вид таблиц (рисунок 3.3).

Date UTC	TIME_H_MIL	geopotential height 500	temperature 500	relative humidity 500	vertical velocity in Pa/s 500	wind u-component 500	wind v-component 500	absolute vorticity 500
2012-01-01 00:00:00	1325376000000	5521.35429766073	249.420294064316	82.7845718419409	7.61726177320882	-19.5066900173506		
2012-01-01 01:00:00	1325379600000	5525.554690000997	249.4322786999886	81.8139888627561	7.955094435297883	-18.2832098320956		
2012-01-01 02:00:00	1325383200000	5531.89551940181	249.560420664853	79.5546761356561	8.07362944454615	-18.8956221417187		
2012-01-01 03:00:00	1325386800000	5533.78395114551	249.517023640159	77.9758915499905	8.08550282554104	-15.564682740322		
2012-01-01 04:00:00	1325390400000	5537.03995227982	249.52944491847	78.4860003498937	7.25441801554057	-15.1506947032842		
2012-01-01 05:00:00	1325394000000	5542.7052529496	249.700606401188	67.1162999507817	7.01347993821291	-14.9294300094978		
2012-01-01 06:00:00	1325397600000	5550.8834281838	249.385232555296	61.8828344781415	6.89683453519344	-15.1129713864706		
2012-01-01 07:00:00	1325401200000	5557.91355907064	250.2488711743	53.139550421778	6.55821177783846	-15.368954640951		
2012-01-01 08:00:00	1325404800000	5564.98824966673	250.614448463511	45.3503936258968	6.38111951223442	-15.6815035174782		
2012-01-01 09:00:00	1325408400000	5573.28524257514	250.94928551821	39.0676954821587	6.768826585252606	-16.1083139363175		
2012-01-01 10:00:00	1325412000000	5581.14742571502	251.373819840526	35.4802448895303	7.12018055860307	-17.1182022563701		
2012-01-01 11:00:00	1325415600000	5591.34852002728	251.895278401931	33.6015458627200	7.30079898836303	-17.4934184480165		
2012-01-01 12:00:00	1325419200000	5591.53407936183	252.086957602389	28.7684630679894	7.41377236570288	-17.5375212351600		
2012-01-01 13:00:00	1325422800000	5597.26193522598	252.456111569529	27.8209896506604	8.010520438204821	-16.9682151668918		
2012-01-01 14:00:00	1325426400000	5602.72899735622	252.846707963914	32.3431945006848	7.93458315697810	-15.033895000181		
2012-01-01 15:00:00	1325430000000	5609.33841112936	253.082782211668	35.7700415009021	8.89482670697785	-13.1931209602704		
2012-01-01 16:00:00	1325433600000	5613.94330732221	252.713851304925	42.0402917807059	9.29771729100237	-12.183529664253		
2012-01-01 17:00:00	1325437200000	5618.48178694701	252.524031309509	69.7458877521458	10.581523418282	-11.4357161943502		
2012-01-01 18:00:00	1325440800000	5626.42738242104	252.208075668984	71.6904000423407	11.886445110526	-10.748685789078		
2012-01-01 19:00:00	1325444400000	5629.85833666013	252.108130722842	99.0886767792648	13.0567643189078	-6.0704789104124		
2012-01-01 20:00:00	1325448000000	5632.39460201003	251.188885557856	83.9020570579405	13.9523151296056	-7.16973879111588		
2012-01-01 21:00:00	1325451600000	5632.37892488425	252.425212586842	64.1348185003141	13.3628913620387	-7.2149944510043		
2012-01-01 22:00:00	1325455200000	5634.70298514923	252.135525895051	60.1450488493915	12.3738703114382	-7.2996405842943		
2012-01-01 23:00:00	1325458800000	5638.04793792852	252.149750940783	57.702720838721	12.3867109991414	-6.89641309076415		
2012-01-02 00:00:00	1325462400000	5638.67908205026	253.12138072257	44.0181288874900	11.2302118020906	-7.07748838423887		
2012-01-02 01:00:00	1325466000000	5637.88948861575	253.583990234438	35.5820915116371	11.0399461788782	-5.17356904817832		
2012-01-02 02:00:00	1325469600000	5637.88613069689	253.901384096526	28.9175714788914	12.2065617848782	-3.4349200321011		
2012-01-02 03:00:00	1325473200000	5636.5561629182	254.330283029857	22.068801739907	13.8123958221012	-2.32301564444423		
2012-01-02 04:00:00	1325476800000	5635.48530826304	254.402206441865	21.151351697774	15.3172628655551	-1.01501506069997		

Рисунок 3.3 – Таблица результатов моделирования для станции Братислава для 1 января 2012 года

В каждом файле для каждой станции имеется 49 строк. В первой строке название переменных моделирования, а в строках 2 – 49 – сами результаты моделирования с дискретностью час на 48 часов.

Первой задачей было собрать в определенной последовательности в один файл все данные за один период. В итоговом файле должны находиться данные сначала за 10 часов первого дня, потом за 11 часов первого дня, ..., 23 часа первого дня, потом 00 часов второго дня и т. д. Заканчиваться файл должен 23 часами второго дня. Анализ начинается с 10 часов, поскольку первые 10 часов моделирования происходит подстройка модели (spin-up) и данные не имеют прогностических перспектив.

Как уже отмечалось, для выполнения задачи был использован язык программирования Python, на котором написано несколько программных

кодов. С помощью первого кода, просматривая каждую папку и файл с конкретным названием станции, создаются файлы для каждого часа отдельно. С помощью второго кода, путём склеивания файлов за каждый час, получается итоговый файл, пример которого приведён на рисунке 3.4.

Date UTC	TIME IN MIL	geopotential height 500	temperature 500	relative humidity 500	vertical velocity in Pa/s 500	west u-component 500	west v-component 500	absolute vorticity 500
2012-01-01 10:00:00	1325412000000	5581.17472571502	251.373819849526	35.4802448995303	7.12019955585207	-17.1192022543701		
2012-01-02 10:00:00	1325498400000	5627.44446617195	252.952545700028	49.3328790771749	14.3540151193859	10.7973448157058		
2012-01-03 10:00:00	1325584800000	5594.49997213253	250.58535156721	27.0511024691085	24.86142235180021	-1.714453403405030		
2012-01-05 10:00:00	1325757600000	5370.26956003984	250.821466099493	87.4684563110464	31.0670158557278	-3.37151672521274		
2012-01-06 10:00:00	1325844000000	5268.07534785418	240.989912115099	77.2133669917875	0.707325221379627	-11.05037200386569		
2012-01-08 10:00:00	1326016800000	5274.21659712958	240.4920845950625	60.3544073818259	4.429986704816932	-1.88244574177159		
2012-01-09 10:00:00	1326103200000	5440.42961188304	242.745304189811	15.534814579711	-1.7112974144775	-17.7227013344001		
2012-01-10 10:00:00	1326189600000	5480.189990189709	242.58164569876	8.3678525902544	-3.45404670746383	1.2178958012903		
2012-01-11 10:00:00	1326276000000	5661.32544053487	249.50085622342	24.4794372399713	2.88851362859223	-20.8037792812668		
2012-01-12 10:00:00	1326362400000	5650.50407063429	250.21972692571	46.2508162911078	-12.332007400143	-8.7900808307889		
2012-01-13 10:00:00	1326448800000	5342.00596001263	237.2571253444347	60.6821679361245	23.84547052155969	-19.324877112679		
2012-01-14 10:00:00	1326535200000	5312.30343638118	236.611291112856	49.419364889471	3.88201038670095	-18.7344334101511		
2012-01-15 10:00:00	1326621600000	5333.74625220377	238.1099138793421	28.8914873428077	-1.75325739671975	-26.73325490317952		
2012-01-16 10:00:00	1326708000000	5286.09700544682	241.600043410959	12.0363786603302	-5.28766479517079	-29.2530935365177		
2012-01-17 10:00:00	1326794400000	5488.975164887	242.00077701448	68.2395379911634	2.08647623312536	-24.1837904182890		
2012-01-18 10:00:00	1326880800000	5540.40915150111	244.276881724279	61.4903662032177	3.252146325701109	-20.1058809957856		
2012-01-19 10:00:00	1326967200000	5582.87781568981	250.548712018177	82.0589440159469	22.5081984827988	-15.7373042252185		
2012-01-20 10:00:00	1327053600000	5353.55997107202	241.436218937012	93.28505604505983	26.5614313840327	3.77806601681437		
2012-01-21 10:00:00	1327140000000	5377.34795289996	240.09396023548	25.4001523733863	6.0303884170039	-30.1064281385379		
2012-01-22 10:00:00	1327226400000	5389.78721258074	242.879480717884	25.799606883972	25.3670952542387	-29.0890007622422		
2012-01-23 10:00:00	1327312800000	5409.37153782247	244.648560617265	11.328814298444	25.15885901383058	-14.0423657841486		
2012-01-25 10:00:00	1327485600000	5380.52399413021	238.881116150037	23.0311126512864	-4.64882522660318	-10.9176134472328		
2012-01-26 10:00:00	1327572000000	5382.20890522096	245.134850900171	3.00188220331716	-0.895818157505151	-28.1147648985608		
2012-01-27 10:00:00	1327658400000	5560.003138166985	249.929670527153	15.7399068890715	6.07829796745467	-8.79356220746626		
2012-01-28 10:00:00	1327744800000	5567.35963388772	245.797066885332	52.7884880392317	4.17918473677253	-0.862952669503584		
2012-01-29 10:00:00	1327831200000	5542.56776778803	245.798814541823	54.8535569420013	-2.5436549828054	-0.207352887543947		
2012-01-30 10:00:00	1327917600000	5512.18635647847	243.143941804205	19.5000517967366	-1.79847464490482	-3.91818917954817		
2012-01-31 10:00:00	1328004000000	5484.71194963317	240.664954924574	16.6975241279007	-2.79232020166703	-1.62848080088821		
2012-02-01 10:00:00	1328090400000	5381.91289030513	238.883728048553	17.8996745683837	6.22412201123143	-2.87917659125071		

Рисунок 3.4 – Отображение итогового файла для результатов моделирования моделью WRF

Таким образом, все необходимые данные собраны в один файл, который и участвовал в дальнейшей обработке.

### 3.1.2 Предварительная обработка данных синоптических сообщений

В отличие от результатов моделирования синоптические сообщения находились в одном файле, но так же для каждого часа. Пример файла с синоптическими данными представлен на рисунке 3.5.

	B	C	D	E
1	Bratislava / M. N. 201203-01-01 08:30:00	Bratislava / M. N. 201203-01-01 09:00:00	Bratislava / M. N. 201203-01-01 09:30:00	Bratislava / M. N. 201203-01-01 10:00:00
2	SYNOPT message	SYNOPT message	SYNOPT message	SYNOPT message
3	2012-03-01 08:30:00	7	2700	8
4	2012-03-01 09:00:00	7	4000	5
5	2012-03-01 09:30:00	5	2300	9
6	2012-03-01 10:00:00	5	1800	6
7	2012-03-01 10:30:00	2	600	6
8	2012-03-01 11:00:00	8	400	5
9	2012-03-01 11:30:00	8	400	5
10	2012-03-01 12:00:00	5	1100	7
11	2012-03-01 12:30:00	4	1300	7
12	2012-03-01 13:00:00	2	900	7
13	2012-03-01 13:30:00	8	1100	7
14	2012-03-01 14:00:00	8	1100	7
15	2012-03-01 14:30:00	2	800	7
16	2012-03-01 15:00:00	8	900	7
17	2012-03-01 15:30:00	8	900	7
18	2012-03-01 16:00:00	8	900	7
19	2012-03-01 16:30:00	8	900	7
20	2012-03-01 17:00:00	8	900	7
21	2012-03-01 17:30:00	4	300	7
22	2012-03-01 18:00:00	1	100	7
23	2012-03-01 18:30:00	1	800	7
24	2012-03-01 19:00:00	8	600	7
25	2012-03-01 19:30:00	1	300	7
26	2012-03-01 20:00:00	3	300	7
27	2012-03-02 00:00:00	4	300	7
28	2012-03-02 05:00:00	1	300	7
29	2012-03-02 10:00:00	1	300	7
30	2012-03-02 15:00:00	4	800	7

Рисунок 3.5 – Данные синоптических наблюдений на станции Братислава

Таким образом, чтобы получить итоговый файл с аналогичной структурой, как и файл результатов моделирования, необходимо было написать программный код только для работы с одним файлом, что существенно упрощает процедуру. Как и в случае с результатами моделирования, сначала из файла с синоптическими сообщениями были созданы файлы для каждого часа отдельно, а затем они склеены в один итоговый файл в необходимой последовательности данных. Пример итогового файла представлен на рисунке 3.6.

TIME2	Clouds 1 amount	Clouds 1 height	Clouds 1 type	C2a	C2b	C2c	C3a	C3b	C3c	Clouds amount	PressOFF	PressStation	RH	DewP	Temp	TempMax	TempMin	Ws
2012-01-01 10:00:00	8	120	7							8	1021.4	1004.4	92.968748	-1.4	0.4			4c
2012-01-02 10:00:00	2	120	7	4	180	7	5	8000	0	7	1017.6	1009.9	91.962996	0.4	1.7			5c
2012-01-03 10:00:00	1	480	8	4	3000	3				6	1025.7	1009.2	76.855383	3.9	7.9			20c
2012-01-05 10:00:00	2	840	8	4	1200	0	0	1380	0	7	1001.3	985	76.80947	0.9	4.8			20c
2012-01-06 10:00:00	1	780	8	5	1200	0				7	1000.3	984.1	98.262019	-1.4	6.2			30c
2012-01-08 10:00:00	2	660	8	4	840	0	5	3300	3	6	1014.6	988.1	74.860679	1.7	5.8			20c
2012-01-09 10:00:00	2	840	8	4	1200	0	7	2100	3	7	1022.9	1006.3	72.228656	0.7	5.3			20c
2012-01-10 10:00:00	1	660	8	0	900	0				7	1025.1	1008.5	73.886302	2	6.3			15c
2012-01-11 10:00:00	6	1200	6							6	1032.5	1015.8	68.88967	1.1	6.4			20c
2012-01-12 10:00:00	1	1500	6							1	1027.3	1010.8	52.082738	-0.6	8.7			20c
2012-01-13 10:00:00	2	1320	8	4	1500	0				5	1017.2	1000.7	56.266199	-1.5	6.6			30c
2012-01-14 10:00:00	4	1080	8	3	1320	0				5	1017	1000.4	61.182763	-3	3.8			20c
2012-01-15 10:00:00	1	630	8	4	2100	3				5	1021.1	1004.3	53.237527	-7	1.5			20c
2012-01-16 10:00:00	1	780	6	7	2100	3				7	1025.8	1008.8	62.004088	-6.8	-0.4			20c
2012-01-17 10:00:00	5	300	7	7	490	6				8	1023.8	1006.9	85.227985	-1.4	0.8			4c
2012-01-18 10:00:00	1	1380	6	7	1500	0				7	1031.4	1014.4	67.022585	-3.2	2.3			15c
2012-01-19 10:00:00	4	390	6	7	1500	0				8	1021.5	1004.8	83.764538	1.5	4			15c
2012-01-20 10:00:00	2	780	6	4	1320	0				7	1008.8	983.4	80.980525	2.2	5.2			20c
2012-01-21 10:00:00	1	720	8	3	2700	3				3	1016.9	1000.3	55.240002	-3.8	4.3			30c
2012-01-22 10:00:00	1	480	8	3	1380	6	7	2400	4	7	1009.2	992.9	67.848295	0.6	6.1			30c
2012-01-23 10:00:00	1	990	8							2	1013.2	996.9	53.150288	-0.6	8.4			50c
2012-01-25 10:00:00	1	960	6							1	1022.3	1009.6	54.148101	-5.1	3.3			25c
2012-01-26 10:00:00	2	900	8							2	1023.2	1006.3	59.061252	-6.2	0.9			15c
2012-01-27 10:00:00	1	810	8							1	1029.3	1012.2	62.583947	-6.6	-0.2			8c
2012-01-28 10:00:00	5	6000	0							5	1032.1	1014.9	73.212025	-5	-0.8			7c
2012-01-29 10:00:00	4	840	6	7	900	6				7	1032.5	1015.3	84.889499	-3.5	-1.3			4c
2012-01-30 10:00:00	2	6000	0							2	1033.1	1015.9	48.109526	-10.8	-1.2			15c
2012-01-31 10:00:00	1	6000	0							1	1031.8	1014.5	55.291981	-11	-3.3			15c
2012-02-01 10:00:00	2	6000	0							2	1028.9	1011.8	42.057095	-14.3	-3.2			10c

Рисунок 3.6 – Пример итогового файла данных синоптических сообщений для станции Братислава

Как можно увидеть из примера, приведённого на рисунке 3.6, названия столбцов переименованы для удобства дальнейшего анализа. В результате последовательность данных полностью аналогична последовательности итогового файла результатов моделирования WRF.

### 3.2 Выбор модели для определения лучших предикторов

Целью исследования была корректировка прогностической температуры, полученной на основе гидродинамической модели WRF.

Массивы предикторов и предиктанта рассматривались, как временные ряды, по которым строилась модель множественной регрессии и оценивалась значимость предикторов и качество построенной регрессионной модели.

При построении регрессионной модели предсказываемой величиной (предиктант) являлась температура воздуха на уровне 2 м над поверхностью земли из данных синоптических сообщений.

В качестве потенциальных предикторов рассматривались переменные из результатов моделирования WRF, а так же несколько переменных, рассчитанных с помощью формул. В итоге было рассмотрено 168 потенциальных предиктора. Они перечислены в п 2.2.

Для выполнения расчетов написан программный код на языке программирования Python. Были использованы несколько моделей для выбора наилучших предикторов. Результаты расчётов сопоставлялись между собой и была выбрана лучшая из них.

### 3.2.1 Описание, использованных программных моделей, и алгоритмов вычислений

#### *Встроенный модуль SelectKBest (библиотекаSKLearn)*

Модуль SelectKBest находится в библиотеке SKLearn. Чтобы использовать данный модуль, следует установить библиотеку SKLearn с помощью команды `pip install sklearn`, а модуль импортировать в самом программном коде с помощью команды `from sklearn.feature_selection import SelectKBest`.

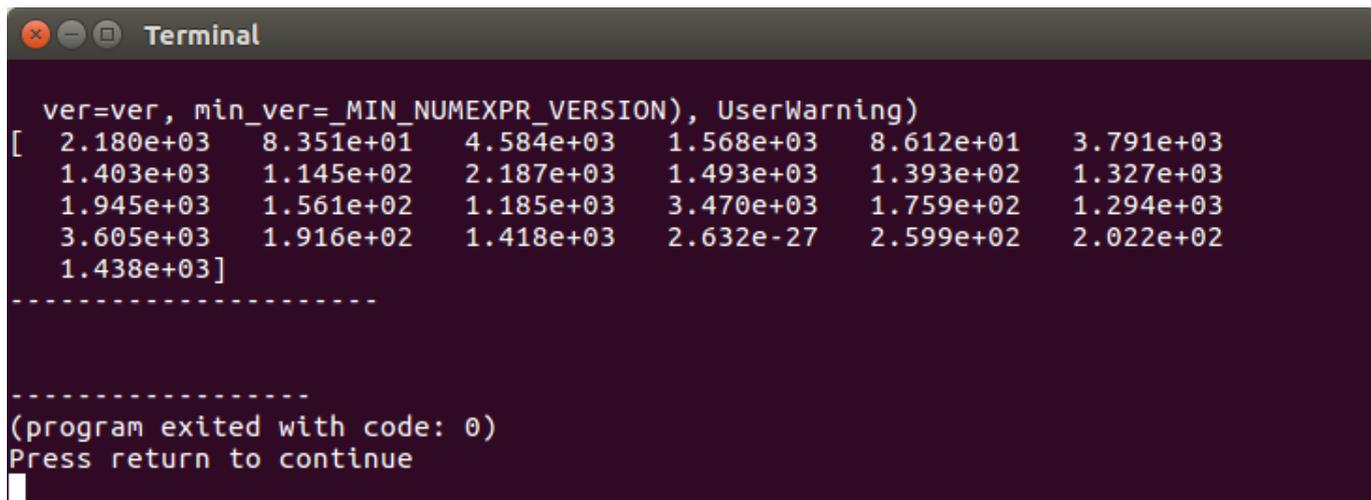
Для использования функции следует выполнить следующую команду:

```
result = SelectKBest(score_func="название функции", k=10)
```

Модуль SelectKBest принимает на вход два массива X и Y, где X — массив потенциальных предикторов с размерность  $168 \times$  на длину ряда (длина ряда зависит от станции, порядка 20000 значений для каждой из станций (минимальное количество значений приходится на станцию Койшовская Хола = 19865, максимальное количество значений приходится на станцию Братислава = 26371), Y — массив-предиктант с размерностью  $1 \times$  на длину ряда, и возвращает массив оценок пригодности данной величины к использованию в качестве предиктора. В скобках `score_func="название функции"` указывается имя оценочной функции, которую можно выбрать (по

умолчанию: *f\_classif*, что соответствует дисперсионной модели ANOVA). Параметр *k* определяет количество предикторов (по умолчанию: *k* = 10).

Выполнение команды *fit = result.fit(X, Y)* позволяет произвести оценку функции для каждого массива. Вывести рассчитанную оценку можно с помощью команды *print(fit.scores\_)*. Пример вывода приведён на рисунке 3.7.



```
ver=ver, min_ver=_MIN_NUMEXPR_VERSION), UserWarning)
[ 2.180e+03  8.351e+01  4.584e+03  1.568e+03  8.612e+01  3.791e+03
 1.403e+03  1.145e+02  2.187e+03  1.493e+03  1.393e+02  1.327e+03
 1.945e+03  1.561e+02  1.185e+03  3.470e+03  1.759e+02  1.294e+03
 3.605e+03  1.916e+02  1.418e+03  2.632e-27  2.599e+02  2.022e+02
 1.438e+03]
-----
(program exited with code: 0)
Press return to continue
```

Рисунок 3.7 – Вывод команды *print(fit.scores\_)* на экран

С помощью команды *fit.get\_support()* можно оценить каждый массив на значения True и False. True — данный параметр может использоваться в данном наборе как предиктор; False — данный параметр не может использовать в данном наборе как предиктор.

Пример результата выполнения команды представлен на рисунке 3.8.

```

Terminal
ver=ver, min_ver=_MIN_NUMEXPR_VERSION), UserWarning)
[ 2.180e+03  8.351e+01  4.584e+03  1.568e+03  8.612e+01  3.791e+03
 1.403e+03  1.145e+02  2.187e+03  1.493e+03  1.393e+02  1.327e+03
 1.945e+03  1.561e+02  1.185e+03  3.470e+03  1.759e+02  1.294e+03
 3.605e+03  1.916e+02  1.418e+03  2.632e-27  2.599e+02  2.022e+02
 1.438e+03]
-----
[False False  True False False False False False False False False False
 False False False False False False False False False False False
 False]
-----
(program exited with code: 0)
Press return to continue

```

Рисунок 3.8 – Вывод команды *fit.get\_support()* (ниже первой пунктирной линии)

Модуль *f\_classif*, установленный по умолчанию, проводит оценку качества регрессионной модели на основе критерия Фишера. Критерий Фишера предполагает оценку качества модели на основе сравнения выборочных дисперсий. Предполагается, что сравниваемые данные имеют нормальное распределение и дисперсия случайных величин одинакова. Тест проводится сравнением рассчитанного значения с критическим, соответствующего распределения Фишера при заданном уровне значимости

$$f = \frac{D_{\text{fact}}}{D_{\text{ost}}} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}, \quad (3.1)$$

где

$D_{\text{fact}}$  – факторная сумма квадратов на одну степень свободы,

$D_{\text{ost}}$  – сумма квадратов остатков,

$R^2$  – коэффициент множественной детерминации,

$m$  – число параметров при переменных  $X$  (В случае линейной регрессии совпадает с числом включенных в модель предикторов),

$n$  – число наблюдений.

При построении модели множественной регрессии качество модели оценивается при добавлении каждого нового предиктора. Добавленный предиктор должен увеличивать долю описываемой регрессионным уравнением дисперсии.

Кроме этого, предикторы могут вводиться в модель в разной последовательности и от этой последовательности может зависеть значимость предиктора. Этот эффект связан с возможной корреляцией предикторов.

Мерой оценки важности учёта предиктора служит частный критерий Фишера

$$f_x = \frac{R_{full}^2}{1-R_c^2} \cdot \frac{n-m-1}{1}, \quad (3.2)$$

где

$R_{full}^2$  – коэффициент множественной регрессии для модели с полным набором предикторов,

$R_c^2$  – коэффициент множественной регрессии для модели без включения фактора.

В результате отбора предикторов получается таблица, пример которой приведён на рисунке 3.9.

	GH000	GH010	GH050	GH025	GH050	GH070	GH000	GH4V1	T300	T700	T900	T825	T900
2	81.2870961524	50.962018046	18.3252039524	8.7249485544	0.47269325141	4.57750299298	4.27684483821	60.3171936796	88.4391052295	100.76185968	158.980993015	234.458418489	300
3	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	False	False	False	False	False	True
10	False	False	False	False	False	False	False	False	False	False	False	False	True
11	False	False	False	False	False	False	False	False	False	False	False	False	True
12	False	False	False	False	False	False	False	False	False	False	False	False	True
13	False	False	False	False	False	False	False	False	False	False	False	False	True
14	False	False	False	False	False	False	False	False	False	False	False	False	True
15	False	False	False	False	False	False	False	False	False	False	False	False	True
16	False	False	False	False	False	False	False	False	False	False	False	False	True
17	False	False	False	False	False	False	False	False	False	False	False	False	True
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													

Рисунок 3.9 – Пример итогового файла с результатами использования ANOVA модели для выбора лучших предикторов

В первой строке итоговой таблицы (рисунок 3.10) указаны потенциальные предикторы. Во второй строке выведено значение критерия Фишера. Далее в строках со 2 по 17 указывается результат *True* или *False*, которые означают: *True* - данный параметр может использоваться в данном наборе как предиктор; *False* - данный параметр не может использовать в данном наборе как предиктор.

#### *Встроенный модуль RFE (библиотека SKLearn)*

Модуль RFE также входит в состав библиотеки SKLearn и может быть импортирован в программный код следующей командой

```
fromsklearn.feature_selection import RFE
```

В данном модуле реализуется методика построения регрессионной модели на основе метода рекурсивного исключения. Оценка качества регрессионной модели определяется по отношению к первому набору предикторов. Для оценки качества используется либо модуль *fit.coef\_* либо модуль *fit.feature\_importances\_*.

Наименее важные предикторы исключаются из текущего набора предикторов. Эта процедура рекурсивно повторяется до тех пор, пока в конечном итоге не будет достигнуто желаемое качество моделирования или желаемое количество предикторов.

Таким образом, чтобы построить регрессионную модель надо сначала определить массивы, которые будут участвовать в анализе командой *fit = result.fit(X, Y)*, а затем получить оценку полученной регрессионной модели с помощью модуля *fit.get\_params*.

Программа создает целочисленный вектор в котором лучший предиктор определяется цифрой 1, а чем хуже качество величины, тем дальше число от 1. На рисунке 3.10 представлен пример выходного файла с результатами использования RFE модели для выбора лучших предикторов. Оценка возможности использования величины в качестве предиктора помещена в колонку G (седьмая колонка). Аналогичным образом составлялась итоговая таблица с наборами лучших предикторов.

Рисунок 3.10 – Пример итогового файла с результатами использования RFE модели для выбора лучших предикторов

### 3.2.2 Оценка потенциальных предикторов корреляционным анализом

Данный метод включает определение коэффициентов корреляции между предиктантом (в нашем исследовании температурой воздуха на уровне 2м по данным синоптических сообщений) и потенциальными предикторами (результатами моделирования с использованием WRF). Решение о включении предиктора в регрессионную модель принимает исследователь, который создаёт модель.

Для корреляционного анализа был создан отдельный модуль, который рассчитывает корреляцию между двумя массивами переменных и ошибку.

Коэффициенты корреляции между предиктантом и потенциальными предикторами представлены в таблице 3.1.

Таблица 3.1 – Результаты корреляционного анализа

Переменная	Коэффициент корреляции
Геопотенциальная высота 500	0.817±0.025
Геопотенциальная высота 700	0.744±0.034
Геопотенциальная высота 850	0.540±0.054
Геопотенциальная высота 925	0.375±0.067
Геопотенциальная высота 950	0.302±0.071
Геопотенциальная высота 975	0.176±0.076
Геопотенциальная высота 1000	0.072±0.079
Температура 500	0.819±0.025
Температура 700	0.847±0.021
Температура 850	0.877±0.017
Температура 925	0.907±0.013
Температура 950	0.923±0.011
Температура 975	0.933±0.010
Температура 1000	0.938±0.009
Относительная влажность 500	-0.241±0.077
Относительная влажность 700	-0.178±0.079
Относительная влажность 850	-0.138±0.079
Относительная влажность 925	-0.134±0.079
Относительная влажность 950	-0.224±0.077
Относительная влажность 975	-0.421±0.068
Относительная влажность 1000	-0.555±0.058
Компонента ветра U500	-0.030±0.080
Компонента ветра U 700	0.036±0.080
Компонента ветра U 850	-0.010±0.080
Компонента ветра U 925	-0.075±0.080
Компонента ветра U 950	-0.098±0.080
Компонента ветра U 975	-0.091±0.080
Компонента ветра U 1000	-0.089±0.080
Компонента ветра V 500	0.130±0.078
Компонента ветра V 700	0.181±0.076
Компонента ветра V 850	0.174±0.076
Компонента ветра V 925	0.140±0.077
Компонента ветра V 950	0.137±0.078
Компонента ветра V 975	0.175±0.076
Компонента ветра V 1000	0.205±0.075
Осадки	-0.014±0.080
Компонента ветра V на уровне 10 м	0.196±0.076
Компонента ветра U на уровне 10 м	-0.081±0.080
Геопотенциальная высота на уровне 1 м	0.000±0.080
Относительная влажность на уровне 2 м	-0.554±0.058
Температура на уровне 1 м	0.937±0.009
Температура на уровне 2 м	0.943±0.008

Помимо вышеперечисленных переменных, так же была рассчитана корреляция между температурой воздуха из синоптических сообщений и дополнительно рассчитанными величинами. Эти коэффициенты корреляции приведены в таблице 3.2.

Таблица 3.2 – Результаты корреляционного анализа

Переменная	Коэффициент корреляции
Геопотенциальная высота слоя 850-1000	0.918±0.012
Геопотенциальная высота слоя 925-1000	0.841±0.022
Геопотенциальная высота слоя 950-1000	0.831±0.023
Геопотенциальная высота в слое 975-1000	0.907±0.013
Геопотенциальная высота в слое 500-700	0.829±0.023
Геопотенциальная высота в слое 500-1000	0.878±0.017
Геопотенциальная высота в слое 700-1000	0.897±0.015
Точка росы 700	0.847±0.021
Точка росы 850	0.877±0.017

### 3.2.3 Выбор статистической модели

Проанализировав полученные результаты выбора наборов наилучших предикторов, была определено использование модели ANOVA. Этот вывод связан с тем, что корреляционный анализ требует экспертной оценки исследователя, чем снижается объективность алгоритма. Рекурсивный метод не удовлетворяет качеством оценки ошибок регрессионной модели. Модель ANOVA была выбрана, т. к. критерий Фишера показывали наилучшие результаты для регрессионной модели, которая построена по этому алгоритму. Коэффициент детерминации также оказался выше, по сравнению с наборами лучших потенциальных предикторов других моделей; СКО оказалась ниже. Проанализировав результаты, отобрали все потенциальные предикторы для каждого набора в определённом порядке, которые представлены таблице 3.3.

Таблица 3.3 – Предикторы, отобранные для использования в модели множественной регрессии

Коэффициент детерминации	Модельная ошибка	Количество предикторов	Предикторы
0.950	4.753	1	TLVL2
0.952	4.643	2	TLVL2, DP1000
0.952	4.637	3	TLVL2, DP1000, T1000
0.952	4.600	4	TLVL2, DP1000, T1000, T975
0.952	4.589	5	TLVL2, DP1000, T1000, T975, DP975
0.952	4.588	6	TLVL2, DP1000, T1000, T975, DP975, TLVL1
0.952	4.587	7	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950
0.952	4.586	8	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950
0.952	4.586	9	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950, GH8501000
0.952	4.561	10	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950, GH8501000, GH9751000
0.952	4.561	11	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950, GH8501000, GH9751000, GH850975
0.953	4.553	12	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950, GH8501000, GH9751000, GH850975, T925
0.953	4.552	13	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950, GH8501000, GH9751000, GH850975, T925, DP925
0.953	4.519	14	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950, GH8501000, GH9751000, GH850975, T925, DP925, GH7001000
0.953	4.519	15	TLVL2, DP1000, T1000, T975, DP975, TLVL1, T950, DP950, GH8501000, GH9751000, GH850975, T925, DP925, GH7001000, GH700975

Для каждого набора предикторов из таблицы 3.3 создана регрессионная модель. Например, для набора под номером 3, регрессионное уравнение будет выглядеть следующим образом:

$$\widetilde{T}_2 = k_0 + k_1 * TLVL2 + k_2 * DP1000 + k_3 * T1000, \quad (3.3)$$

где

$\widetilde{T}_2$  – прогнозируемая температура на уровне 2 метра над землей,

$k_0, k_1, k_2$  и  $k_3$  – коэффициенты регрессионного уравнения,

TLVL2 – температура воздуха на уровне 2 метра над землей по результатам гидродинамического моделирования WRF в К,

DP1000 – температура точки росы на уровне 1000 гПа по расчетам, используя результаты гидродинамического моделирования WRF в К,

T1000 – температура на уровне 1000 гПа по результатам гидродинамического моделирования WRF в К.

Чтобы оценить какой набор потенциальных предикторов все же является лучшим, так же были рассчитаны коэффициенты детерминации и СКО всех регрессионных уравнений с каждым из наборов потенциальных предикторов.

Чтобы определить какой из наборов потенциальных предикторов лучше, было использовано основное ограничение: количество предикторов должно быть в количестве 5-7 штук. Такое число предикторов было выбрано после анализа научной литературы, посвященной статистическому моделированию. Кроме этого, проведенные исследования показали, что качество регрессионной модели не улучшается при увеличении количества предикторов.

Чтобы рассчитать коэффициент детерминации и СКО для каждого набора предикторов, использовалась так же библиотека *sklearn* со встроенным модулем *linear\_model*.

### 3.3 Расчет ошибок регрессионных моделей

После использования модели дисперсионного анализа ANOVA, был выбран набор наилучших потенциальных предикторов под номер 5. В этот набор входят следующие метеорологические величины:

- 1) модельная температура воздуха на уровне 2 м,
- 2) модельная температура воздуха на уровне 1000 гПа,
- 3) рассчитанная по модельным данным температура точки росы на уровне 1000 гПа,
- 4) модельная температура воздуха на уровне 975 гПа,
- 5) рассчитанная по модельным данным температура точки росы на уровне 975 гПа.

На следующем шаге находились коэффициенты уравнения множественной регрессии. Были составлены модели для всех станций. Приведем для примера регрессионную модель для станции Братислава.

$$\tilde{T}_2 = 19,9 + b \cdot X^T, \quad (3.4)$$

где  $b$  – вектор коэффициентов,

$$b = \begin{pmatrix} 0,774 \\ -1919 \\ 1916 \\ 1766 \\ -1762 \end{pmatrix},$$

$X$  – вектор предикторов,

$$X = \begin{pmatrix} T_2 \\ T_{1000} \\ T_{d1000} \\ T_{975} \\ T_{d975} \end{pmatrix},$$

С помощью этого же программного кода были найдены коэффициенты регрессионного уравнения, коэффициент детерминации и СКО и для других станций Словакии. Данные представлены в Приложении Б в таблице Б.1.

Следующей задачей была оценка эффективности выбранных предикторов для каждой станции. Для этого оценивались ошибки прогноза температуры воздуха на уровне 2м по результатам моделирования и после корректировки этих результатов с помощью разработанной регрессионной модели. После совместного анализа ошибок был сделан вывод о том, что лучше использовать для прогнозирования температуры: результаты пост-процессинга с использованием техники MOS или прямые результаты моделирования температуры на уровне 2 м. Расчёты проводились для всех 18 станций.

В таблицах 3.4 и 3.5 представлены результаты расчётов ошибок моделирования для станции Братислава. Для более наглядного представления для каждой станции были построены графики изменения тех или иных ошибок во времени. Результаты для других станций приведены в Приложении В.

Таблица 3.4 – Ошибки моделирования после MOS для станции Братислава

Час	СКО	Систематическая ошибка (смещение)	Средний модуль отклонения	Максимальный модуль отклонения	Коэффициент корреляции
10	1.885	0.170	1.417	7.145	0.983
11	2.022	0.317	1.488	8.792	0.981
12	2.061	0.316	1.501	8.586	0.981
13	2.160	0.315	1.582	9.155	0.980
14	2.251	0.204	1.630	9.566	0.978
15	2.198	0.078	1.644	10.457	0.980
16	2.109	-0.063	1.606	7.826	0.981
17	2.146	-0.248	1.686	7.464	0.980
18	2.136	-0.351	1.659	7.552	0.979
19	2.070	-0.183	1.575	8.800	0.976
20	2.063	-0.014	1.541	11.468	0.973
21	2.065	0.169	1.574	10.591	0.972
22	2.052	0.202	1.571	11.730	0.970
23	2.142	0.191	1.675	10.673	0.966
0	2.151	0.114	1.687	9.837	0.966
1	2.113	0.113	1.677	9.613	0.965
2	2.078	0.019	1.644	8.998	0.965
3	2.158	-0.039	1.688	8.825	0.963
4	2.137	-0.131	1.655	9.458	0.964
5	2.117	-0.162	1.623	8.796	0.969
6	2.088	-0.276	1.603	9.065	0.974
7	2.009	-0.243	1.539	9.421	0.977
8	2.061	-0.192	1.556	10.333	0.978
9	2.234	-0.162	1.688	11.289	0.975
10	2.354	-0.047	1.732	12.220	0.973
11	2.489	0.102	1.831	12.869	0.971
12	2.491	0.066	1.797	13.325	0.972
13	2.636	0.043	1.930	13.746	0.969
14	2.612	0.014	1.924	13.774	0.970
15	2.607	-0.141	1.933	14.080	0.971
16	2.508	-0.343	1.931	10.627	0.973
17	2.454	-0.534	1.951	9.638	0.975
18	2.260	-0.570	1.823	9.040	0.977
19	2.144	-0.427	1.725	10.088	0.975
20	2.107	-0.255	1.677	9.469	0.973
21	2.062	-0.083	1.650	8.191	0.972
22	2.089	-0.058	1.654	9.596	0.969
23	2.232	-0.041	1.774	9.294	0.964

Таблица 3.5 – Ошибки моделирования для Братиславы. Результаты WRF

Час	СКО	Систематическая ошибка (смещение)	Средний модуль отклонения	Максимальный модуль отклонения	Коэффициент корреляции
10	2.040	-0.646	1.573	8.803	0.983
11	2.104	-0.463	1.574	11.171	0.981
12	2.153	-0.481	1.620	10.994	0.981
13	2.233	-0.506	1.689	12.036	0.979
14	2.376	-0.691	1.772	12.377	0.978
15	2.451	-0.943	1.895	12.001	0.979
16	2.491	-1.206	1.979	10.064	0.980
17	2.697	-1.560	2.196	10.051	0.979
18	2.849	-1.830	2.329	8.898	0.977
19	2.781	-1.753	2.282	10.793	0.973
20	2.701	-1.622	2.190	10.946	0.971
21	2.616	-1.480	2.063	10.627	0.970
22	2.580	-1.460	2.043	11.098	0.969
23	2.659	-1.497	2.103	9.390	0.965
0	2.727	-1.606	2.149	9.590	0.965
1	2.696	-1.619	2.143	8.023	0.964
2	2.726	-1.723	2.200	7.440	0.964
3	2.823	-1.800	2.284	7.462	0.962
4	2.837	-1.858	2.317	7.780	0.964
5	2.742	-1.733	2.224	8.205	0.968
6	2.667	-1.658	2.146	8.976	0.974
7	2.492	-1.429	1.960	10.246	0.977
8	2.434	-1.219	1.897	12.188	0.978
9	2.519	-1.092	1.958	13.095	0.976
10	2.551	-0.883	1.948	13.998	0.973
11	2.616	-0.697	1.964	14.640	0.971
12	2.633	-0.753	1.971	15.082	0.972
13	2.780	-0.795	2.107	15.454	0.969
14	2.789	-0.891	2.124	15.480	0.970
15	2.900	-1.175	2.242	15.807	0.972
16	2.934	-1.506	2.353	11.452	0.974
17	3.047	-1.865	2.518	10.641	0.974
18	3.019	-2.065	2.503	9.997	0.977
19	2.938	-2.004	2.447	11.618	0.974
20	2.845	-1.876	2.337	11.274	0.972
21	2.731	-1.737	2.182	9.811	0.972
22	2.745	-1.721	2.192	9.530	0.969
23	2.854	-1.726	2.268	9.447	0.963

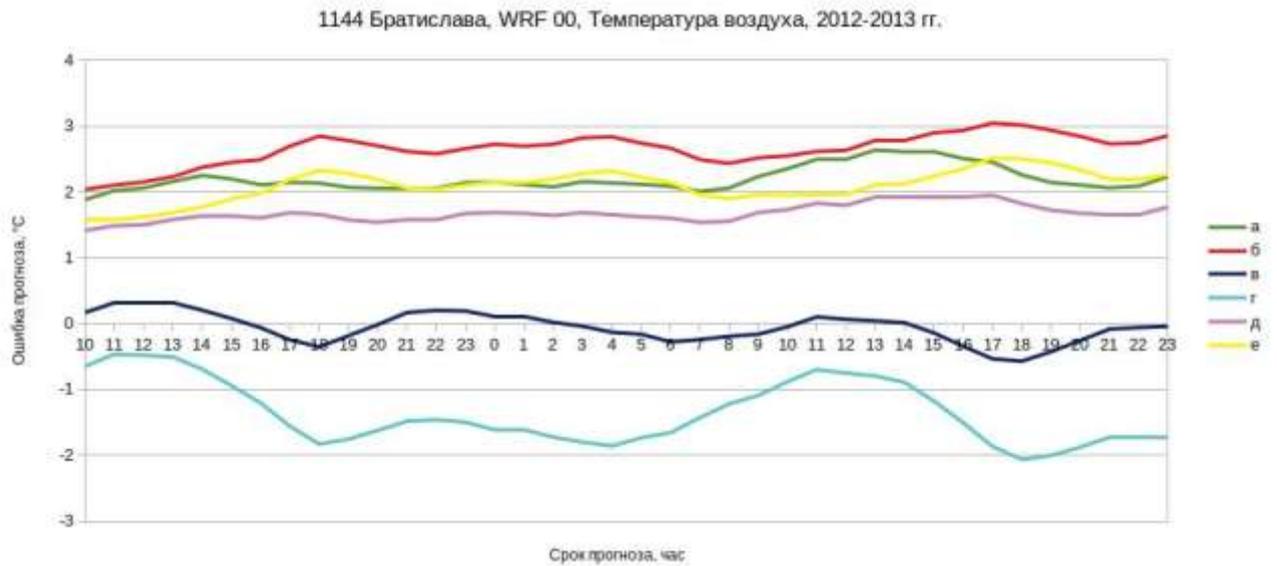


Рисунок 3.11 – Ошибки моделирования для станции Братислава.  
а - СКО MOS; б - СКО WRF ; в – систематическая ошибка MOS; г - систематическая ошибка WRF; д – CAO MOS; е - CAO WRF

## 4 Анализ результатов

### 4.1 Анализ полученных результатов

#### *Братислава*

Анализ результатов позволяет увидеть, что значения СКО MOS несколько ниже, СКО WRF. Значения систематической ошибки MOS колеблется в пределах  $\pm 0.4$ , в то время как по результатам WRF ошибки находится в пределах  $\pm 2.1$ . Средняя абсолютная ошибка так же ниже после обработки техникой MOS. Значения корреляции между полученной температурой после обработки результатов гидродинамического моделирования техникой MOS и температурой из синоптических сообщений выше 0.96, что свидетельствует об отличной связи временных рядов.

Проведя подобный анализ по другим станциям, результаты которых представлены в Приложении В, можно сделать вывод о том, что техника MOS улучшает результаты численного прогнозирования модели WRF. Улучшение прогноза связано с уменьшением ошибки прогнозирования на несколько градусов. Для наглядности представим наборы предикторов для каждой станции и высоту станции в метрах в таблице 4.1. Анализируя результаты, было замечено, что особенно улучшается прогноз для станций, которые находятся выше уровня моря на несколько сотен или тысяч метров, так как в наборе потенциальных предикторов преобладает использование переменных геопотенциальной высоты или разницы между некоторыми уровнями геопотенциальной высоты, а не температуры воздуха. В таблице 4.2 перечислены все предикторы, встречающиеся на всех 18 станциях, в порядке убывания частоты их применения.

Таблица 4.1 – Названия станций, их высота, количество и названия предикторов

Станция	Количество Предикторов	Предикторы	Н, м
Братислава	5	TLVL2, T1000, DP1000, T975, DP975	133
Дудинце	5	TLVL2, DP1000, T1000, TLVL1, T975	140
Жилина Хричов	6	TLVL2, T1000, DP1000, T975, DP975, TLVL1	310
Каменица на Цирохоу	7	TLVL2, T1000, DP1000, TLVL1, GH9751000, GH9501000, T975	177
Койшовская Хола	5	T925, DP925, T1000, DP1000, TLVL2	1242
Кошице	7	TLVL2, T1000, DP1000, T975, DP975, GH9501000, GH9751000	232
Ломницки-Штит	6	GH700925, GH700950, GH700975, GH700850, GH7001000, GH5001000	2633
Лученец	5	TLVL2, DP1000, T1000, T975, DP975	214
Льесек	6	GH9501000, GH9251000, TLVL2, GH950975, T950, DP950	692
Милхостов	5	TLVL2, T1000, DP1000, T975, DP975	104
Попрад-Гановце	5	GH9501000, GH9251000, TLVL2, T950, DP950	703
Попрад-Татры	4	GH9501000, TLVL2, GH9251000, DP950	694
Прьевидза	7	TLVL2, DP1000, T1000, DP975, T975, GH9501000, GH9751000	259
Пьештяни	6	TLVL2, DP1000, T1000, T975, DP975, TLVL1	163
Стропков Тисинец	7	TLVL2, T1000, DP1000, T975, DP975, GH9501000, GH9751000	216
Телгарт	4	TLVL2, GH9251000, GH9501000, GH8501000	901
Хопок	6	GH700975, T850, DP850, GH7001000, GH700950, T950	2008
Штрбске Плесо	6	TLVL2, GH8501000, GH9251000, GH7001000, T925, DP925	1355

Таблица 4.2 – Все используемые предикторы

Предиктор	Количество станций, где используется в качестве предиктора
Температура воздуха на уровне 2 м	16
Температура воздуха на уровне 1000 гПа	11
Температура точки росы на уровне 1000 гПа	11
Температура воздуха на уровне 975 гПа	10
Температура точки росы на уровне 975 Гпа	8
Разница геопотенциальных высот между уровнями 950 и 1000 гПа	8
Разница геопотенциальных высот между уровнями 925 и 1000 гПа	5
Разница геопотенциальных высот между уровнями 975 и 1000 Гпа	5
Температура воздуха на уровне 1 м	4
Температура воздуха на уровне 950 гПа	3
Температура точки росы на уровне 950 гПа	3
Разница геопотенциальных высот между уровнями 700 и 1000 гПа	3
Температура воздуха на уровне 925 гПа	2
Температура точки росы на уровне 925 гПа	2
Разница геопотенциальных высот между уровнями 700 и 950 гПа	2
Разница геопотенциальных высот между уровнями 700 и 975 гПа	2
Разница геопотенциальных высот между уровнями 850 и 1000 гПа	2
Температура воздуха на уровне 850 гПа	1
Температура точки росы на уровне 850 гПа	1
Разница геопотенциальных высот между уровнями 500 и 1000 гПа	1
Разница геопотенциальных высот между уровнями 700 и 850 гПа	1
Разница геопотенциальных высот между уровнями 700 и 925 гПа	1
Разница геопотенциальных высот между уровнями 950 и 975 гПа	1

На рисунке 4.1 представлен график частоты использования рассмотренных метеорологических величин в качестве предикторов. Анализируя график, можно сделать вывод о том, что практически во всех регрессионных моделях участвует прогностическая температура на 2 метрах (89 %). Так же довольно часто используются характеристики температуры и влажности на самых нижних изобарических поверхностях: 1000 и 975 гПа (61 %). Реже всего для уточнения прогноза используются характеристики поверхности 925 и 850 гПа (11 – 6%).

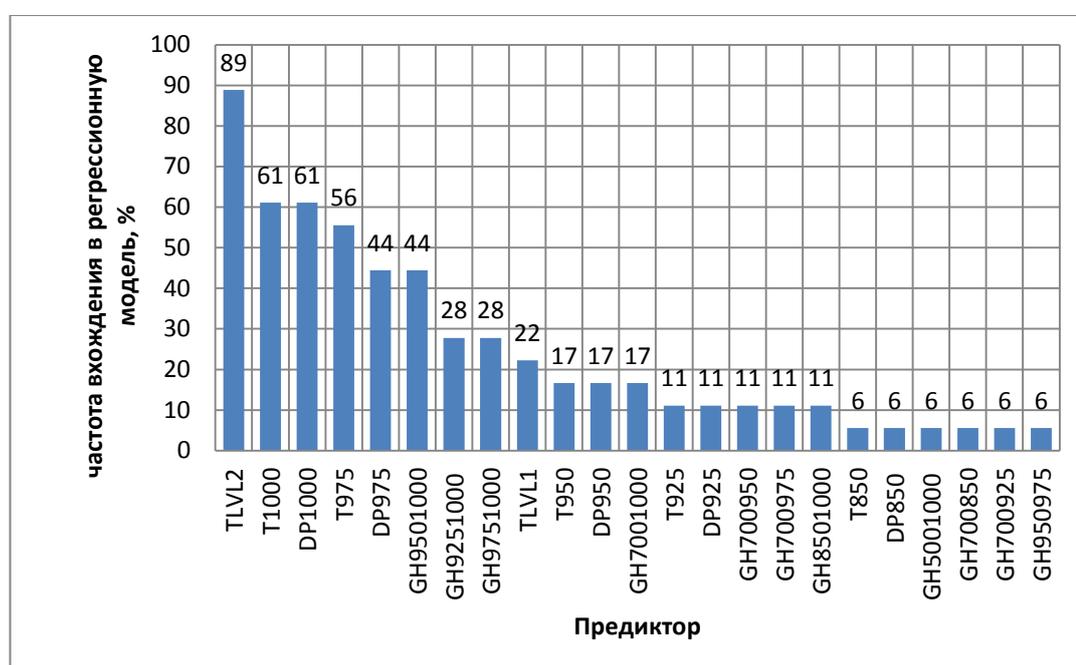


Рисунок 4.1 – Частота использования переменной в качестве предиктора

Практически для всех станций сохраняется один и тот же набор предикторов. Исключение составляют станции, чья высота над уровнем моря превышает 1000 м: Койшовская Хола, Ломницки-Штит, Хопок, Штребске Плесо.

Можно выделить станцию Хопок. Применение техники пост-процессинга MOS позволило уменьшить ошибки прогноза с 5 °С до 1.5 °С. На станции Жилина Хричов применение техники MOS так же позволило

улучшить качество моделирование, но не значительно с 2.9 °C до 2.4 °C. Эта ситуация требует дополнительных исследований.

#### 4.2 Применение

Результаты исследованию будут применены в оперативной практике компании MicroStep-MIS (Словакия, г.Братислава). Результаты прогноза численной гидродинамической модели WRF с уточнениями по результатам данного исследования будут приводиться на сайте *azet.sk*.

Как было указано ранее, одним из недостатков MOS является ее зависимость от конфигурации модели. Поэтому в данном исследовании разработан алгоритм пост-процессинга, который будет использоваться в дальнейшем при любых модельных модификациях.

## ЗАКЛЮЧЕНИЕ

В данной выпускной квалификационной работе магистра реализована методика уточнения результатов гидродинамического прогноза с использованием концепции Model Output Statistics (MOS), то есть цель, поставленная перед началом исследования, достигнута.

В начале исследований были поставлены несколько задач, все из которых были успешно выполнены. А именно:

1) Разработана система обработки синоптических сообщений с метеорологических станций с помощью программных кодов на высокоуровневом языке программирования Python.

2) Разработана система обработки результатов прогнозирования модели WRF с помощью программных кодов на высокоуровневом языке программирования Python.

3) Изучена научная литература по технике пост-процессинга MOS.

4) Разработан программный код с применением специальных модулей языка программирования Python для расчета всех необходимых параметров техники MOS.

5) Произведено исследование возможности улучшения результатов прогноза WRF для температуры воздуха на уровне 2 метра для 18 метеорологических станций, используя технику MOS.

6) Произведена верификация результатов применения техники MOS с результатами прогнозирования численной модели WRF.

7) Разработана технология внедрения техники MOS в оперативную практику.

В результате проделанной работы можно сделать выводы о том, что техника Model Output Statistics улучшает результаты численного гидродинамического прогноза WRF. Таким образом, ее смело можно

использовать в качестве схемы пост-процессинга для оперативного прогнозирования.

Для большинства станций главным предиктором является прогностическая модельная температура воздуха на уровне 2 метров. Исключения составляют станции, высота которых над уровнем моря превышает 1000 метров.

В среднем качество прогноза улучшилось на 2 °С. Практически на всех исследуемых станциях после пост-процессинга прогноз относится к категории успешных, согласно синоптическим наставлениям.

Разработанный алгоритм может быть использован в любых оперативных подразделениях, занимающихся прогнозами погоды на основе гидродинамического моделирования.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. В.А. Тищенко, В.М. Хан, М.А. Толстых, Е.Н. Круглова, И.А. Куликова, А.Н. Гельфан Применение статистической коррекции месячных и сезонных детерминистских прогнозов температуры воздуха и осадков по модели плав для отдельных районов России // Труды Гидрометеорологического научно-исследовательского центра Российской Федерации. –2015. – №358. –С. 212-132.
2. Armstrong, J. Scott Illusions in Regression Analysis // International Journal of Forecasting (forthcoming). –2012. –v. 28. № 3. –p. 689–694.
3. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani An Introduction to Statistical Learning // Springer New York. –2017. –440 p.
4. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных [Электронный ресурс] // Режим доступа: [www.machinelearning.ru/wiki/index.php?title=Критерий\\_Акаике](http://www.machinelearning.ru/wiki/index.php?title=Критерий_Акаике), свободный (23.05.2018).
5. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных [Электронный ресурс] // Режим доступа: [www.machinelearning.ru/wiki/index.php?title=ВКС](http://www.machinelearning.ru/wiki/index.php?title=ВКС), свободный (23.05.2018).
6. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных [Электронный ресурс] // Режим доступа: [www.machinelearning.ru/wiki/index.php?title= Коэффициент\\_детерминации](http://www.machinelearning.ru/wiki/index.php?title=Коэффициент_детерминации), свободный (23.05.2018).
7. Harry R. Glahn, Dale A. Lowry The use of MOS in objective weather forecast // Journal of applied meteorology. –1972. –v.11. –p. 1203–1211.

8. Model Output Statistics (MOS) - Objective Interpretation of NWP Model Output [Электронный ресурс] // Режим доступа: <https://www.weather.gov/media/mdl/Antolik2012.pdf>, свободный (23.05.2018)
9. Jordan G. Powers, Joseph B. Klemp, William C. Skamarock The weather research and forecasting model // Overview, System Efforts, and Future Directions. –2017. –p. 1717–1737.
10. Воробьева О.В. Статистическая обработка результатов моделирования для улучшения качества гидрометеорологического прогноза // Современные тенденции и перспективы развития гидрометеорологии в России. –2018. – С. 104-106.
11. Python for beginners [Электронный ресурс] // Режим доступа: <http://www.pythonforbeginners.com/learn-python/what-is-python/>, свободный (23.05.2018)