**РЕФЕРАТ**

**ВВЕДЕНИЕ**

Одна из задач совершенствования оперативной системы Черноморского Центра Морских Прогнозов связана c созданием комплексных систем морских прогнозов, основанных на одновременной, совместной работе различных прогностических систем. К числу таких систем относится система прогноза циркуляции и волнения Черного моря.

Совместный прогноз волнения и циркуляции позволит повысить точность расчетов как скоростей морских течений, температуры и солености в верхнем слое моря, так и определение характеристик морского волнения. Это достигается за счет более корректного учета эффектов взаимодействия морских волн и течений при использовании комплексной прогностической модели этих явлений.

**Актуальность**. Прежде чем, создавать комплексную систему необходимо оценить точность расчетов морского волнения при работе в автономном режиме. Оценка точности прогнозов волнения позволяет выяснить насколько созданная система удовлетворяет требования в точности предъявляемой со стороны специалистов, занимающихся изучением и моделированием состояния океана и атмосферы, а также потребителей-практиков, использующих эти результаты для решения прикладных задач. Кроме того, при разработке комплексной системы также потребуется решать ту же задачу.

**Цель данной работы** – Разработка системы, автоматизированной валидации результатов оперативных прогнозов морского волнения. В ходе создания данной системы были решены следующие задачи:

* Изучена организация хранения данных в Черноморском центре морских прогнозов ФГБУН МГИ
* Изучены литературные источники, в которых приведены сведения о математической модели SWAN используемой для расчета волнения, приведена информация о методах дистанционного определения высоты волн по данным альтиметрических измерений.
* Разработаны алгоритмы оценки точности прогноза волнения
* Разработан модуль оперативного загрузки и распаковки данных получаемые с серверов AVISO и ФГБУН МГИ
* Разработан модуль формирования совместной выборки данных результатов прогнозов и дистанционных измерений высоты морских волн
* Разработан модуль оценки точности расчетов

**Предмет и объект исследования.** Предметом исследования является система прогноза морского волнения. Объектом исследования является автоматизированная система оценки точности прогноза морского волнения с использования спутниковых альтиметрических измерений

**Научная новизна.**

**Практическое значение работы.** Введена в состав системы морских прогнозов ЧЦМП автоматизированная система оценки прогноза морского волнения по модели SWAN для дальнейшего совместного использования с моделью циркуляции Черного моря и улучшения достоверности прогнозов скоростей течений и волнения на этой основе.

**Структура работы.**

1. АНАЛИЗ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ В ЗАДАЧЕ
   1. Стандарты описания процессов в рамках прикладной предметной области
      1. IDEF0-диаграмма основного процесса

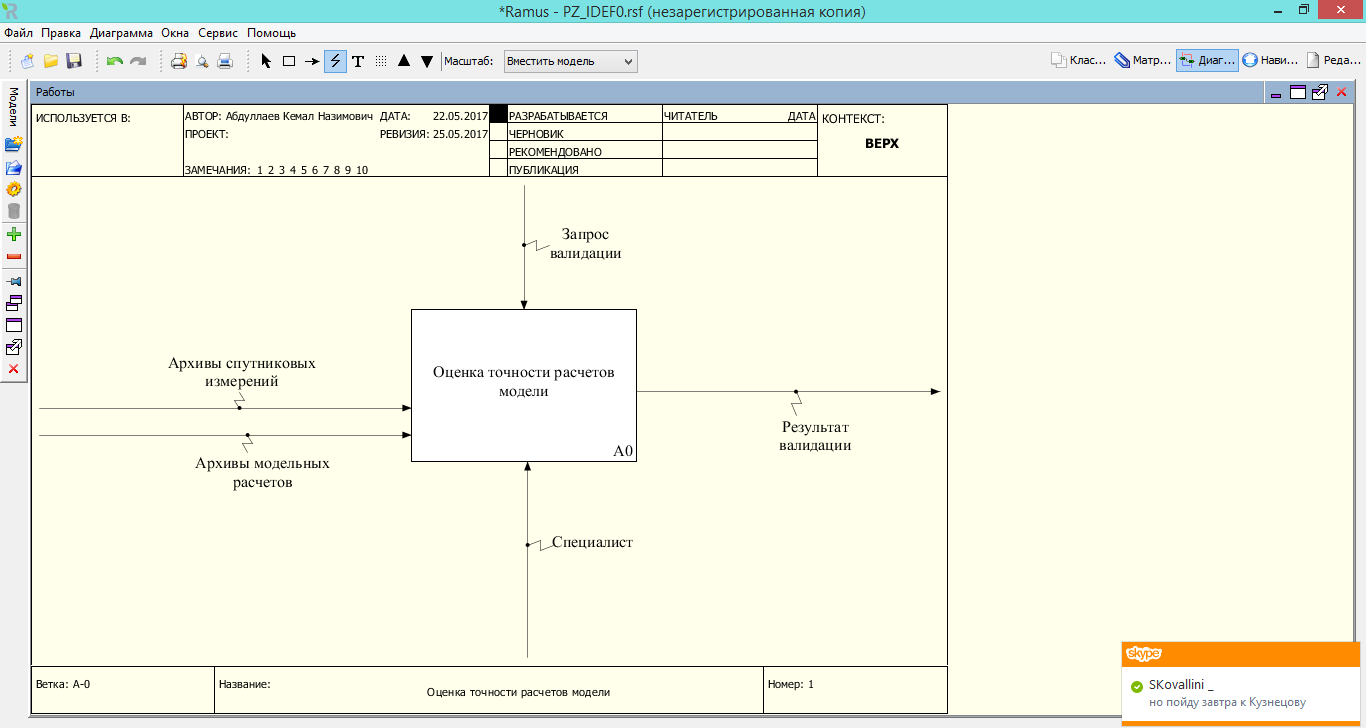


Рисунок 1 – Диаграмма IDEF0 основного процесса

Описание процессов на основной диаграмме представлено в таблице 1.

Таблица 1Описание процесса на контекстной диаграмме

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Шифр** | **Название процесса** | **Входные данные** | **Управляющие данные** | **Механизм** | **Результат процесса (работы)** |
| А-0 | Оценка точности расчетов модели | Архивы спутниковых измерений, архивы модельных расчетов, | Запрос валидации | Специалист | Результат валидации |

1.1.2 IDEF0-диаграмма декомпозиции основного процесса

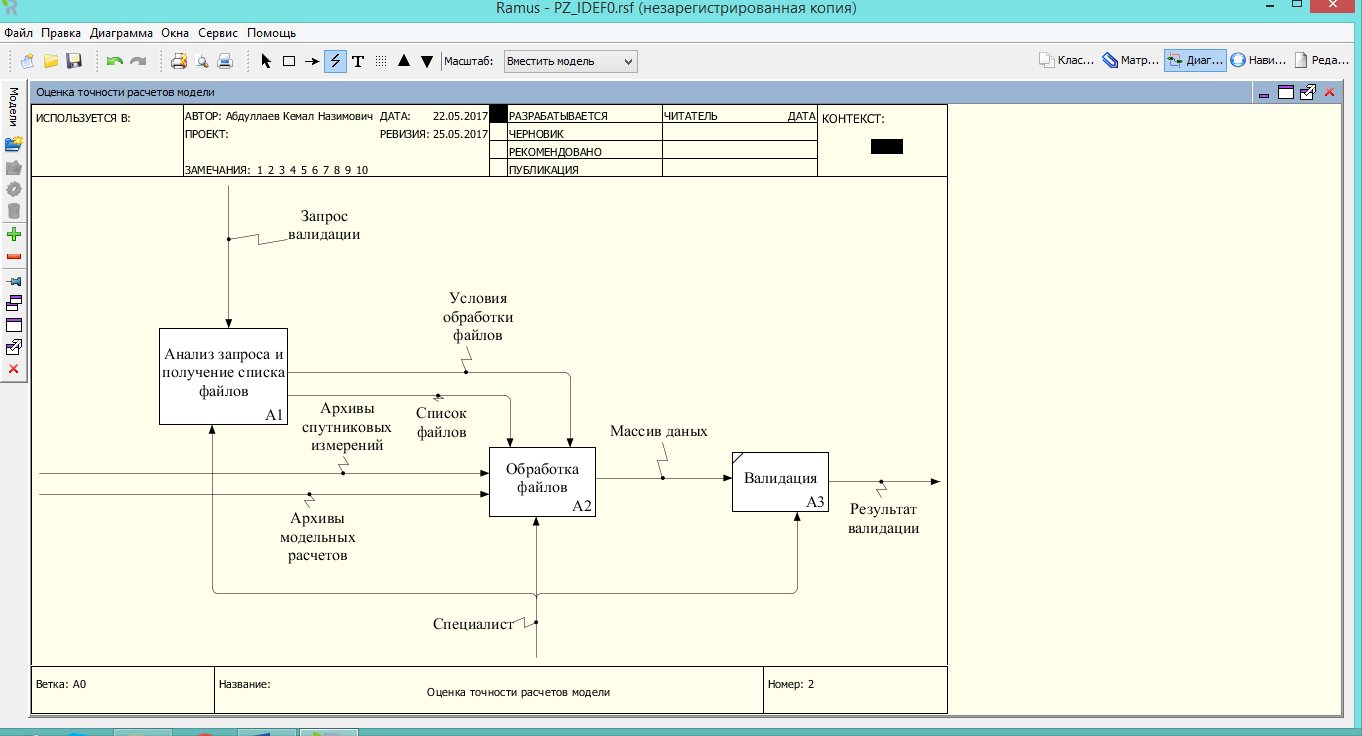


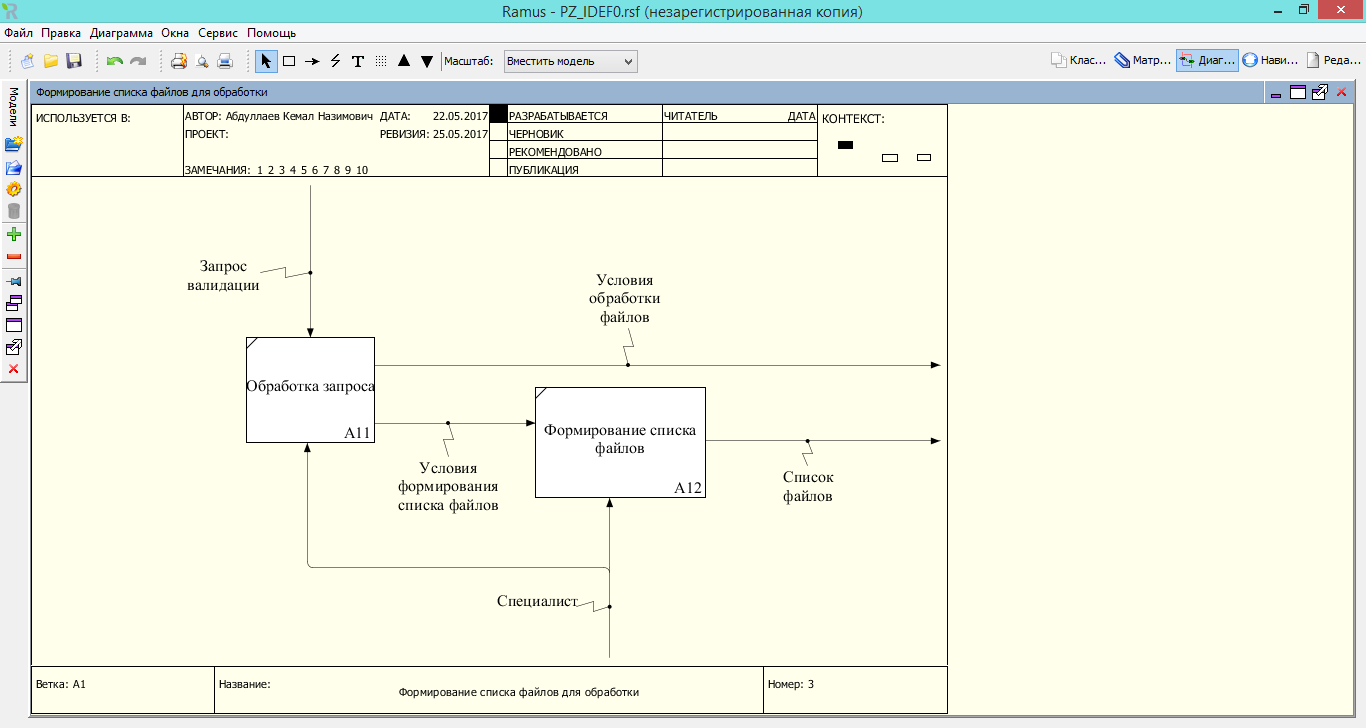
Рисунок 2 – Декомпозиция основного процесса

Описание процессов на диаграмме декомпозиции представлено в таблице 2.

Таблица 2Описание декомпозиции основного процесса

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Шифр** | **Название процесса** | **Входные данные** | **Управляющие данные** | **Механизм** | **Результат процесса (работы)** |
| А0 | Формирование списка файлов для обработки |  | Запрос | Специалист | Список файлов |
| Обработка данных | Архивы спутниковых измерений, архивы модельных расчетов | Условия обработки файлов, список файлов | Массив данных |
| Валидация | Обработанный массив данных |  | Результат валидации |

* + 1. IDEF0-диаграмма декомпозиции процесса формирования списка файлов для обработки



1.1. DFD-диаграмма основного процесса

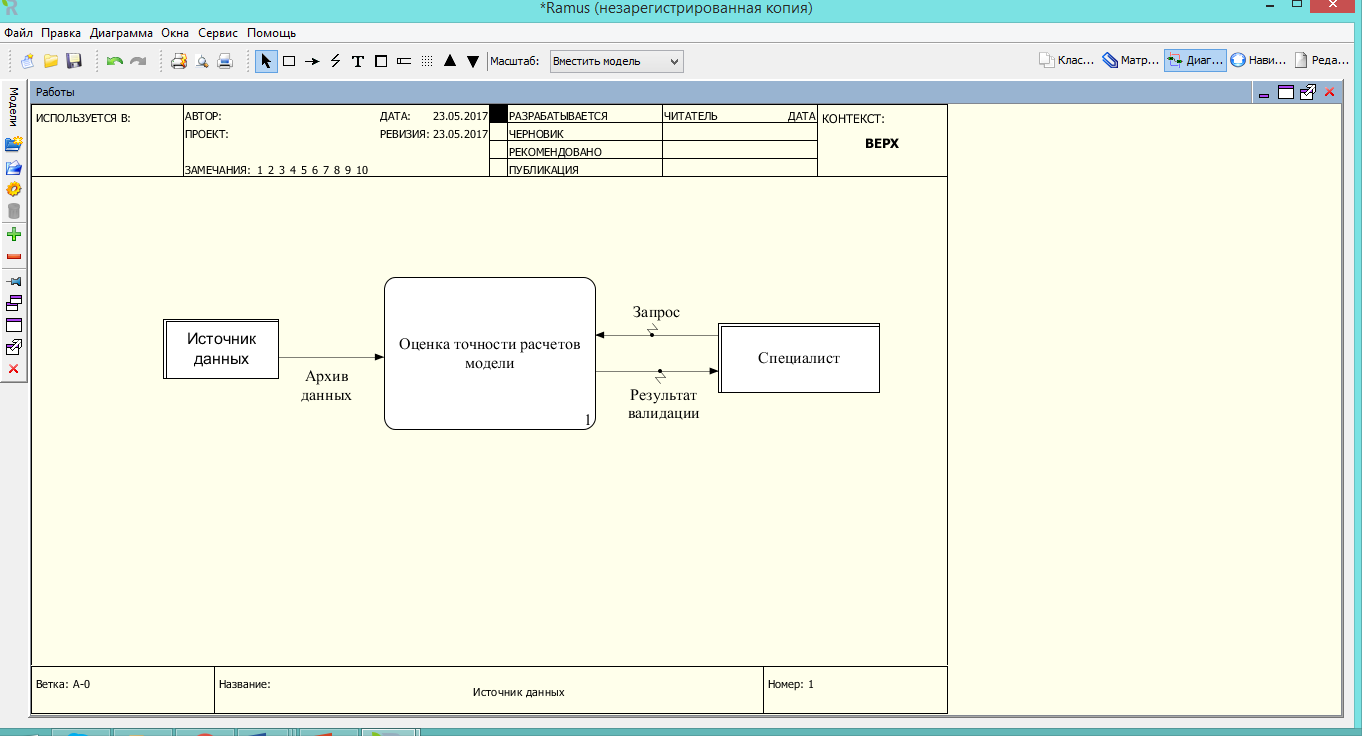


Рисунок 3 – DFD диаграмма основного процесса

Описание основного процесса представлено в таблице 3

Таблица 3Описание основного процесса

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Внешняя сущность** | **Событие(описание взаимодействия** | **Тип события** | **Основной процесс** | **Реакция системы на события** |
| Специалист | Запрос | типичный | Оценка точности расчетов модели | Результат валидации |
| Источник данных | Архив данных |

1.1.2.2 DFD-диаграмма декомпозиция основго процесса

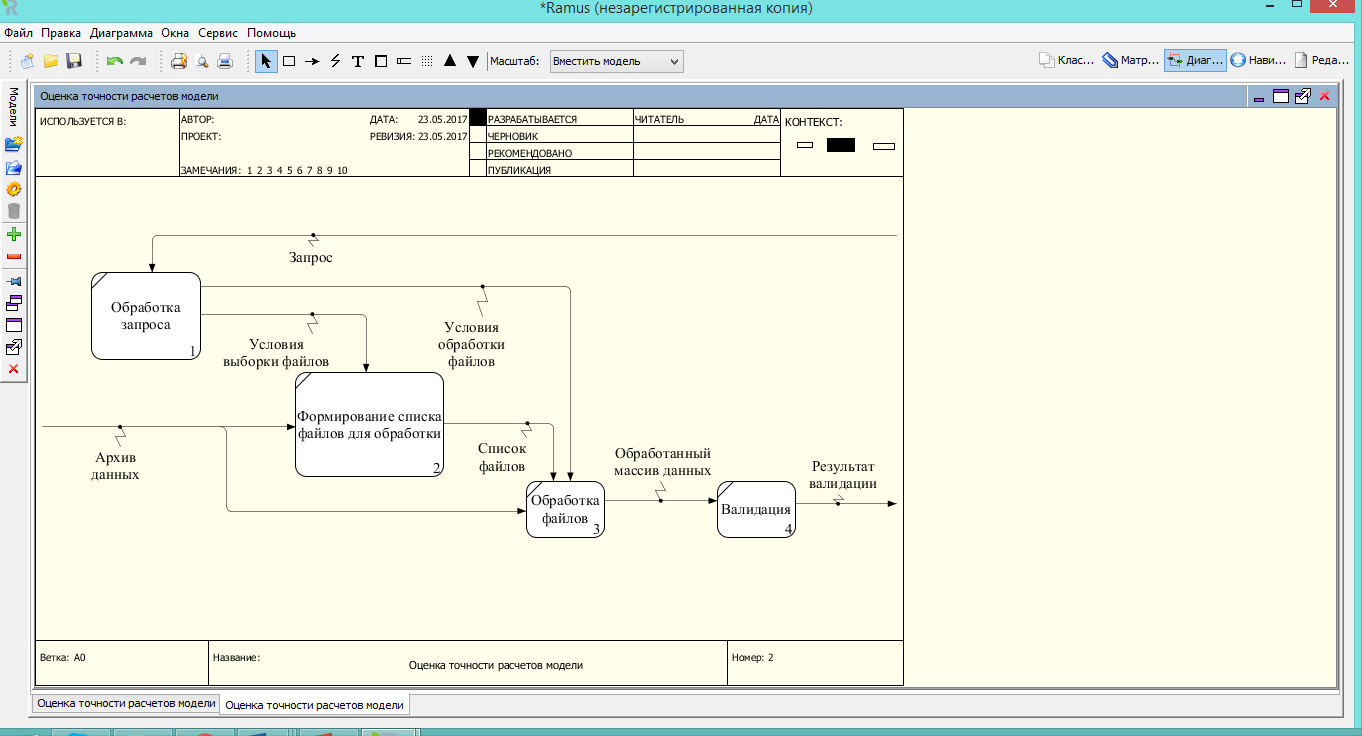


Рисунок 4 – DFD-диаграмма декомпозиции основного процесса

Таблица 4Описание декомпозиции основного процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Событие(описание взаимодействия** | **Тип события** | **Основной процесс** | **Реакция системы на события** |
| Запрос | типичный | Обработка запроса | Результат валидации |
| Условия выборки файлов | Формирование списка файлов для обработки |
| Условия обработки файлов | Обработка файлов |
| Обработанный массив данных | Валидация |

* 1. Обзор существующих подходов к решению задачи
  2. Выбор средств и методов решения задачи на основе вариантного анализа

Для выполнения поставленной задачи был рассмотрен ряд программных средств. В качестве языка программирования был выбран Python, по следующим причинам:

- Удобный синтаксис, простота в освоении, наличие гибких и мощных языковых конструкций, ускоряющих разработку ПО;

- Относительная простота кода что упрощает дальнейшее сопровождение ПО.

- Большое количество библиотек, которые включают работу с файлами научного формата.

В качестве среды разработки (IDE) был выбран программный пакет PyCharm от компании JetBrains. По сравнению с такими аналогами, как IDE Spyder, Komodo, PyDev (Eclipse), PyCharm содержит более удобный, легко настраиваемый и минималистичный интерфейс, возможность интеграции Git-репозитория, встроенный отладчик и синтаксический помощник — IntelliSence.

В качестве версии Python был выбран стандартный языковой пакет версии 3.6.0, загружаемый с официального ресурса [1]. Преимуществом версии языка версии 3x по сравнению с 2x является наличие более гибких языковых конструкций и всякого рода «синтаксического сахара». Минусом языка Python версии 3x является более низкая скорость работы.

* 1. Постановка задачи

Цель задачи – создать систему автоматической оценки точности расчетов данных по модели SWAN. Требуется по запросам произвести выборку данных и произвести оценку точности расчетов.

Для оценки точности необходимо использовать данные спутниковых альтиметрических измерений и прогнозов модели высоты волн. Эти данные записаны в файлы формата NetCDF и сжаты с помощью Zip архиватора. Архивы необходимо скачивать с FTP-серверов AVISO и МГИ РАН, распаковывать и производить выборку согласно требованиям МГИ РАН. После выбранные данные альтиметрических измерений и модельных прогнозов необходимо объединить в массив. По объединенному массиву необходимо оценить точность в рамках методологии, действующей в МГИ. Визуализировать результаты.

1. СИСТЕМОТЕХНИЧЕСКИЙ АНАЛИЗ ИНФОРМАЦИОННОЙ СИСТЕМЫ
   1. Архитектура ИС

Архитектура ИС – концепция, определяющая модель, структуру, выполняемые функции и взаимосвязь компонентов информационной системы.

Проект имеет файл-серверную архитектуру, что обеспечивается интерфейсами каждого отдельного модуля. Данная архитектура трехслойной структурой: слой представления, слой бизнес-логики и слой управления данными. Достоинствами данной архитектуры являются низкая стоимость разработки, высокая скорость разработки и низкая стоимость обновления ПО.

Система состоит из множества, подсистем, которые разбиваются на еще более простые подсистемы. Данный метод позволяет уменьшить сложность, а также повысить гибкость и расширяемость системы.

* 1. Критерий оценки эффективности ИС

Для оценки качественной и количественной эффективности ИС предполагает использование таких критериев как достоверность функционирования ИС и эффективность системы.

Система использует два источника данных таких как реальные измерения показателей и расчетные данные. Расчетные данные основываются на реальных измерения, что свидетельствует о достоверности данных.

Также система выполняет возложенные на нее задачи согласно поставленным задачам и требованиям.

* 1. Проектирование ИС

Научные данные имеют огромный объем, что осложняет их обработку и может привести к потере информации. По этой причине были разработаны открытые форматы научных данных такие как CDF (Common Data Format), CERIF (Common European Research Information Format), CIF (Crystallographic Information File), HDF (Hierarchical Data Format), NetCDF (Network Common Data Format) и т. д.

CDF –формат данных для хранения скалярных и многомерных данных независимым от платформы и дисциплины способом. Пакет управления научными данными позволяет разработчикам приложений управлять этими массивами данных Прозрачный доступ к данным и метаданным через интерфейсы прикладного программирования (API). Встроенная поддержка сжатия данных, таких как gZip, RLE, Huffman.

CERIF – Модель ориентированная на данные позволяет представлять исследовательские организации, их деятельность и их результаты, а также высокую гибкость с формальными отношениями, обеспечивает качественное обслуживание, архивирование. Обеспечивает доступ и обмен исследовательской информацией и поддержки передачи знаний лицам, принимающим решения, для оценки исследований, руководителей исследований, стратегов, исследователей, редакторов и широкой общественности.

CIF – это общий, гибкий и легко расширяемый архив в свободном формате. Может быть отредактирован простым текстовым редактором. CIF – предназначен для электронной передачи кристаллографических данных между отдельными лабораториями, журналами и базами данных: он был принят Международным союзом кристаллографии в качестве рекомендуемого средства для этой цели. Файл состоит из имен данных и элементов данных, а также объекта цикла для повторяющихся элементов. Имена данных, построенные иерархически для формирования категорий данных, являются самоописательными в пределах 32-символьного предела. Сортированный список имен данных вместе с их точными определениями представляет собой словарь CIF. Основной словарь CIF представлен в полном объеме и охватывает основные и наиболее часто используемые элементы данных, относящиеся к анализу кристаллической структуры. Словарь также доступен в виде электронного файла, подходящего для компьютерных приложений CIF. Будущие расширения Словаря будут включать элементы данных, используемые в более специализированных областях кристаллографии.

Формат иерархических данных (HDF) - это набор форматов файлов (HDF4, HDF5), предназначенных для хранения и организации большого количества данных. Первоначально разработанный в Национальном центре суперкомпьютерных приложений, он поддерживается некоммерческой корпорацией HDF Group, целью которой является дальнейшее развитие технологий HDF5 и постоянная доступность данных, хранящихся в HDF. В соответствии с этой целью библиотеки HDF и связанные с ними инструменты доступны в рамках либеральной BSD-подобной лицензии для общего использования. HDF поддерживается многими коммерческими и некоммерческими программными платформами.

NetCDF – это набор программных библиотек, не зависящих от машинных форматов данных, которые поддерживают создание, доступ и совместное использование массивных научных данных. Домашняя страница проекта размещена в рамках программы Unidata в Университетской корпорации по исследованию атмосферы (UCAR). Они также являются основным источником программного обеспечения netCDF, разработки стандартов, обновлений и т. д. Формат является открытым стандартом. NetCDF Classic и 64-битный формат смещения являются международным стандартом Open Geospatial Consortium.

Так как научные данные используемые в приложении записаны в файлы формата NetCDF и хранятся на FTP-серверах, была выбрана файл-серверная архитектура. На рисунке ниже представлен общий вид структуры системы:

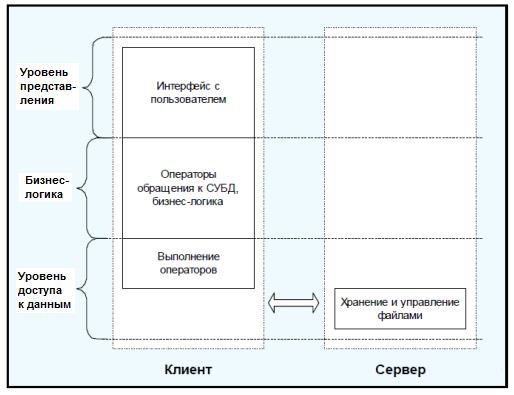


Рисунок 5 – Структура файл-серверной архитектуры

Достоинствами такой структуры являются:

* Низкая стоимость разработки
* Высокая скорость разработки
* Невысокая стоимость обновления и изменения ПО

Недостатками данной структуры являются:

* Низкая надежность системы
* Ограничение количества трафика
* Высокая нагрузка на сервер

Для повышения гибкости и расширяемости системы будет использоваться модульная структура и принцип декомпозиции.

Модульная структура позволит динамически подключать различные программные модули по мере необходимости, а принцип декомпозиции позволит разбить сложные задачи на более простые компоненты.

* 1. Принципы верификации и тестирования ИС

Верификация – это процесс для определить выполнение программными средствами и их компонентами поставленных задач и удовлетворение условий.

Цели верификации ПС:

* Определить функции информационной системы, которые должно выполнять ПС

Выводы по разделу 2

1. СТРУКТУРНЫЙ АНАЛИЗ ИС
   1. Разработка ПО

Альтиметрические данные используемые в приложение являются результатом измерений спутника Jason2, их последующего анализа и обработки. Полученные измерения записываются в файлы формата NetCDF и имеют расширение «.nc». На рисунке 3.1 изображена структура файла формата NetCDF.

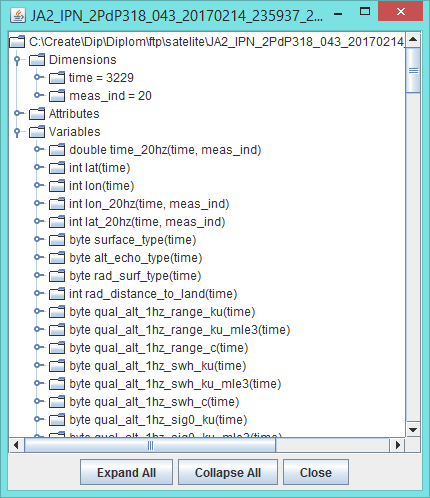


Рисунок 3.1 – Структура файла формата NetCDF

В имени файла NetCDF указывается тип спутника производившего измерения, семейство данных, тип данных, номер цикла, номер трека, дата начала измерений и дата окончания измерений.

Всего есть три семейства данных:

* ODGR (Operational Geophysical Data Record) – оперативные геофизические данные, поступают несколько раз в день с задержкой от 3 до 5 часов. Имеют наименьшую точность среди всех представленных.
* IGDR (Interim Geophysical Data Record) – промежуточные геофизические данные, поступают каждые день с задержкой от 1 до 2 дней.
* GDR (Geophysical Data Record) – геофизические данные, поступают каждые 10-35 (в зависимости от спутника) дней с задержкой в 2 месяца). Наиболее точные данные.

Также существует три типа данных указывающих на тип получения:

- N:Native

- R:Reduced

- S:Sensor

В приложение используются данные семейства IGDR и типа Native, так как имеют относительно высокую точности и поступают с небольшой задержкой. На рисунке 3.2 показаны примеры файлов формата NetCDF с альтиметрическими измерениями.

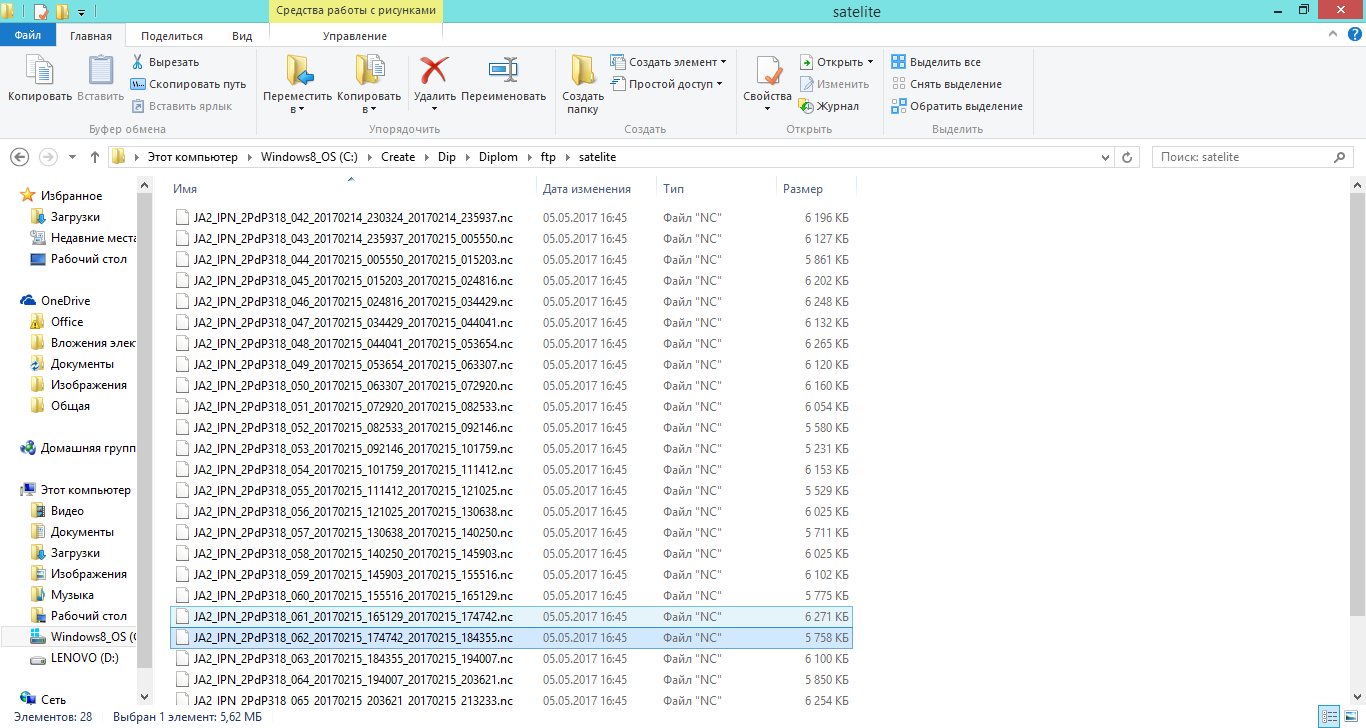


Рисунок 3.2 – Пример файла формата NetCDF

Данные модельных расчетов используемые приложением являются результатом прогноза ветрового волнения спектральной моделью SWAN (Simulating Waves Nearshore).

Базовая версия системы SWAN была разработана МГИ РАН в 2016. Причиной создания данной системы послужила сложность адаптации прогнозов, получаемых с помощью других систем, работающих в оперативном режиме, совместно с моделями, функционирующими в составе ЧЦМП. В ходе ее создания были поставлены и решены следующие задачи.

1. Введена в состав системы морских прогнозов ЧЦМП модель прогноза морского волнения SWAN для дальнейшего совместного использования с моделью циркуляции Черного моря и улучшения достоверности прогнозов скоростей течений и волнения на этой основе.
2. Создан пользовательский интерфейс для оперативного доступа сотрудников МГИ и представителей других заинтересованных научных организаций к данным прогноза волнения для использования при выполнении исследовательских работ.
3. Выполнена валидация модели SWAN, адаптированной для прогнозов волнения Черного моря в автономном квазиоперативном режиме работы за период 2012–2015 гг. и сравнение полученных результатов с аналогичными работами других авторов.

Модель SWAN основана на численном решении уравнения баланса волновой энергии в спектральной форме [3]:

1.  . (1)

Здесь *N = E/σ* – плотность волнового действия; *E* – энергетический спектр волнения; *x, y, t* – пространственные координаты и время; *σ, θ* – частотная и угловая координаты. Величины *cx, cy, cσ, cθ* – скорости переноса плотности волнового действия вдоль соответствующих пространственных и частотно-угловых координат [10]. Функция источника в правой части уравнения (1) имеет вид

*S = Sin + Snl + Swc+ Sbf + Sdib* (2)

где *Sin* – источник генерации волн ветром; *Snl*– нелинейные взаимодействия спектральных гармоник; *Swc* – диссипация энергии вследствие обрушения волновых гребней; *Sbf*– диссипация энергии, обусловленная донным трением; *Sdib* – обрушение волн на критических глубинах.

Применялась модель SWAN версии 42.10, использующая технологию OpenMP для многоядерных процессоров. Волновые характеристики определялись на сетке 238 x 132, покрывающей Черное море с шагом ~ 4,5 км. Рельеф дна задавался на той же сетке. Частотный интервал (0,05 Гц ≤ *σ* ≤1Гц) покрывался неравномерной сеткой с 31 узлом. Разрешение модели по угловой координате составляло 10°. Шаг интегрирования по времени в уравнении (1) равнялся 30 мин. Донное трение определялось на основе теории Гранта-Мадсена [11]. На входе в модель задавались скорость ветра на стандартной высоте 10м с дискретностью по времени 3 ч.

Основными выходными характеристиками модели являются: высота значительных волн (среднее значение высот 1/3 наиболее высоких волн в спектре волнения); среднее направление распространения волн; средний период волн.

В ходе разработки системы была выполнена адаптация модели SWAN для учета особенностей ее функционирования в составе ЧЦМП. Было улучшено представление данных по батиметрии Черного моря и осуществлено согласование входных и выходных данных модели с номенклатурой и форматами представления данных ЧЦМП.

Автоматическая система прогноза волнения в настоящее время работает в автономном оперативном режиме и выполняет стандартную последовательность действий, предусмотренных при выполнении прогнозов – от планирования вычислений до выдачи результатов расчетов высоты, направления и периода значимых волн на серверы ЧЦМП. Атмосферное воздействие (скорость ветра) задается согласно данным греческого центра атмосферных прогнозов SKIRON, регулярно принимаемым ЧЦМП по каналам Internet и подготавливаемым для использования в моделях ЧЦМП.

Функциональная схема системы показана на рис.1.



Рисунок 3.3 – Функциональная схема

Управление ее работой осуществляется с помощью скрипт-файлов, разработанных на языке CalcmanPack и исполняемых программами-интерпретаторами пакета программ CalcMan [13]. В процессе функционирования система выполняет следующую последовательность действий:

- планирование выполнения прогнозов;

- ожидание времени запуска очередного цикла прогноза;

-подготовка исходной информации, необходимой для выполнения очередного цикла прогноза;

- прогноз характеристик волнения с использованием модели SWAN;

- архивирование входных и вспомогательных данных, использованных и созданных при выполнении очередного цикла прогноза, в локальном файловом архиве, размещенном на файл-сервере ЧЦМП;

- подготовка выходных NetCDF-файлов данных прогноза волнения, архивирование результатов в локальном файловом архиве;

- выгрузка результатов прогноза на THREDDS-сервер для обеспечения пользователям доступа к данным;

- завершение цикла работы, переход к планированию следующего цикла вычислений.

В настоящее время, по соображениям экономии ресурсов дисковой памяти с файл-сервера ЧЦМП могут быть загружены данные за любой промежуток времени в пределах с 90 дня в прошлое, до пяти дней в будущее от текущей даты.

Данные прогнозов модели SWAN записываются в файлы формата NetCDF. Структура файла изображена на рисунке 3.4. В имени файла содержащих результаты прогноза модели указывается дата, на которую был осуществлен прогноз и дата, когда прогноз был составлен. Примеры файлов модельных расчетов приведены на рисунке 3.5.

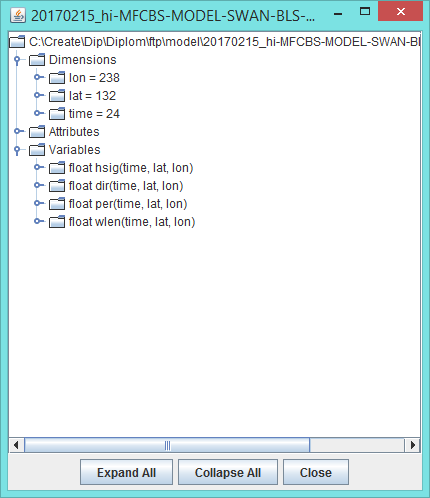


Рисунок 3.4 – Структура файла

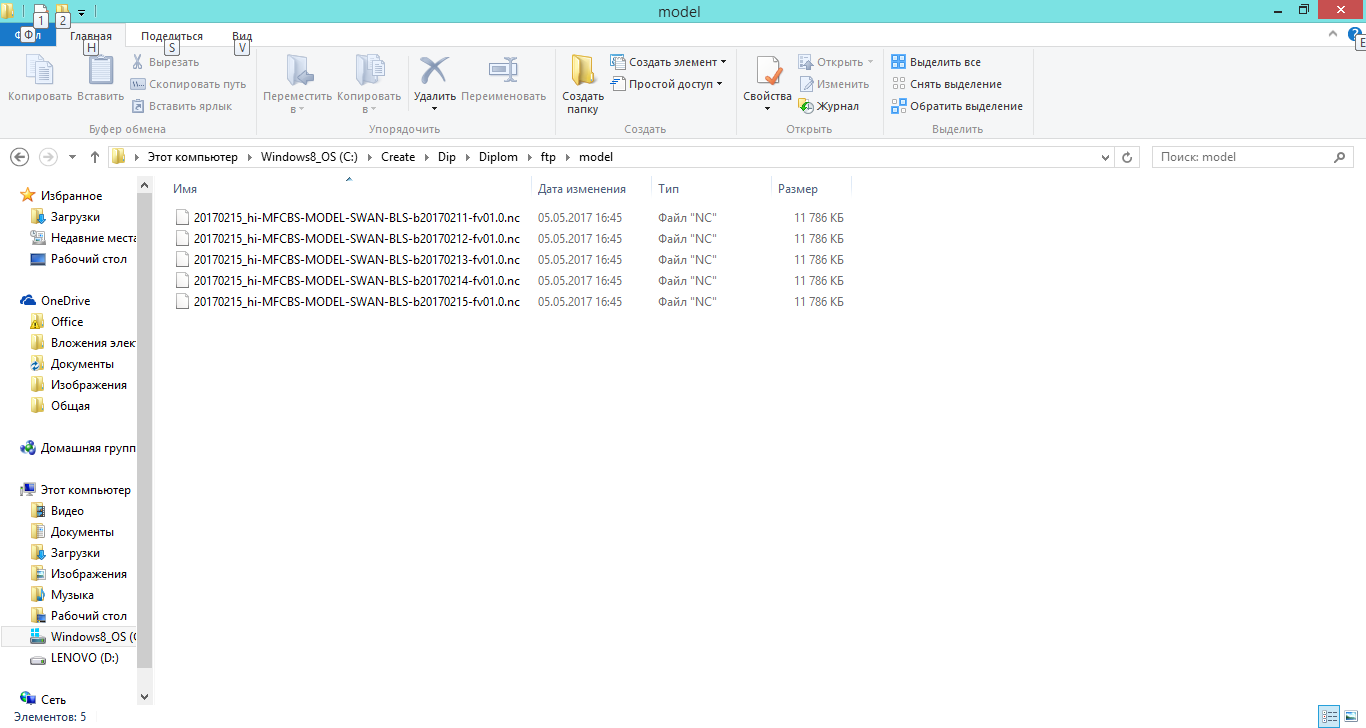


Рисунок 3.5 – Пример файлов

Приложение состоит из двух частей – главного модуля «wave\_model\_validation» и подключаемых модулей. На рисунке 3.1 представлена структура проекта.

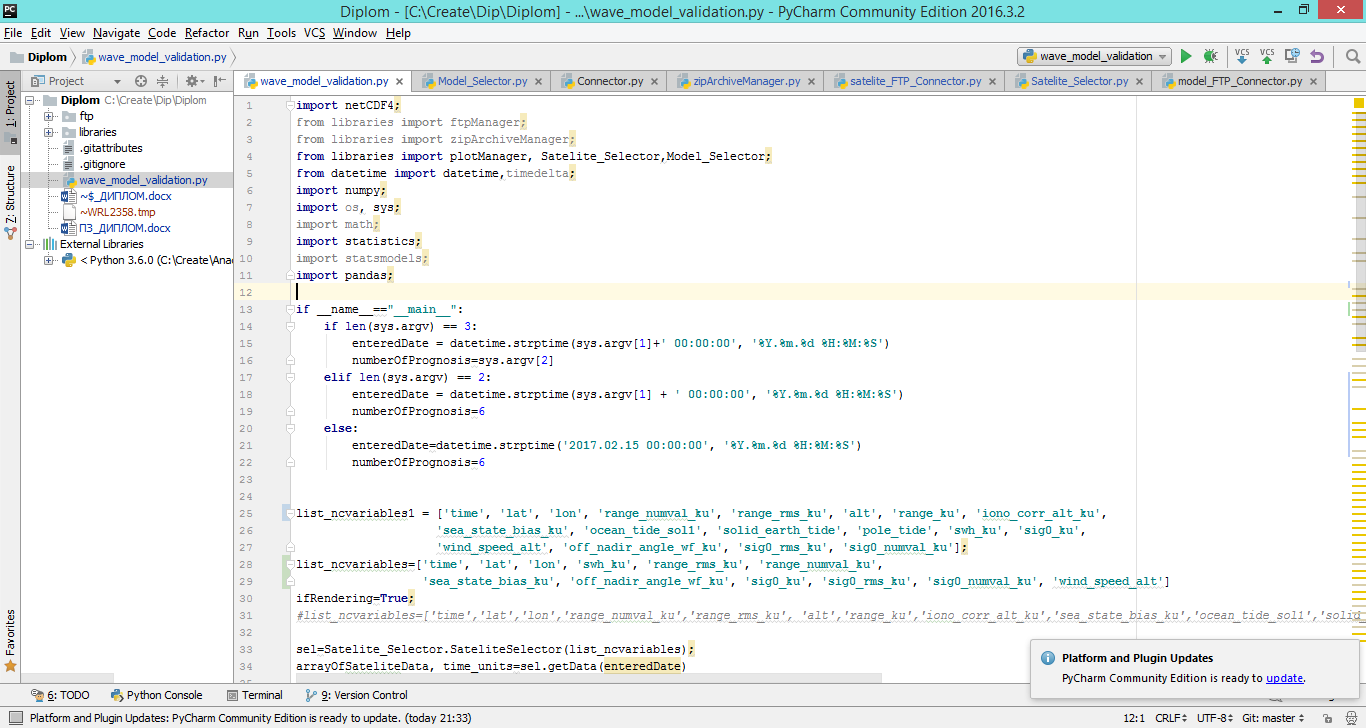


Рисунок 3.1 – Структура проекта

При вызове производится через командную консоль, где указывается название главного модуля «wave\_model\_validation» и два параметра. Первый параметр является обязательным и представляет собой дату, на которую необходимо произвести оценку точности расчетов модели. Второй параметр является диапазоном прогнозов модели на указанную дату, диапазон прогнозов находится в пределах от 1 суток до 5.

В модуле «wave\_model\_validation» производится анализ запроса и формируются условия для выборки и обработки файлов.

Для того что бы произвести оценку точности в модуля «wave\_model\_validation» производится инициализация классов SateliteSelector и ModelSelector

После обработки запроса и формирования условий обработки файлов альтиметрических измерений производится загрузка архивов данных.

Для загрузки архивов данных с FTP серверов AVISO и МГИ РАН были разработаны классы SateliteFTPConnector и ModelFTPConnector, соответственно. Оба класса наследуются от интерфейса ConnectorInterface. На рисунке 3.2 изображена UML-диаграмма классов SateliteFTPConnector и ModelFTPConnector.

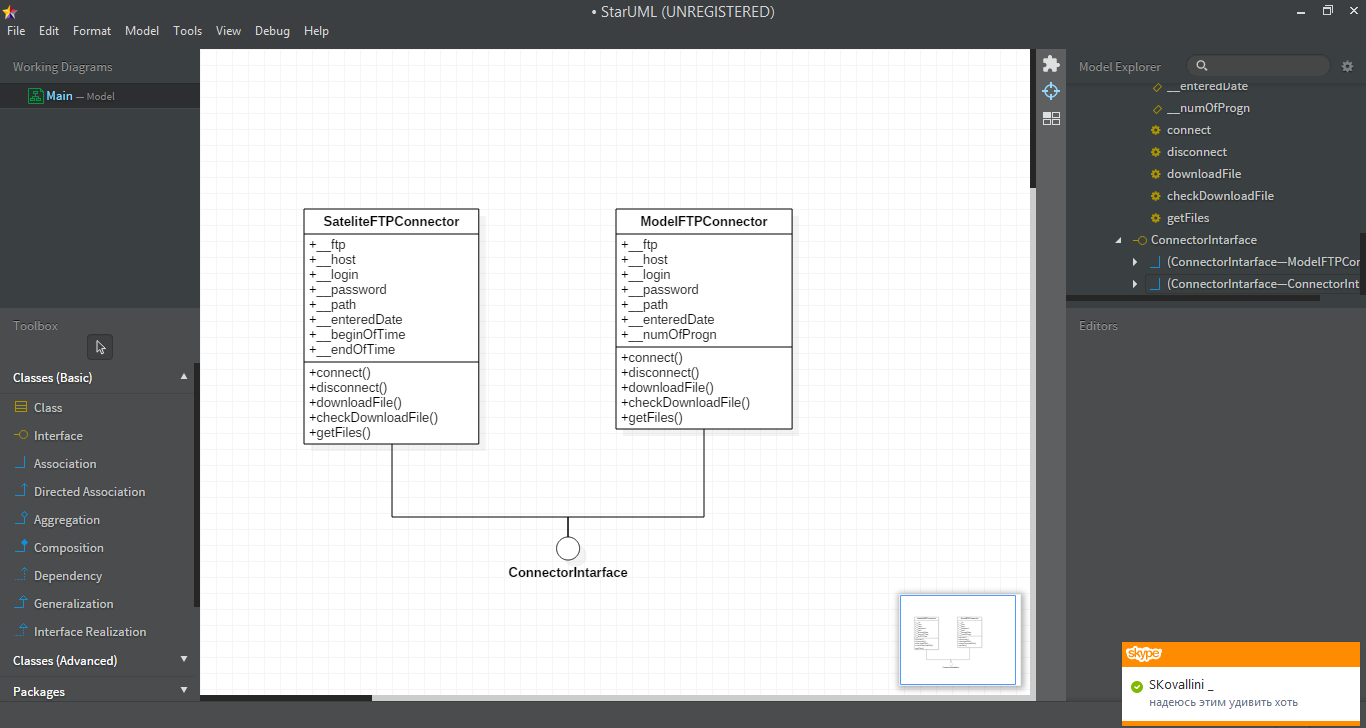


Рисунок 3.2 – Диаграмма классов

При инициализации экземпляра класса SateliteFTPConnector передаются обязательные параметры: имя хостинга FTP-сервера, логин, пароль и необязательный параметр директория загрузки файлов. Для загрузки архивов данных альтиметрических измерений необходимо вызвать метод getFiles() с параметром задающим директорию с необходимыми архивами. После вызова метода getFiles() вызывается метод connect() и устанавливается соединение с FTP-сервером. Производится поиск необходимых архивов по директориям и их последующая загрузка с помощью метода downloadFile(). Производится проверка загруженных архивов на наличие потерь. После завершения всего цикла загрузки вызывается метод disconnect() и соединение с FTP-сервером обрывается. На рисунке 3.3 представлена диаграмма класса SateliteFTPConnector.

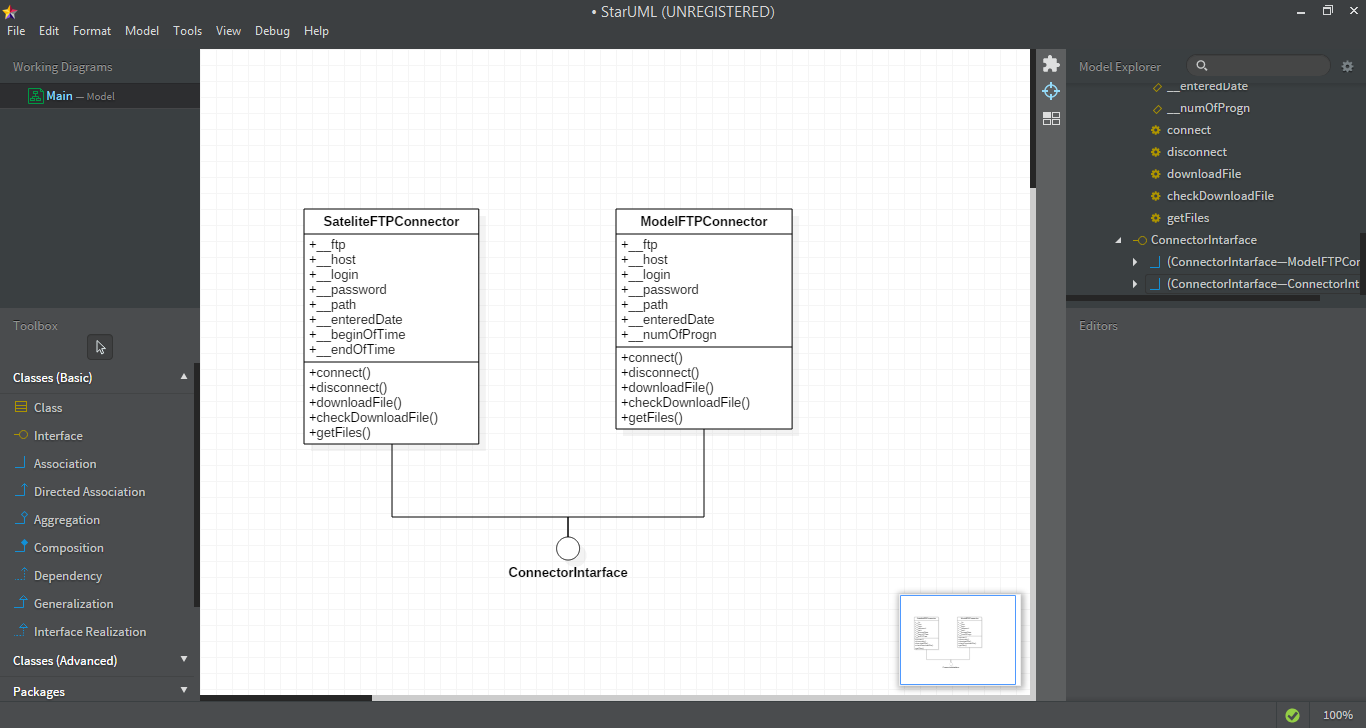


Рисунок 3.3 – Диаграмма класса SateliteFTPConnector

* 1. Разработка компонентов
  2. Разработка интерфейсов
  3. Разработка возможностей и эксплуатации
  4. Тестирование, верификация, валидация ПО (ИО, АО, МО) ИС

1. АНАЛИЗ ПРОЦЕССОВ НА ОСНОВЕ ИС