



АНАЛИЗ СТРУКТУРЫ ДАННЫХ СТРАТОСФЕРНОГО ПОЛЯРНОГО ВИХРЯ С ПРИМЕНЕНИЕМ САМОКОНТРОЛИРУЕМОГО ОБУЧЕНИЯ



Демидова Е.А. ¹, Криницкий М.А. ^{1,2}
Московский физико-технический институт
Институт океанологии имени П.П. Ширшова

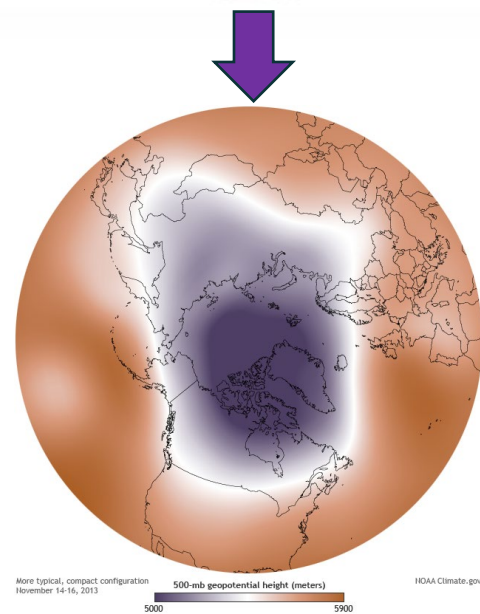
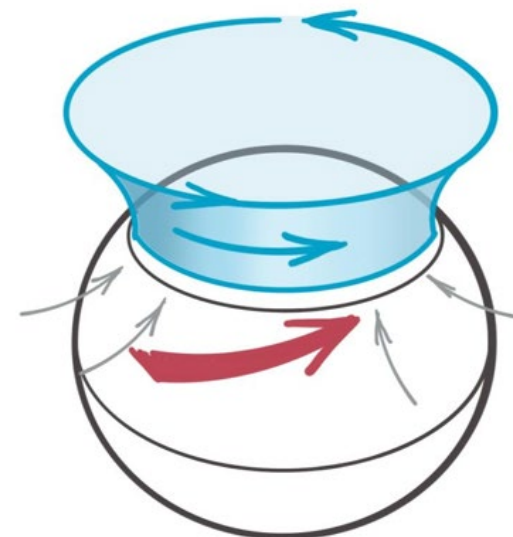


ВВЕДЕНИЕ

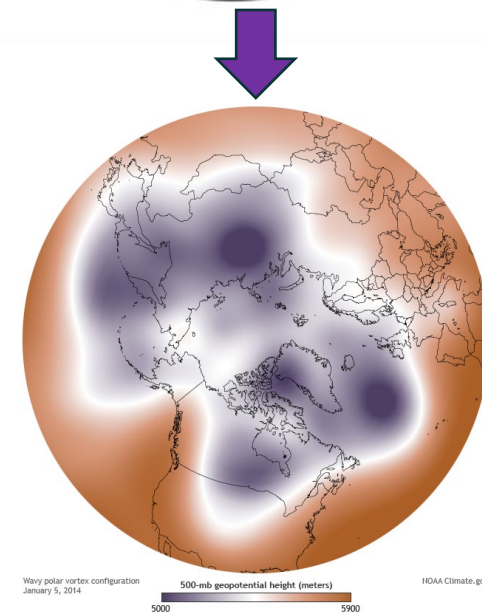
Стратосферный полярный вихрь (СПВ) - это крупномасштабный циклон, который формируется в стратосфере над полюсами зимой. Его возникновение вызвано сильным меридиональным температурным градиентом во время полярной ночи.

- **СИЛЬНЫЙ/УСТОЙЧИВЫЙ** (с центром над полюсом)
- **СМЕЩЁННЫЙ**
- **РАЗДЕЛЕННЫЙ ВИХРЬ** (несколько циклонов)

Внезапные стратосферные потепления(ВСП) - ослабление или разрушение СПВ. Они приводят к смещению траекторий циклонов и экстремальным погодным явлениям, таким как холодные вторжения в средние широты



Стабильный полярный вихрь, ноябрь 2013



Ослабленный полярный вихрь, январь 2014

Актуальность исследования:

Актуальность данного исследования обусловлена необходимостью создания объективного, автоматизированного метода классификации состояний стратосферного полярного вихря, который сохраняет его пространственную структуру. Такой инструмент важен для:

- Улучшения понимания динамики и изменчивости СПВ.
- Исследования стратосферно-тропосферных взаимодействий.
- Повышения качества сезонных прогнозов (S2S).
- Изучения влияния климатических изменений на полярную стратосферу.



«Beast from the East»(Европа), 2018. Образовался из – за шторма вызванного ВСП.

4



Техас, 2021. Полярный вихрь был растянут и потоки холодных ветров распространялись гораздо дальше над американским и евразийским континентами



ПРОКЛЯТИЕ РАЗМЕРНОСТИ

РАЗМЕРНОСТЬ ИСХОДНЫХ ДАННЫХ:

131 072 – очень высокая



евклидово расстояние теряет
свою информативность



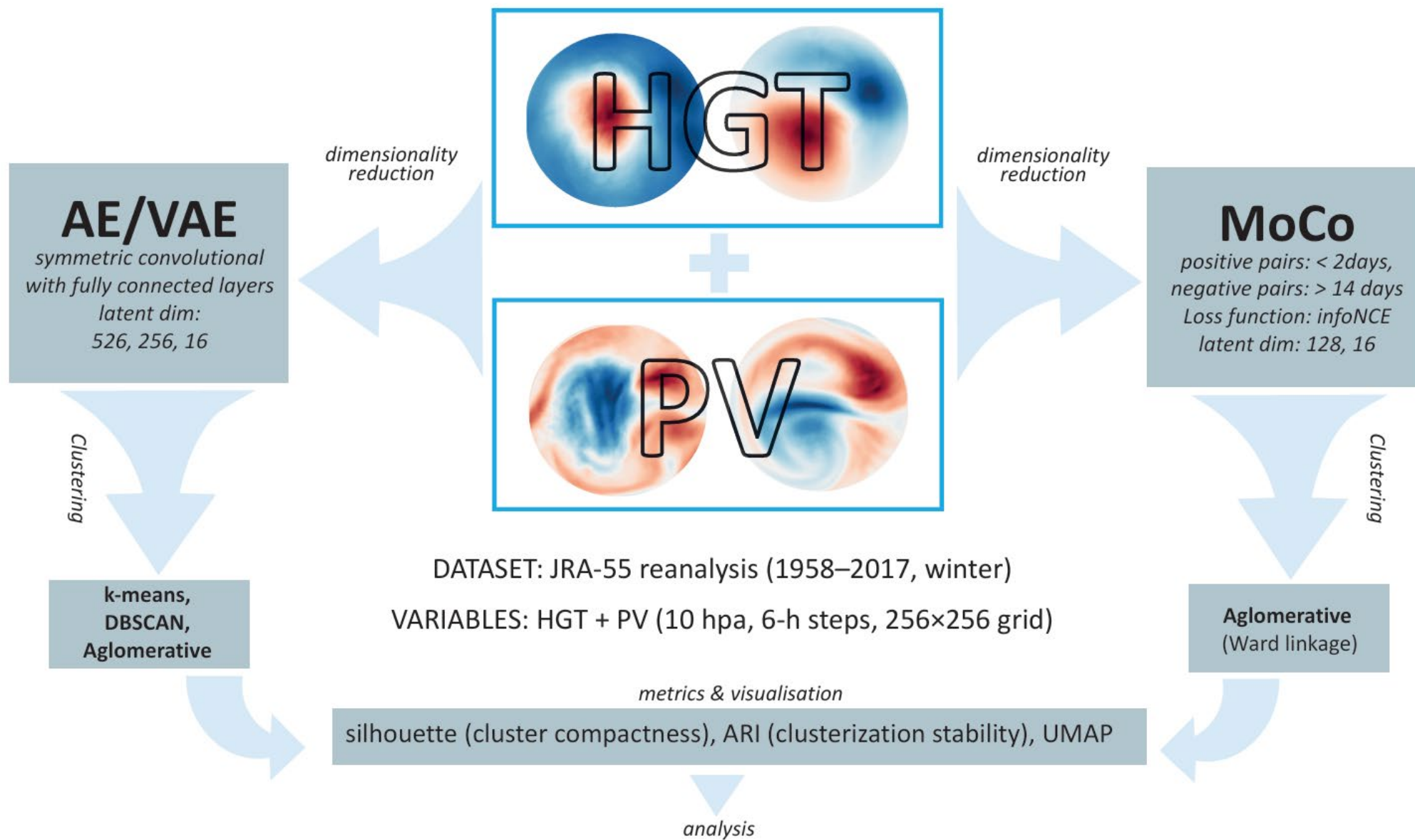
Решение: нелинейное
снижение размерности



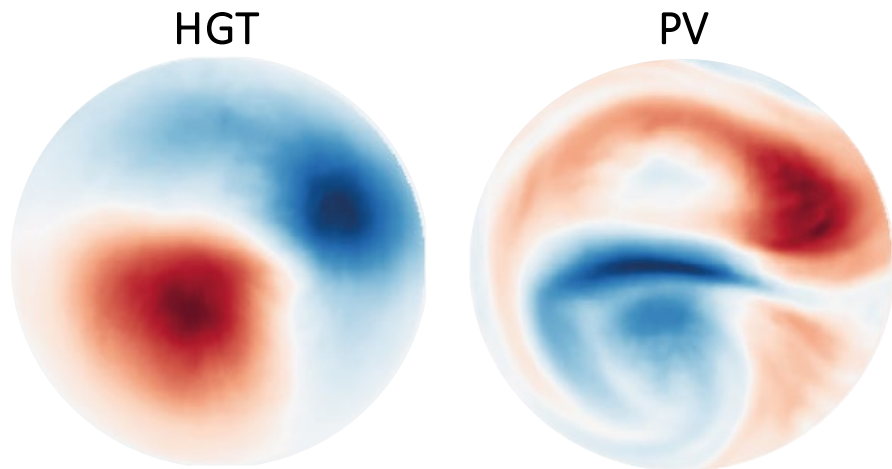
ПОЛУЧАЕМАЯ РАЗМЕРНОСТЬ ДАННЫХ:

16, 32, ..., 512 с сохранением семантической близости

СХЕМА ИССЛЕДОВАНИЯ



ДАННЫЕ



HGT - это высота определённой изобарической поверхности над уровнем моря. Это "рельеф" атмосферы, где низкий геопотенциал на полюсе означает, что холодный, плотный воздух занимает меньший вертикальный столб.

PV — это мера вращательного движения воздушной частицы, "взвешенная" на её стабильность. Граница полярного вихря — это не плавный переход, а резкий градиент PV. Эта граница действует как барьер для перемешивания между полярным и среднеширотным воздухом.

$$PV = -g (\zeta + f) \frac{\partial \theta}{\partial p}$$

где g - ускорение свободного падения,
 ζ — относительная завихренность,
 f - параметр Кориолиса,
 θ - потенциальная температура,
 p - давление.

Источник данных: реанализ JRA-55

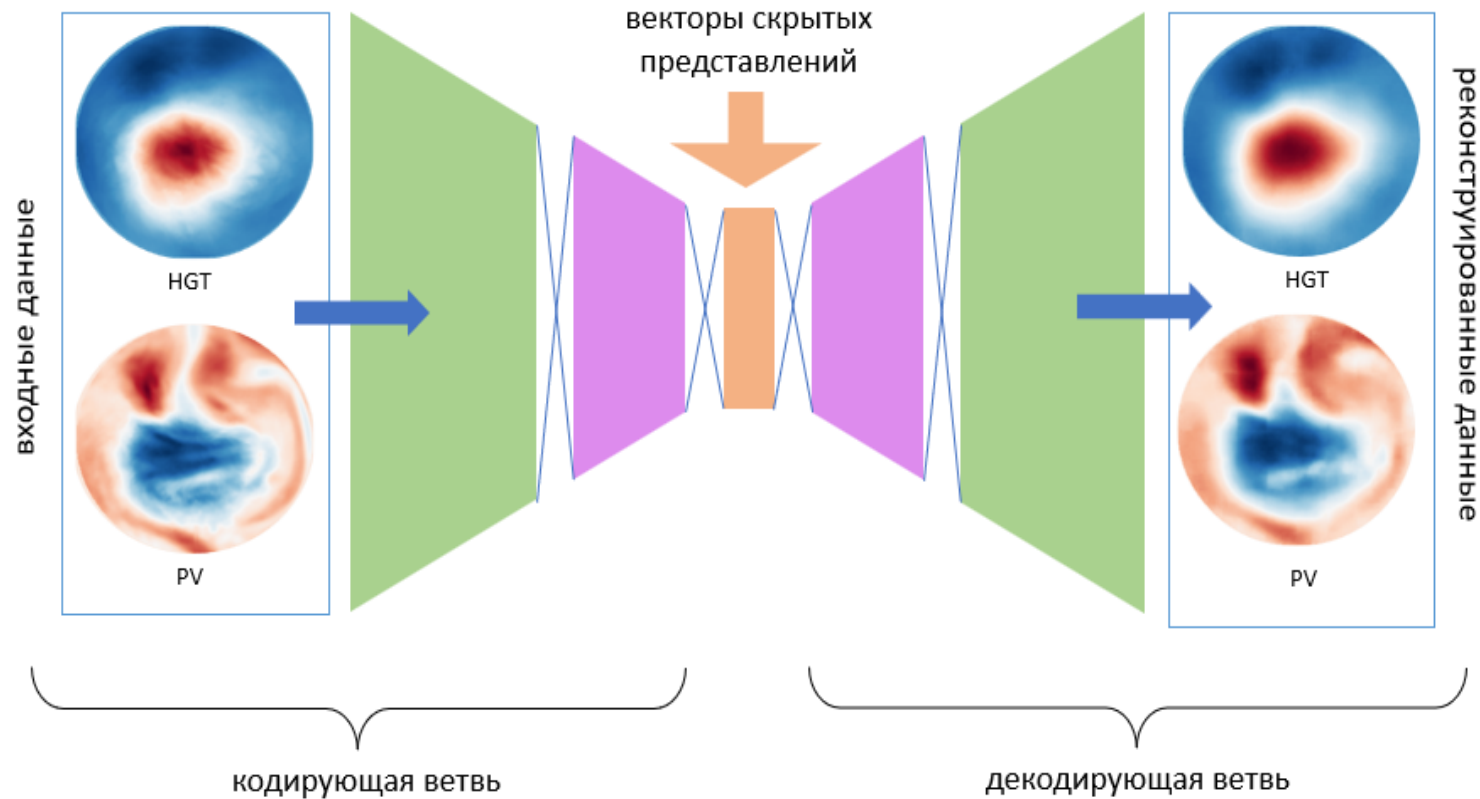
Поля: геопотенциальная высота (HGT) и
потенциальная завихренность (PV) на уровне 10 гПа.

Период: Зимы (декабрь-февраль) с 1958 по 2017 гг.

Временной интервал: 6 часов.

Пространственное разрешение: Выделение области к северу от 40°с.ш., преобразование в полярную стереографическую проекцию, интерполяция на регулярную сетку 256x256

АВТОКОДИРОВЩИКИ



Автокодировщик:

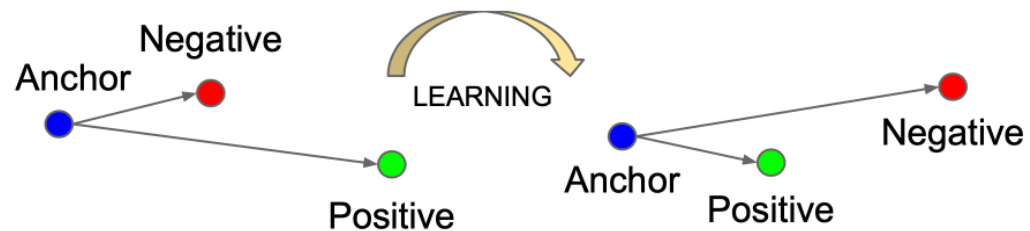
сжимает входные данные в «узкое горлышко», а затем восстанавливает их обратно. Учится только воспроизводить вход.

Вариационный автокодировщик:

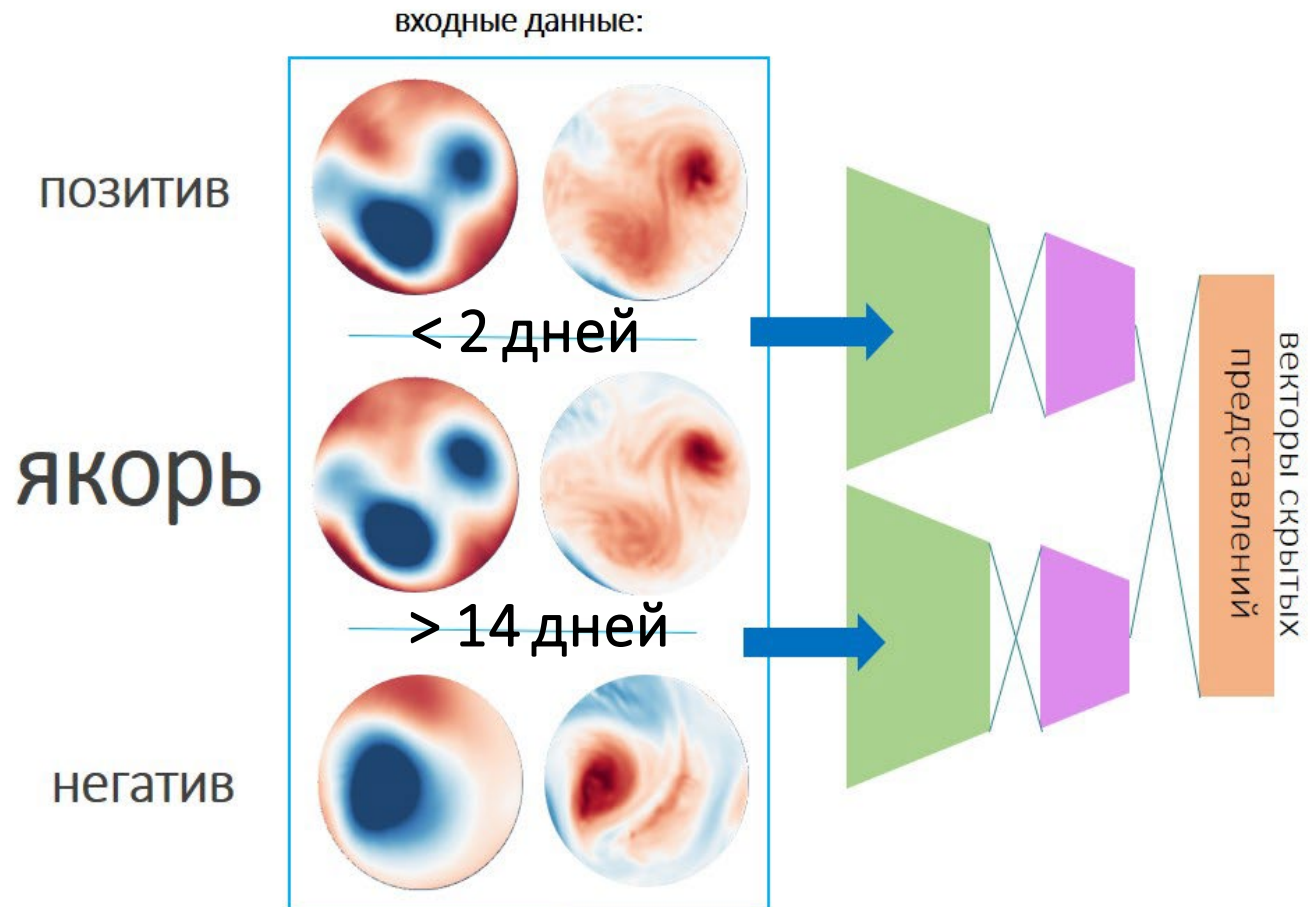
учится не просто сжимать, а представлять данные в виде вероятностного распределения.

КОНТРАСТНОЕ ОБУЧЕНИЕ

Это метод машинного обучения, направленный на умение модели различать похожие и непохожие объекты, формируя скрытые векторные представления,



где «позитивные» пары сближаются, а «негативные» — удаляются друг от друга в пространстве признаков.



Силуэтный коэффициент

— это метрика качества кластеризации, которая показывает, насколько хорошо каждый объект соответствует своему кластеру по сравнению с другими кластерами,

где 1 – идеальное соответствие
-1 – совпадение хуже случайного.

$$s = \frac{b - a}{\max(a, b)}$$

a - внутрикластерное расстояние,
 b - межкластерное расстояние

K-means

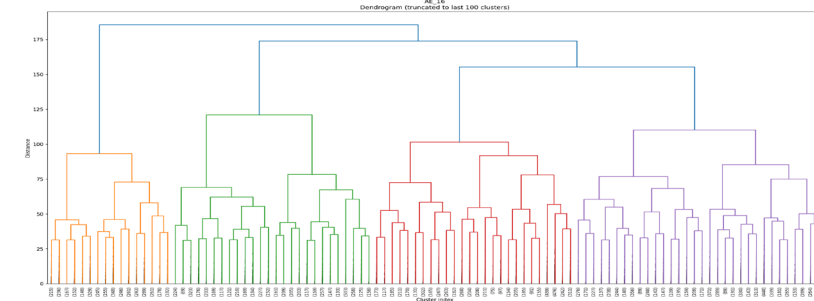
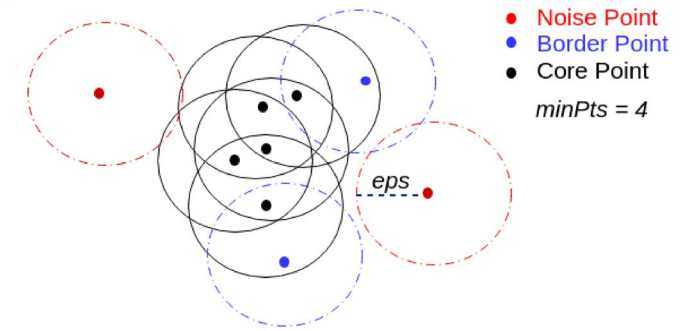
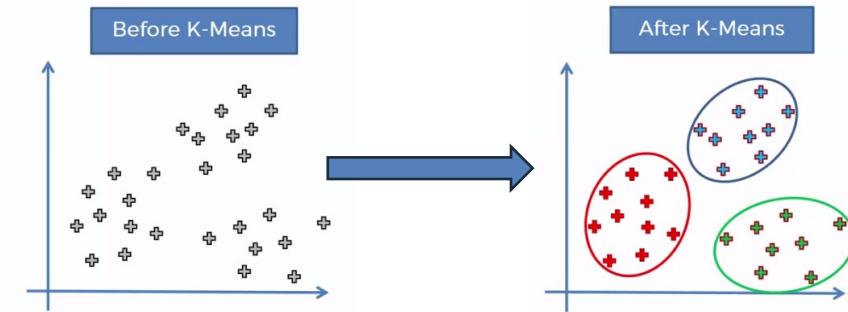
Разбиение на k сферических кластеров

DBSCAN

Поиск плотных областей, выделение выбросов

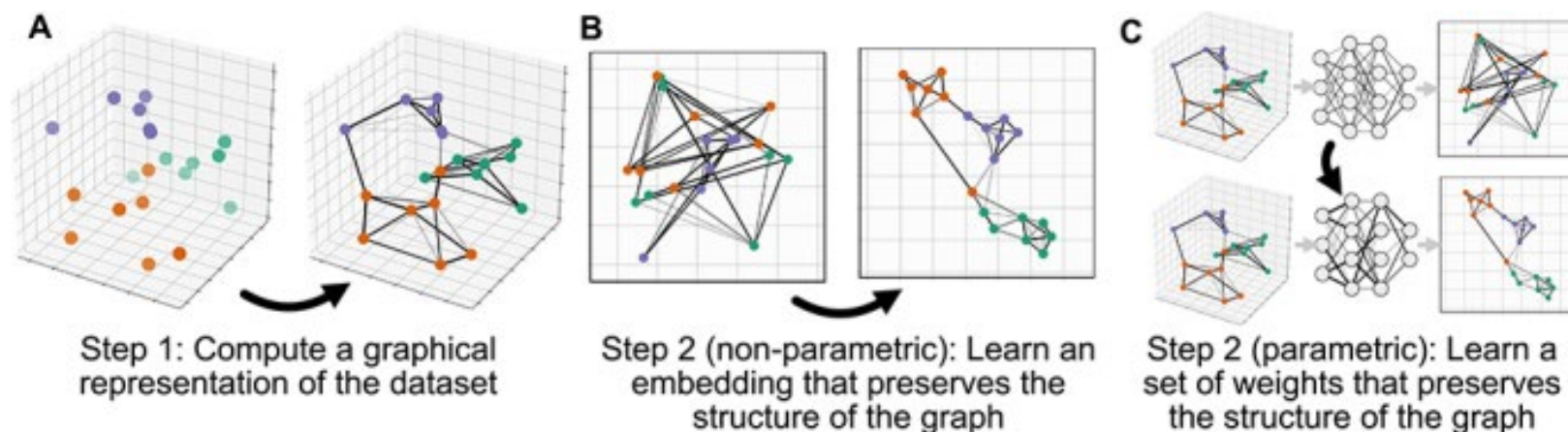
Agglomerative

Иерархическое объединение ближайших кластеров



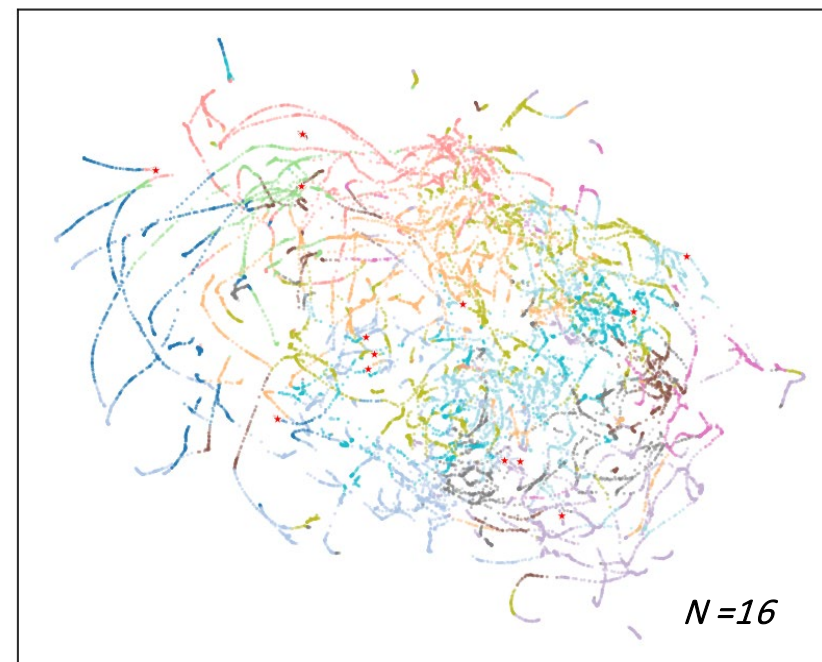
УМАР ВИЗУАЛИЗАЦИИ:

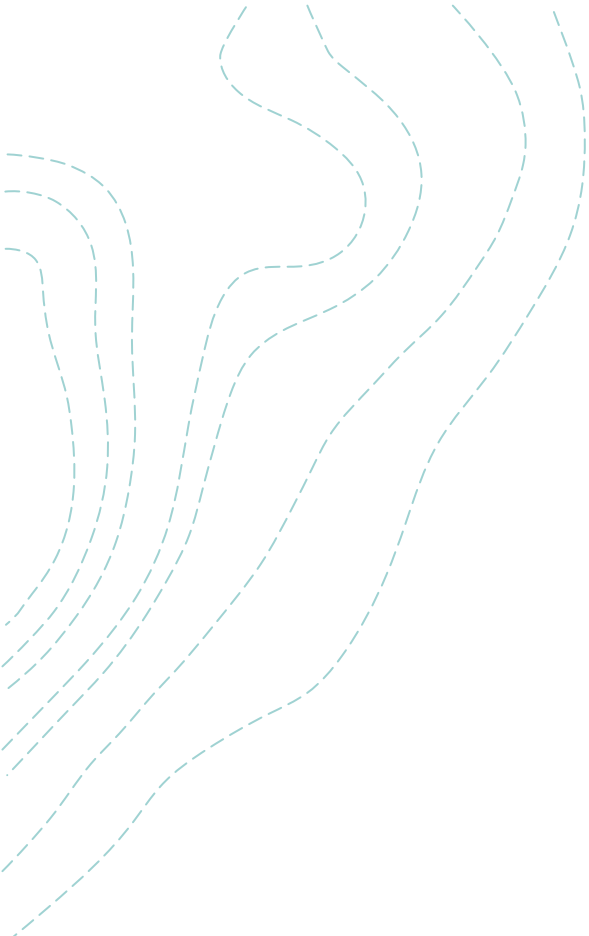
Нелинейный метод снижения размерностей позволяющий сохранить близость между точками данных в многомерном и двухмерном пространстве. ➡



Пример визуализации скрытого представления с помощью UMAP ➡

Ленточная структура отражает плавные переходы между состояниями СПВ и может отражать временную зависимость данных





РЕЗУЛЬТАТЫ

ЗНАЧЕНИЯ СИЛУЭТА:

при агломеративной кластеризации

Автокодировщик:

N	K = 3	K = 12	K = 30
<u>16</u>	<u>0.034</u>	0.019	0.027
256	0.008	0.005	0.07
512	0.003	-0.019	-0.015

Вариационный

N	K=3	K=12	K=30
16	0.001	-0.005	-0.008
128	-0.003	-0.009	-0.014

Контрастное обучение:

Архитектура, N	K = 3	K = 12	K = 30
<u>Contrast 16</u>	<u>0.081</u>	0.062	0.072
Contrast 128	0.068	0.06	0.077
MoCo 16	0.038	0.029	0.039

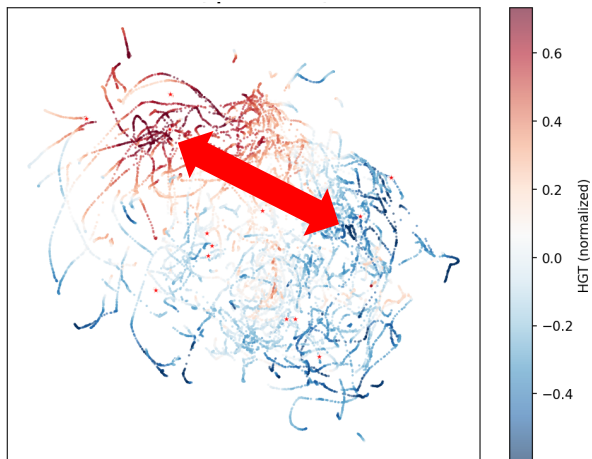
Значения случайного
коэффициента силуэта < 0

Значения коэффициента силуэта
близки к нулю или отрицательны

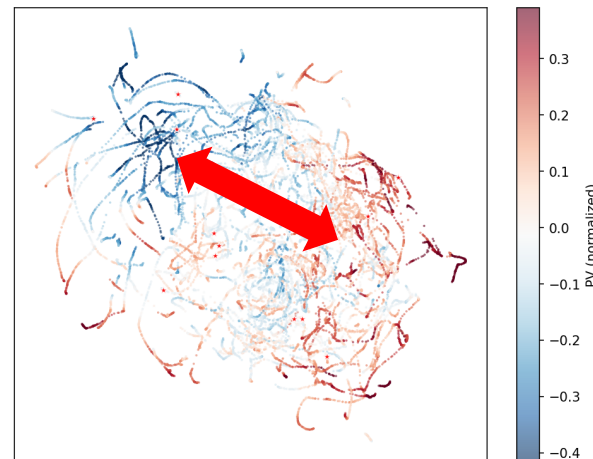
**четкие кластера
не выделяются**

N – размерность скрытого пространства
K – количество кластеров

среднее HGT

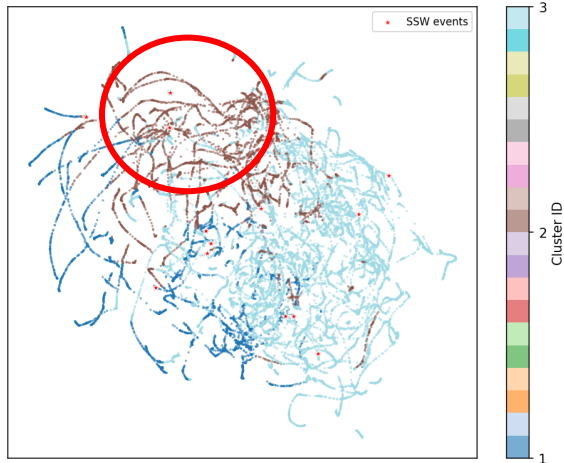


среднее PV

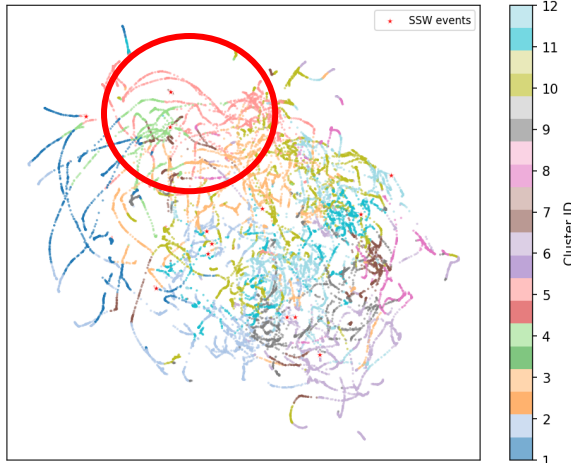


Разные значения усредненного
физического поля в пространстве
скрытого представления
разносятся на дальние расстояния

AE_16, k=3
Clusters



AE_16, k=12
Clusters

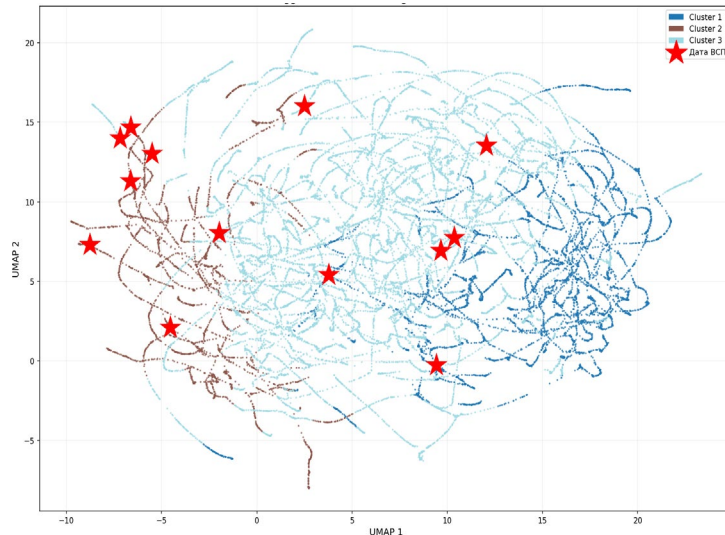


AE_16, k=30
Clusters

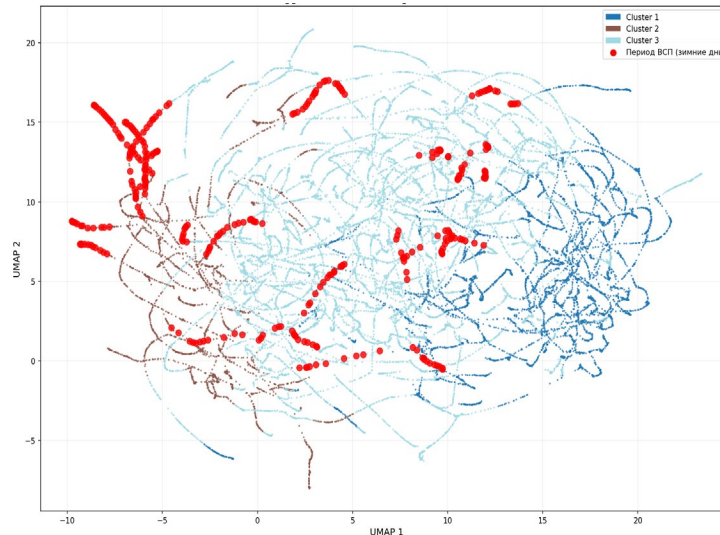


Кластеризации стремятся
выделить
неустойчивый/разрушенный
вихрь в отдельный класс

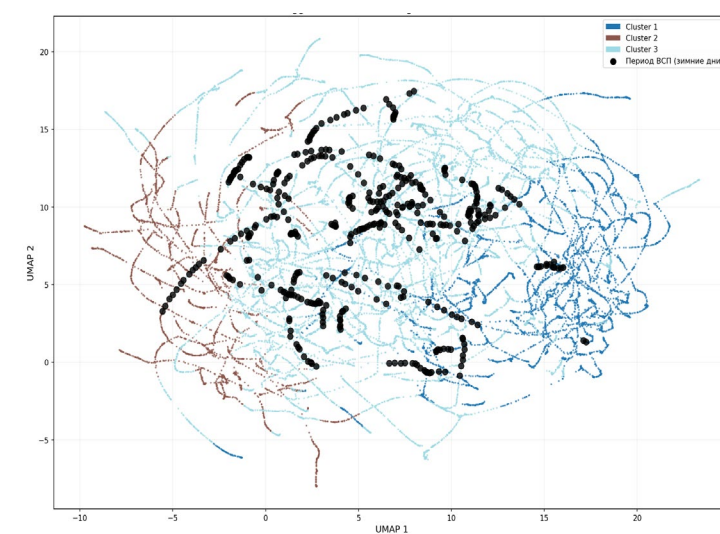
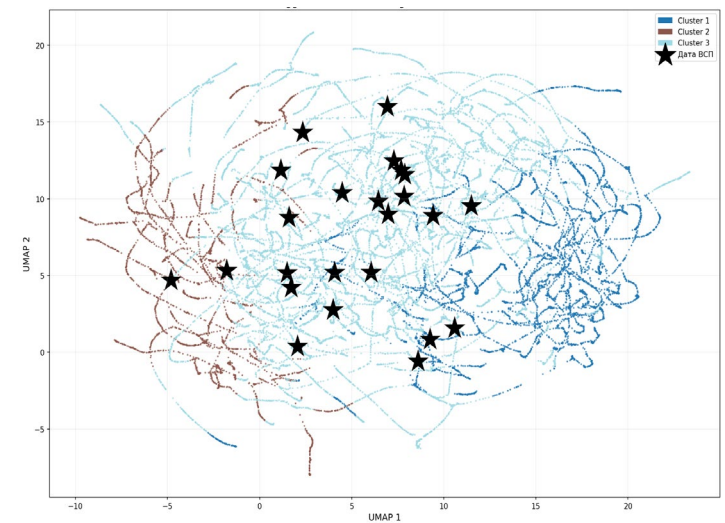
ВНЕЗАПНЫЕ СТРАТОСФЕРНЫЕ ПОТЕПЛЕНИЯ



даты максимальной температуры



все данные между началом и концом ВСП



← **МАЖОРНЫЕ
(ГЛАВНЫЕ) ВСП**

разрушение или сильное
смещение вихря, с резким
повышением температуры в
стратосфере и изменением
направления ветра с западного
на восточный

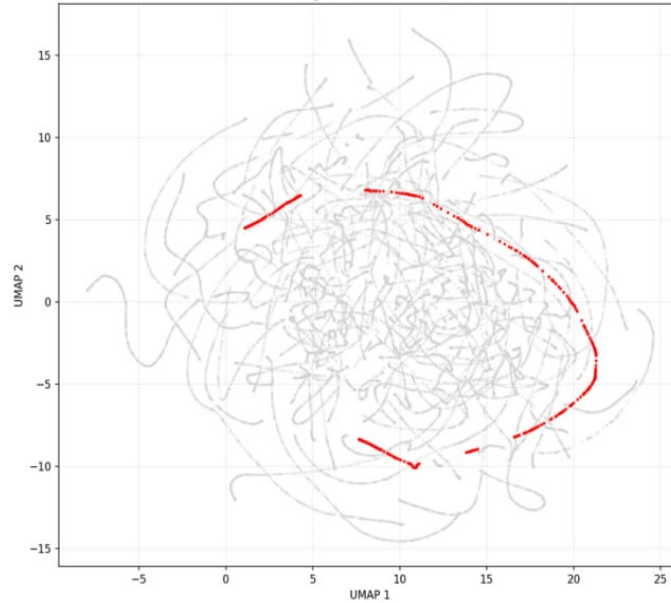
← **МИНОРНЫЕ
ВСП**

частичное ослабление вихря без
полного разрушения или смены
направления ветра, температура
повышается не так резко

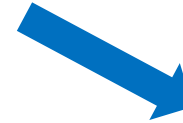
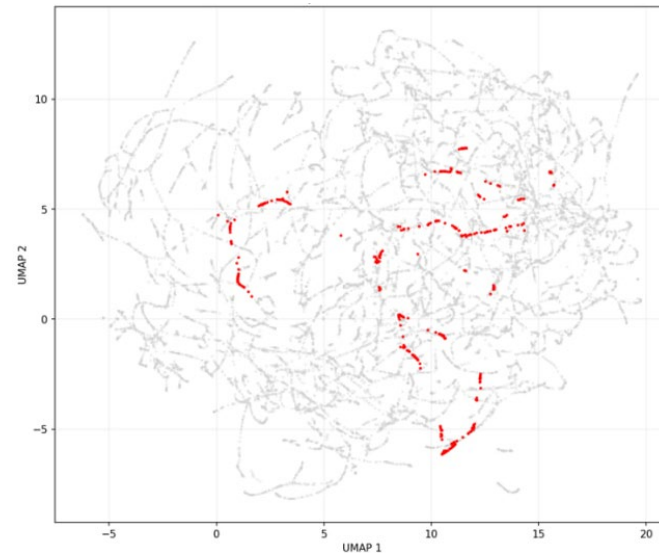
UMAP проекции:

*Зима 1980 - 1981

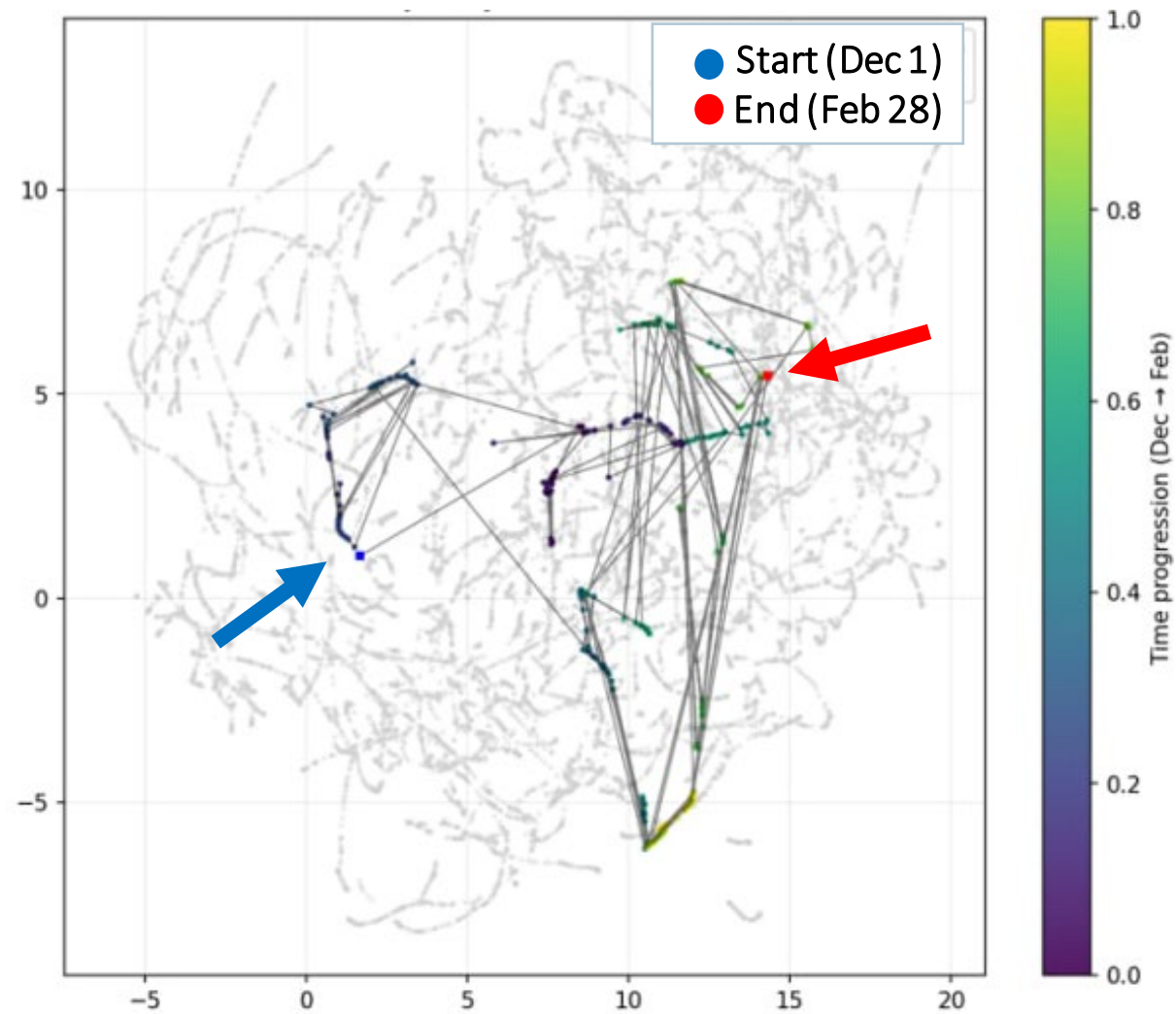
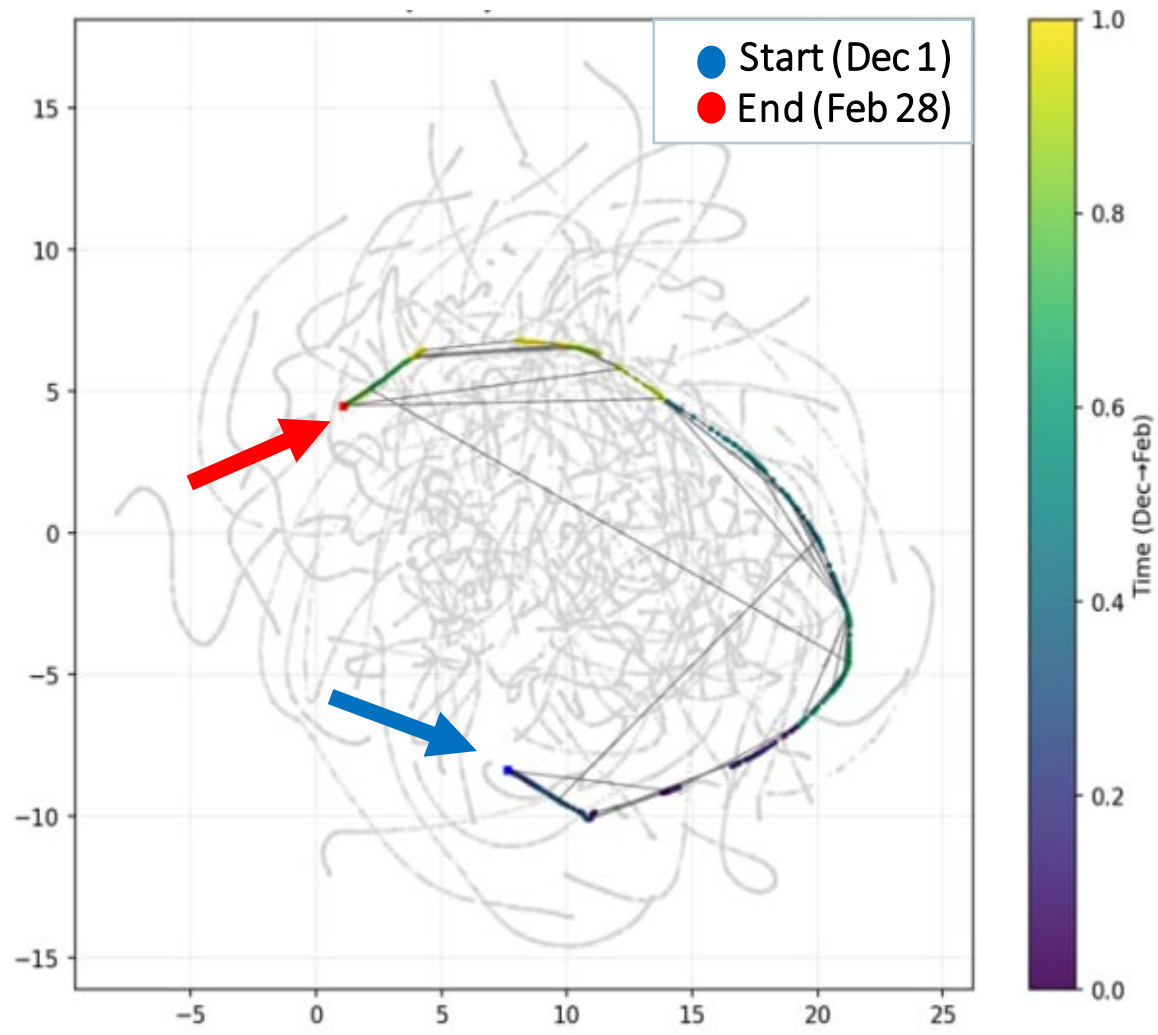
АВТОКОДИРОВЩИК



КОНТРАСТНОЕ
ОБУЧЕНИЕ



Возможно более высокое
значение силуэта при
контрастном подходе
кроется в разреженности
последовательных данных
скрытого представления





ЗАКЛЮЧЕНИЕ

- Переходы между состояниями стратосферного полярного вихря носят плавный характер, без чётких границ между классами.
- Снижение размерности с использованием нейросетевых моделей позволяет сохранить физическую интерпретируемость исходных полей (HGT и PV) в скрытых представлениях.
- Внезапные стратосферные потепления (ВСП) не образуют отдельного кластера, но мажорные и минорные ВСП тяготеют к разным областям в пространстве скрытых представлений, что говорит о сохранении физических инвариант.
- Выбросы в скрытых представлениях могут содержать данные с ярко выраженной структурой вихря, что указывает на потенциальную способность модели выделять кластеры в будущих исследованиях.
- Контрастное обучение показывает более высокие значения силуэтного коэффициента.
- В зависимости от предпосылок, заложенных в основу метода снижения размерности, получаемые представления могут использоваться в различных задачах — от прогнозирования до анализа структуры данных.
- Модели снижения размерности и кластеризации устойчивы к случайной инициализации и изменению параметров, что подтверждает надёжность полученных результатов.

СПАСИБО ЗА ВНИМАНИЕ!

