

Слайд 1. Я хотел бы показать вам слайды по статье Spatio-temporal filling of missing points in geophysical data sets. В геофизике проблема пропущенных значений является очень актуальной. Например, при измерении температуры поверхности воды можно получить очень точные оценки с помощью наблюдения в инфракрасном диапазоне, но для конкретной точки из-за облачности может быть неизвестна температура в 70 процентах моментов времени. Часто пропущенные значения можно восстановить с высокой точностью с помощью различных моделей, так как в геофизических данных много закономерностей и сильные автокорреляции между одним показателем во времени. Для решения задачи заполнения пропущенных значений используется в частности метод анализа сингулярного спектра (SSA), который сворачивает временной ряд в матрицу, содержащую фрагменты исходного ряда с каким-то сдвигом, а затем применяет к этой матрице метод главных компонент. В статье авторы предлагают процедуру надстройки над этим базовым методом с помощью итеративной процедуры интерполяции и оценки параметров модели кросс-валидацией, на этих вводных слайдах я скорее покажу почему базовая модель хорошо подходит для решения этой задачи.

Слайд 2. Результат работы SSA может быть интерпретирован как разложение временного ряда в сумму компонент, у каждой из которых есть осмысленная интерпретация. На картинке вы видите синий исходный временной ряд, разложенный как сумма оранжевого тренда, красного и зеленого периодического сигнала и фиолетового шума. Поскольку каждый из новых рядов является интерпретируемым и описывается какой-то аналитической функцией, мы можем интерполировать найденные главные компоненты, чтобы получить оценку на пропущенные значения в наших данных. В статье авторы предлагают итерационную процедуру, чтобы более точно оценить пропущенные значения при выбранных параметрах модели во внутреннем цикле и кросс-валидацию для подбора оптимальных параметров модели (сдвига) во внешнем цикле, но лежащая в основе идея остается такой же. В итоге авторам удастся получить хорошие оценки на пропущенные значения для различных геофизических наборов данных, что они показывают дальше в статье, рекомендую почитать, достаточно интересно. У меня все, спасибо за внимание.