

Исследование взаимосвязи структурной сложности и прогнозных свойств сигналов СВИ

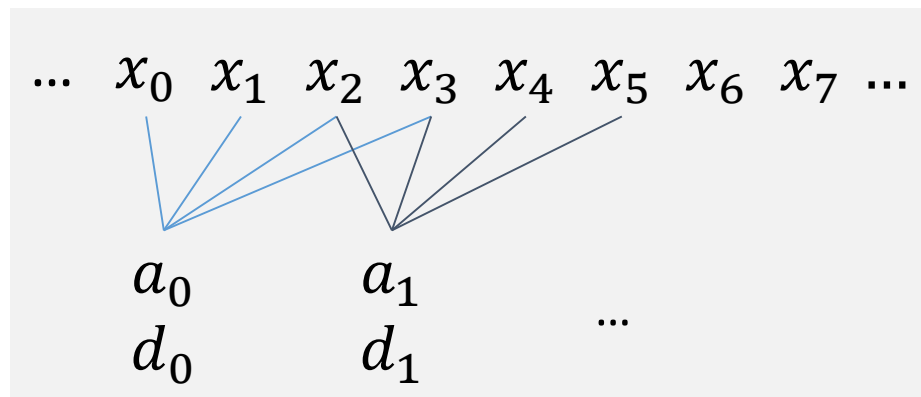
Попов А.И., Бутин К.П., Угрюмов И.А. Родионов А.В.
ООО «Инженерный центр «Энергосервис», Архангельск



Введение

- Поиск применений машинного обучения при обработке данных СВИ приводит к идее о **прогнозировании временных рядов**. Это согласуется с постоянным нарастанием объёмов потоковых данных в СМПР.
- Прогнозирование сигналов СВИ:
 - предиктивный анализ,
 - выявление аномалий и артефактов,
 - восстановление пропусков и искажений.
- На значения измерительных сигналов влияет множество факторов:
 - изменения нагрузки,
 - работа автоматики,
 - переходные процессы,
 - низкочастотные колебания,
 - измерительный тракт.
- **Выбор параметров и методов прогнозирования должен учитывать структурную сложность данных СВИ.**

Дискретное вейвлет-преобразование



Декомпозиция

$$x \xrightarrow{H,G} (a_1, d_1); a_1 \xrightarrow{H,G} (a_2, d_2) \dots$$

$$a_0 = h_0 x_0 + h_1 x_1 + h_2 x_2 + h_3 x_3$$

$$d_0 = g_0 x_0 + g_1 x_1 + g_2 x_2 + g_3 x_3$$

$$a = Hx; d = Gx$$

Реконструкция

$$\dots (a_2, d_2) \xrightarrow{H^*, G^*} a_1; (a_1, d_1) \xrightarrow{H^*, G^*} x$$

$$A_2 = h_2 a_0 + h_0 a_1 \quad D_2 = g_2 d_0 + g_0 d_1$$

$$A_3 = h_3 a_0 + h_1 a_1 \quad D_3 = g_3 d_0 + g_1 d_1$$

$$A = H^* a \quad D = G^* d$$

$$x = H^* a + G^* d$$

Haar

$$H = \left\{ \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right\}$$

$$G = \left\{ \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \right\}$$

Daub2

$$H = \left\{ \frac{1 - \sqrt{3}}{4\sqrt{2}}, \frac{3 - \sqrt{3}}{4\sqrt{2}}, \frac{3 + \sqrt{3}}{4\sqrt{2}}, \frac{1 + \sqrt{3}}{4\sqrt{2}} \right\}$$

$$G = \left\{ \frac{1 + \sqrt{3}}{4\sqrt{2}}, -\frac{3 + \sqrt{3}}{4\sqrt{2}}, \frac{3 - \sqrt{3}}{4\sqrt{2}}, -\frac{1 - \sqrt{3}}{4\sqrt{2}} \right\}$$

Энтропия Колмогорова-Сина (КС-энтропия)

$$K = \lim_{\tau \rightarrow 0} \lim_{l \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{\tau N} \sum_{n=0}^{N-1} (K_{n+1} - K_n)$$

τ – время между наблюдениями

l – точность определения положения системы в фазовом пространстве, разбитом на ячейки l^m

$$K_n = - \sum_{i_0, \dots, i_n} P_{i_0 \dots i_n} \ln P_{i_0 \dots i_n}$$

$P_{i_0 \dots i_n}$ – совместная вероятность того, что при $t = 0, \tau, 2\tau, \dots, n\tau$ система находится в ячейках $i_0, \dots, i_n$

K_n – информация, необходимая для определения положения системы на траектории $i_0, \dots, i_n$ с точностью l

$K_{n+1} - K_n$ – **дополнительная информация для прогнозирования** ячейки фазового пространства, в которой окажется система после прохождения этой траектории

Выборочная энтропия (Sample Entropy)

$$\{x_0, \boxed{x_1, x_2, x_3}, x_4, x_5, \dots\}$$

$$p_i^{(m)} = [x_i, \dots, x_{m-1+i}]$$

$$p_i^{(m+1)} = [x_i, \dots, x_{m-1+i}, x_{m+i}]$$

$$0 \leq SampEn \leq \infty$$

Предельные случаи

- Константа: $SampEn = 0$
- Белый шум: $SampEn \rightarrow \infty$

Функция схожести

$$E_{i,j}(L) = E(p_i^{(L)}, p_j^{(L)}) = \begin{cases} 1, \forall k \in \{0, \dots, L-1\}: |x_{i+k} - x_{j+k}| \leq r \\ 0, \text{ иначе} \end{cases} \quad (\text{на практике } r \sim 0,2\sigma)$$

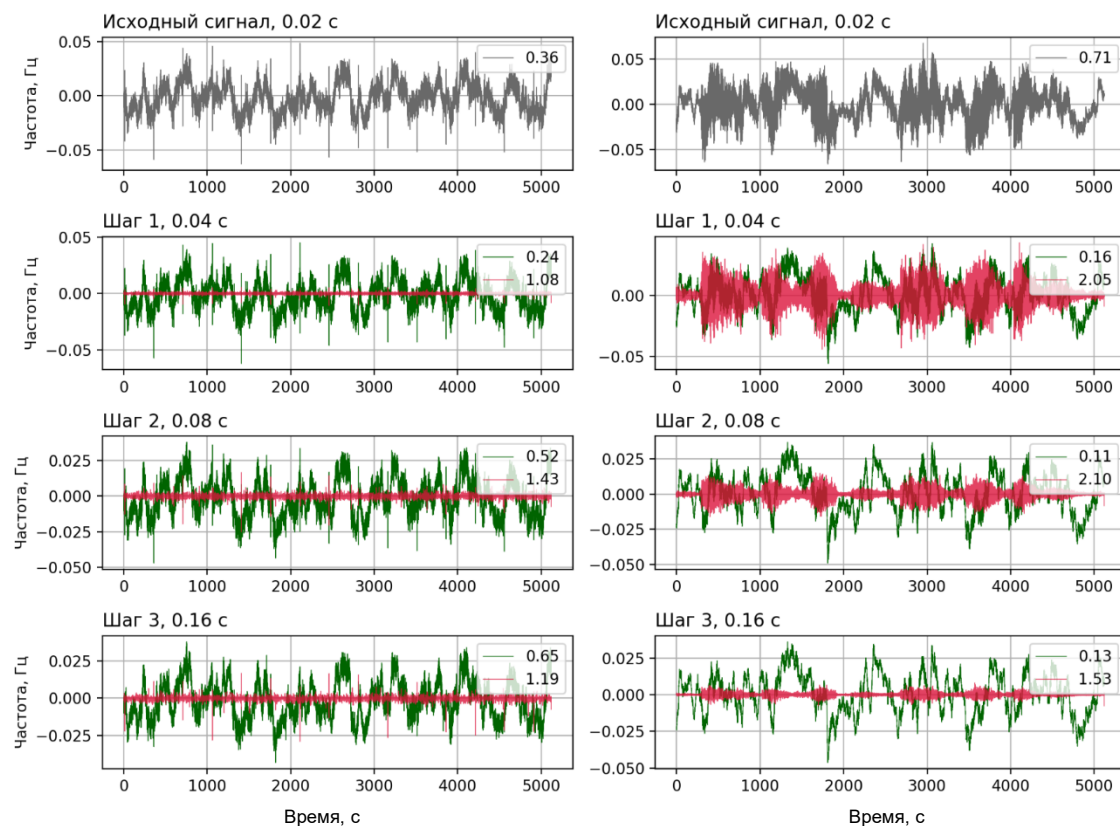
$$SampEn = -\ln \frac{\sum_{i < j} E_{i,j}(m+1)}{\sum_{i < j} E_{i,j}(m)}$$

Разложение сигнала частоты

Каскады с 1 по 3

Локальный КСВД (220 В)

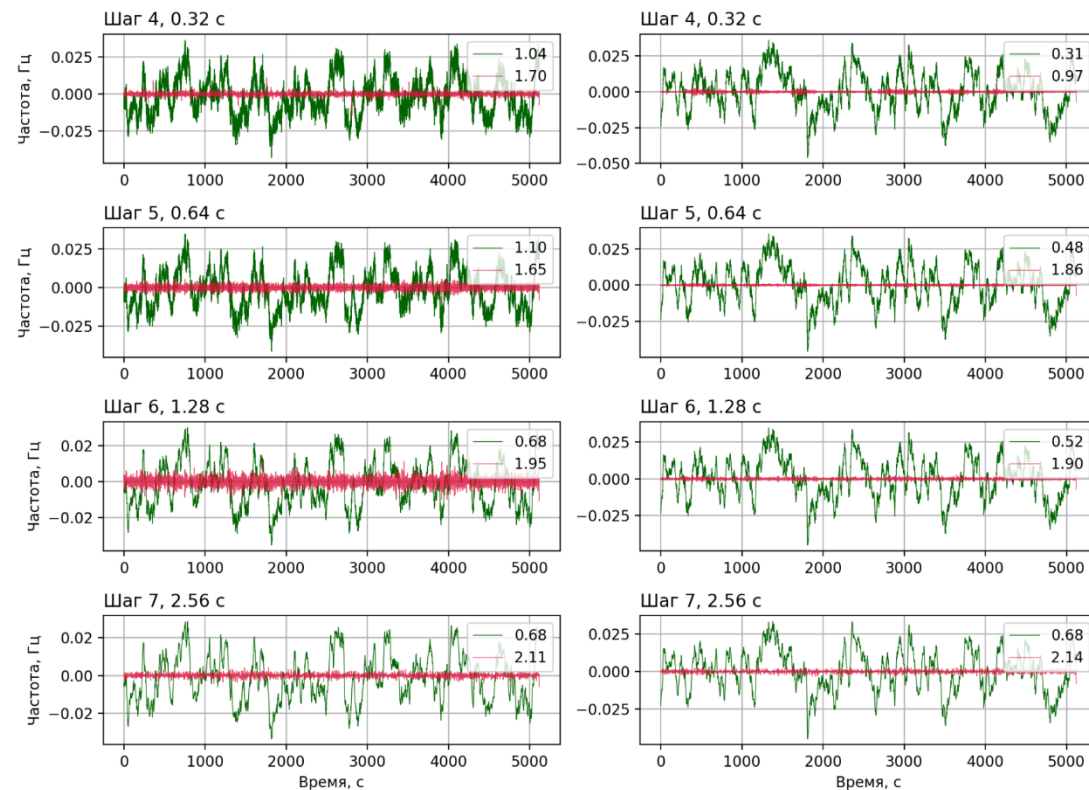
СМПР (линия 500 кВ)



Каскады с 4 по 7

Локальный КСВД (220 В)

СМПР (линия 500 кВ)



Прогнозирование

Выборки

$$\{a_0, a_1, \dots, a_{m-1} \mid a_m, a_{m+1}, \dots, a_{n-1}\} \quad m = \left\lceil \frac{n}{2} \right\rceil$$

обучение тестирование

Схемы

$$(2 \rightarrow 1): \{a_k, a_{k+1}\} \rightarrow \{a_{k+2}\}$$

$$(4 \rightarrow 2): \{a_k, \dots, a_{k+3}\} \rightarrow \{a_{k+4}, a_{k+5}\}$$

$$(8 \rightarrow 4): \{a_k, \dots, a_{k+7}\} \rightarrow \{a_{k+8}, a_{k+11}\}$$

Гиперпараметры

- FCNN
- 2 скрытых слоя по 64 нейрона
- Функция активации ReLU
- Оптимизатор Adam, 200 эпох

Оценка точности

$$MAE = \frac{1}{N} \sum_{i=1}^N |a_i - p_i|$$

a_i – целевое значение

p_i – прогнозное значение

Выборочная энтропия и точность прогноза

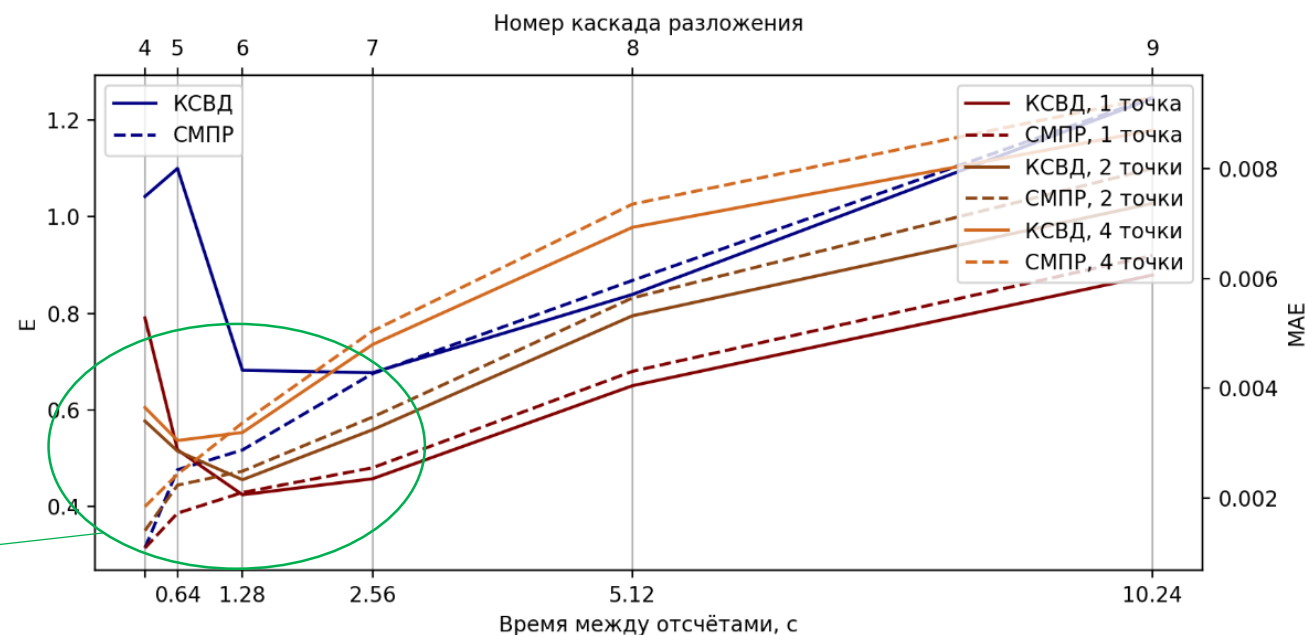
Распределительная сеть (локальный КСВД)

Каскад	T	SampEn	MAE, 1	MAE, 2	MAE, 4
4	0,32	1,042	0,005281	0,003401	0,003649
5	0,64	1,1	0,002878	0,002856	0,003048
6	1,28	0,682	0,00206	0,002334	0,003194
7	2,56	0,677	0,002351	0,003243	0,004795
8	5,12	0,839	0,004046	0,00532	0,006931
9	10,24	1,245	0,006061	0,007368	0,008699

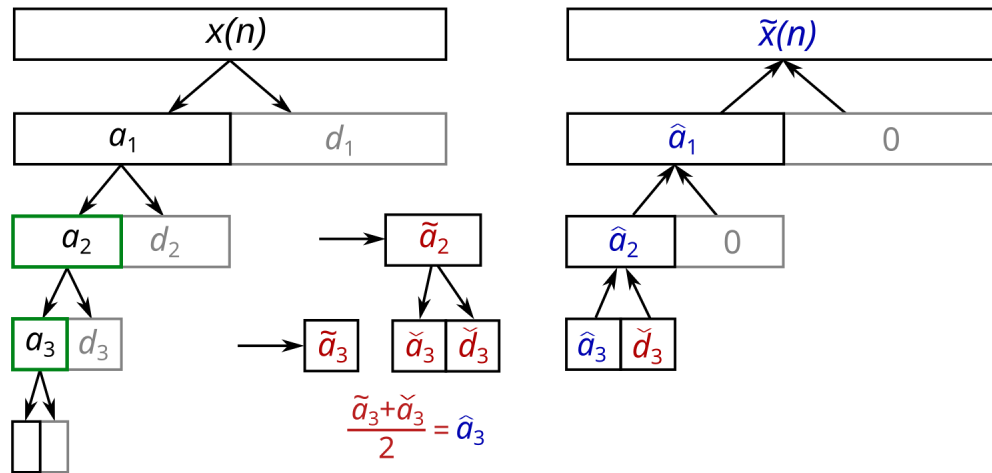
Линия 500 кВ (СМПР)

Каскад	T	SampEn	MAE, 1	MAE, 2	MAE, 4
4	0,32	0,314	0,00109	0,001407	0,001844
5	0,64	0,476	0,001724	0,002234	0,002433
6	1,28	0,517	0,002101	0,002489	0,003369
7	2,56	0,675	0,002551	0,003474	0,005042
8	5,12	0,868	0,004308	0,005645	0,007354
9	10,24	1,247	0,006419	0,008002	0,009297

Область наилучшего прогноза

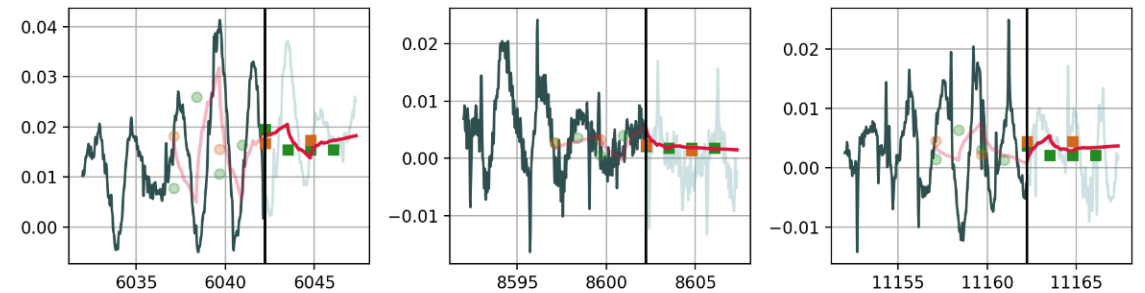
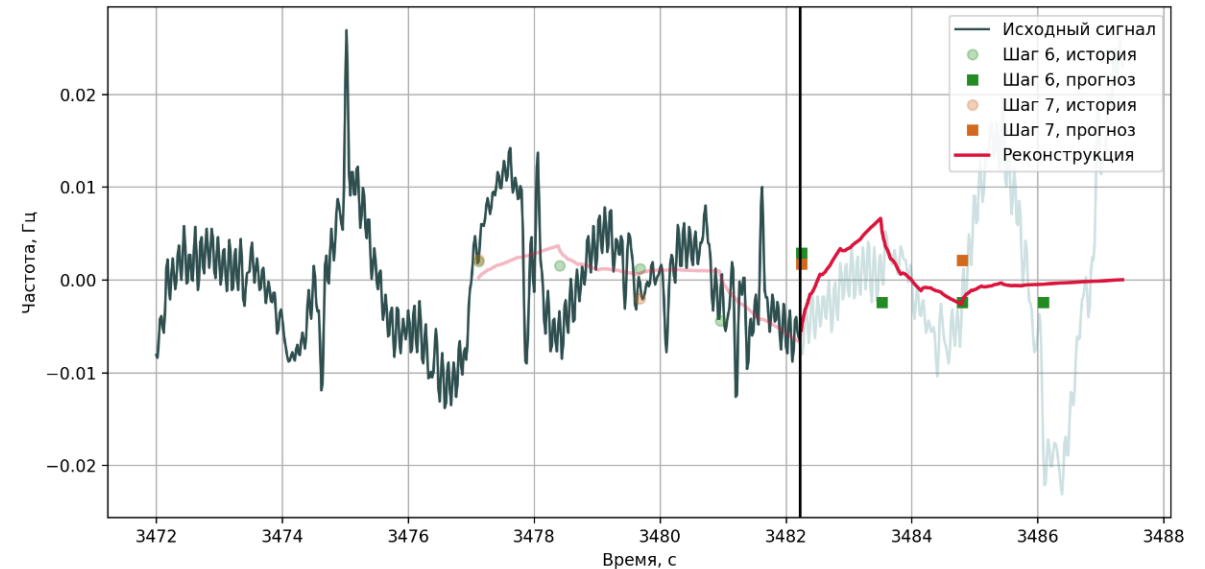


Разложение, прогнозирование, реконструкция



- $a_2 \rightarrow \tilde{a}_2, a_3 \rightarrow \tilde{a}_3$ (прогнозирование)
- $\tilde{a}_2 \rightarrow \check{a}_3, \check{d}_3$ (разложение)
- $\tilde{a}_3, \check{a}_3 \rightarrow \hat{a}_3$ (согласование)

Реконструкция выполняется так, чтобы разложение $\tilde{x}(n)$ приводило к $\check{a}_3$ и $\check{d}_3$.



Заключение

- Применение набора доступных методов к набору доступных данных и констатация того, что одни методы оказались лучше, чем другие с точки зрения выбранных метрик, представляется методологически недостаточным.
- **Сложный сигнал СВИ, отражая в деталях работу большой ЭС, сочетает в себе порядок и хаос.** Важно найти структуры, наиболее пригодные к прогнозированию.
- **Разработана процедура**, включающая разложение сигнала на аппроксимирующие и детализирующие компоненты, оценку КС-энтропии (прогнозируемости), выборочное прогнозирование и реконструкцию.
- **Показана корреляция** энтропии и точности прогноза для аппроксимирующих последовательностей на разных масштабах разложения.

Попов Александр Игоревич
a.popov@ens.ru

Сложность, структура и хаос

00b1н2р3а4м0 1v263r4ы0k1л2е3а4а0 1d2н3с4е0h1т2s3о4у0р1т2о3т4п0w1л2р3и4q0в1w2а3с4,0s1 2v3н4b
0e1н2 3b4x0k1о2q3л4k0о1л2д3м4н0л1а2g3,4x0 1h2н3е4е0р1 2d3г4и0о1z2в3с4о0b1р2м3л4л0и1b2в3л4а0z1

