

РОБАСТНЫЕ ИНТЕРВАЛЬНЫЕ ПРОГНОЗНЫЕ МОДЕЛИ В СТРУКТУРНО-СЛОЖНЫХ СИСТЕМАХ

Червяков А.А.

Федеральное казначейство Министерства финансов Российской Федерации, Москва, Россия
МИРЭА – Российский технологический университет, Москва, Россия
achervyakov@roskazna.ru

Аннотация. Системы поддержки принятия решений имеют информационно-технологическое обеспечение, способное аккумулировать большое количество наблюдаемых и управляемых параметров объектов и процессов, объединенных технической информационной инфраструктурой и вычислительными комплексами. В работе обсуждается подход к разработке робастных интервальных моделей в больших структурно-сложных системах федерального уровня.

Ключевые слова: системы поддержки принятия решений (СППР), интервальный прогноз, робастные модели, снижение размерности, архитектура СППР.

Введение

Основой принятия решений и управления в системах поддержки принятия решений (СППР) являются прогнозные значения, полученные на основе математических моделей. Важным классом задач является динамическое распределение ресурсов на основе прогнозных значений, это, например: балансировка вычислительных ресурсов; распределение трафика в системах массового обслуживания; распределение каналов связи; распределение финансовых ресурсов. Одной из особенностей данного класса задач является важность точности прогноза с точки зрения достоверности и надежности границ прогнозной оценки. Так, например, если реальное значение будет выше верхней оценки прогноза, то ресурса зарезервируется недостаточно и система не сможет функционировать. Таким образом, построение моделей, обеспечивающих построение надежных оценок прогнозируемых значений имеет важное значение в системах принятия решений.

В настоящее время развиваются методы, основанные на так называемых прогнозных интервалах (далее — ПИ) (англ. — Prediction Intervals), способные обеспечивать заданные требования к качеству прогноза временных рядов (Quan H.; Srinivasan D.; Khosravi A.; Taormina R.; Chau, K.W. и др. [1-3]). Применение ПИ к большим структурно-сложным системам требует проведение теоретических и экспериментальных исследований. Одним из ключевых методов является непараметрический метод нахождения оценок нижней и верхней границ ПИ, которые получил название LUBE (Lower Upper Bound Estimation, Khosravi A.; Nahavandi S.; Creighton D.; Atiya A.F., 2011 [3]). Согласно методу LUBE ставится многокритериальная задача с противоречивыми критериями качества, компромисс между которыми определяется фронтом множества Парето, который ищется численными методами.

ПИ, как подход к моделированию, привлек большое внимание, поскольку такой подход является интуитивно понятным [4]. Например, метеопрогноз представляется, традиционно, как диапазон прогнозируемых температур на каждый момент времени или по суткам. Методы широко применяются как для погодных данных, прогнозирования ветра [5], в энергетике [6] и в других прикладных областях [7-9]. Общность применения можно охарактеризовать как: (1) задача обладает ярко выраженными трендами, сезонными компонентами; (2) является портфелем нескольких сигналов; (3) к прогнозам, предъявляются требования к достоверной надежности верхней или нижней границы.

В работе, на примере построения СППР для таргетирования остатка на едином казначейском счете (совокупности остатков на счетах всех клиентов, ЕКС ФК) [10], рассматривается комплексный подход к разработке методов построения робастных интервальных прогнозных моделей в СППР структурно-сложных систем, включающий в себя:

- метод построения по временным рядам робастных моделей в форме прогнозных интервалов, обеспечивающих заданные требования к качеству; отличие метода состоит в разработанном подходе, основанном на объединении прогнозных интервалов [9];
- методику построения статистических моделей группы временных рядов для задачи распределения однотипных ресурсов, основанной на гипотезе о рассмотрении каждого временного ряда из группы как реализации одного случайного процесса; предложенная гипотеза увеличивает достоверность, сужает интервальный прогноз каждого ряда и снижает размерность структурно-сложной системы;

- состав, структуру и архитектуру информационно-аналитического обеспечения СППР, способной анализировать и осуществлять построение робастных интервальных прогнозов на заданный горизонт для большого количество временных рядов.

1. Метод построения робастных прогнозных интервалов

Существует два подхода к моделированию динамических систем – прямые методы, основанные на физических свойствах явлений или систем; обратные – основанные на решении некорректной задачи, называемой обратной задачей нелинейной динамики. В качестве нового подхода, объединяющего преимущества прямых и обратных методов, предлагается использовать модели, нечувствительные к вариациям собственно временного ряда, заменяя точечное прогнозирование диапазоном, то есть получая модель, которая строит ПИ.

Подход построения робастных ПИ не является новым, доверительные интервалы в регрессионных моделях представляют тоже в некотором смысле прогнозный интервал, но в таких моделях интервал не является объектом манипуляций. Иными словами, доверительный интервал — вычисляемая характеристика многих моделей, который необходимо учитывать, но не использовать полученный интервал для изменения модели. В настоящей работе, в соответствии с имеющимися предпосылками, предлагается решать задачу моделирования на основе манипуляций прогнозными интервалами.

Предложена операция объединения интервалов. Пусть имеется две или более приближительных модели, решение которые есть графики с интервальными решениями. Тогда робастная интервальная модель представляет собой объединение интервалов в каждой точке моделируемого или прогнозируемого временного ряда.

Определение 1. Пусть задан временной ряд $X = x(0), x(1), \dots, x_n, \dots$, пусть существует эволюционная функция $y(n) = f(x(n))$, (n – дискретное время), при этом $x(n+1)$ принадлежит интервалу $y \pm \varepsilon: x(n+1) \in [y - \varepsilon; y + \varepsilon]$. Тогда при $n \rightarrow \infty$, если для $\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0$, такое что для всякого состояния $\forall x(n)$ из $\|x(0) - y(0)\| < \delta$ следует $\|x(n) - y(n)\| < \varepsilon$.

Такую модель будем называть робастной интервальной моделью.

Определение 2. Для интервальных моделей f_1, f_2 задана операция объединения моделей, при этом исходный интервал в каждой точке $x(n)$ определяется как $[\min_{i=1,2} \varepsilon_i, \max_{i=1,2} \varepsilon_i]$.

Лемма 1. Пусть $f_1(x)$ – не удовлетворяет условию определения 1. Тогда может существовать функция $f_2(x)$, являющаяся робастной интервальной моделью исходной системы. Если рассматривать операцию объединения моделей, то результирующая модель тоже будет являться робастной. Доказательство очевидно.

На рис. 1 схематично показан процесс получения интервальных моделей для отдельных объектов структурно-сложных систем.

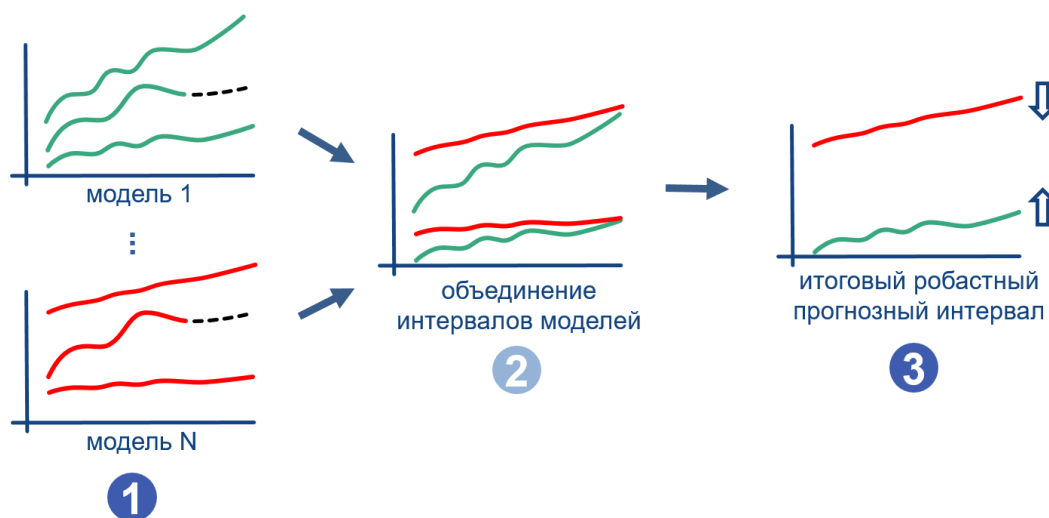


Рис. 1. Метод построения робастные интервальных прогнозных моделей

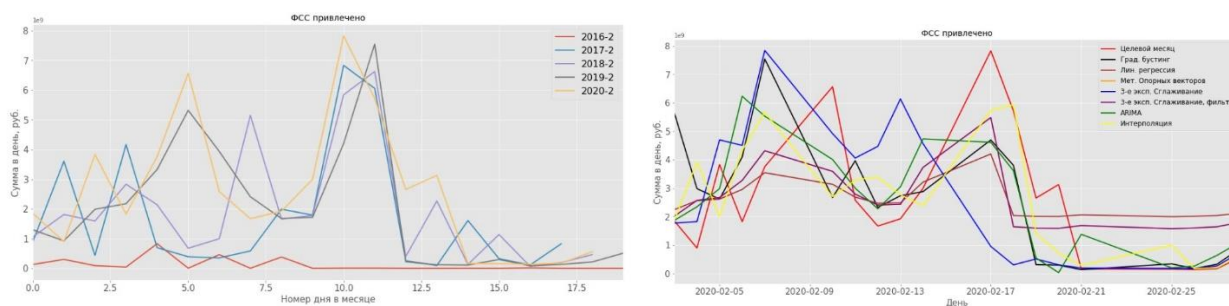
Введенные понятия позволяют сформировать метод построения по временным рядам робастных интервальных моделей в форме прогнозных интервалов, обеспечивающих требования к качеству управления в структурно-сложных системах. Метод состоит из 3 шагов.

Шаг 1. Построение нескольких динамических моделей временного ряда с получением интервальных прогнозов.

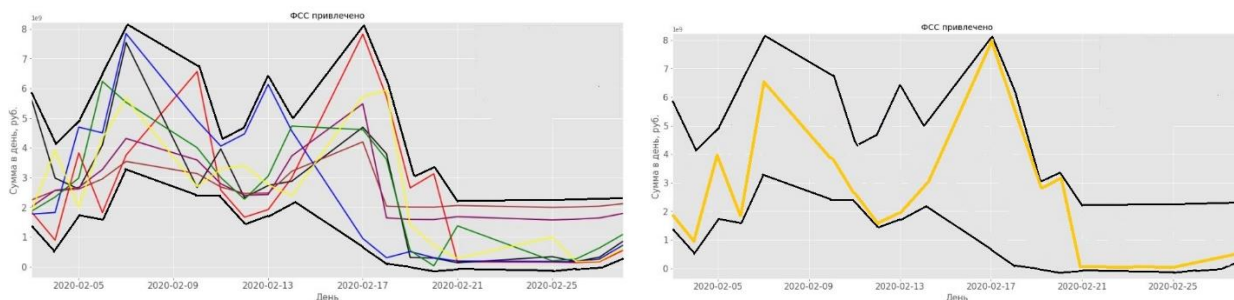
Шаг 2. Объединение интервалов моделей.

Шаг 3. Если требуется, то сужение интервалов на основе заданных критериев оценки качества интервальных решений [1-3].

В качестве примера рассмотрены временные ряды, отражающие расходы клиентов Федерального казначейства (ФК). Ряды представляют собой значения, которые соответствуют запросу денежных ресурсов необходимым клиентам для бесперебойного осуществления своих обязательств. Таким образом, задача состоит в построении прогнозной модели, обеспечивающей гарантированное значение ресурсов, которые могут потребоваться клиентам в следующих диапазонах времени. Рассмотрим применение метода прогнозирования на примере остатков на счетах Фонда социального страхования (ФСС). Для моделирования взяты исторические данные за период 2015–2020 гг. Было применено 6 разных моделей.



а) плановые значения с февраля 2016 по февраль 2020 б) графики прогнозных значений моделей



в) объединение прогнозных интервалов (черные линии) г) итоговый робастный прогнозный интервал для значений февраля 2020

Рис. 2. Робастный интервальный прогноз остатков ФСС

Как видно из приведенных графиков все модели дают в среднем одну величину ошибки, но тем не менее, в совокупности итоговый робастный прогнозный интервал включает все точки прогнозируемых остатков ФСС за февраль 2020.

2. Снижение размерности структурно сложной системы

Рассмотрим задачу распределения ресурсов в следующей постановке. Пусть несколько объектов расходует единый ограниченный ресурс, необходимо потребление этого ресурса каждым объектом зарезервировать в совокупности для всей системы на основе прогноза. При этом для каждого объекта, во временных рядах, описывающих потребление ресурса, выражена сезонная компонента. Примерами таких задач являются: распределение вычислительных ресурсов между виртуальными машинами, расход электроэнергии, расход бюджетных средств на примере ЕКС ФК. В этих задачах есть типовые характеристики – существуют пиковые нагрузки в определенное время суток, дня недели или месяца, в остальное время существуют спады, рост и т.д., но резервирование должно быть гарантированно надежно и включать пиковые или иные общие тенденции. Для однотипных задач (например, расходование ресурсами одинаковых вычислительных сервисов или выделение трафика

сети для одинаковых компаний) вид временного ряда похожий (всплеск расходов ресурсов в понедельник утром, после обеденное время, падение потребности ресурсов в вечернее время). А под структурно-сложной системой будем понимать совокупность однотипных или близких к однотипным объектам, описываемых своими робастными интервальными моделями, о которых говорили в первом разделе.

Тогда построение робастной интервальной модели для структурно сложной системы, основанное на операции объединения интервалов, будет аналогично описанному выше, только объединению будут подлежать не разные интервалы одного объекта, а уже найденные робастные интервалы объектов.

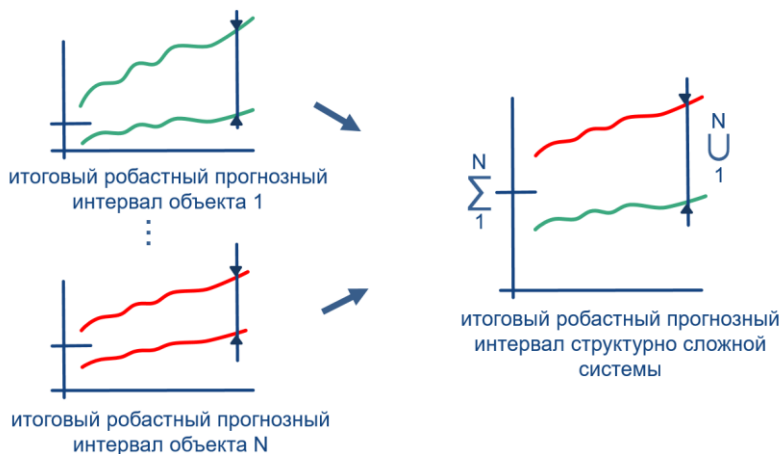


Рис. 3. Построение робастного прогнозного интервала структурно-сложной системы

Имея исторические данные за 2015–2020 гг., эмулируем поведение системы в 2020 г. На базе 20 наиболее крупных клиентов ФК. Последовательно добавляя в систему данные за следующий месяц, будем экстраполировать увеличивающиеся ряды с помощью скользящих средних, агрегировать в итоговый прогноз с интервалом в 99%. На выходе (рис. 4) получим соответственно 11 пар верхних и нижних интервалов в виде набора отдельных рядов с размерностью от 11 до 1 точки.

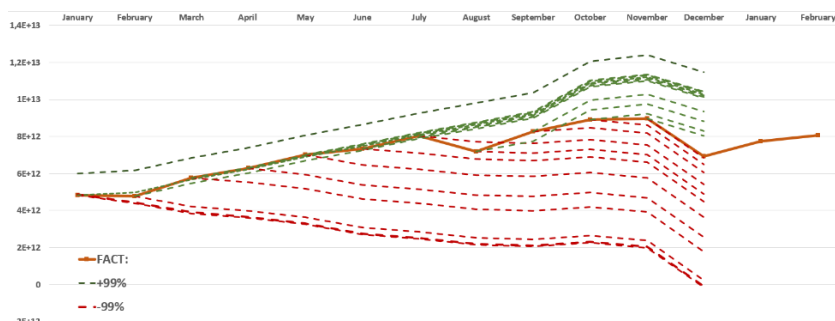


Рис. 4. Построение робастной интервальной модели на 2020 год

Объединяя соседние первые точки верхних и нижних рядов доверительных интервалов, получим набор интервалов (коридор) в котором должны находиться фактические значения итогового прогноза (рис. 5).

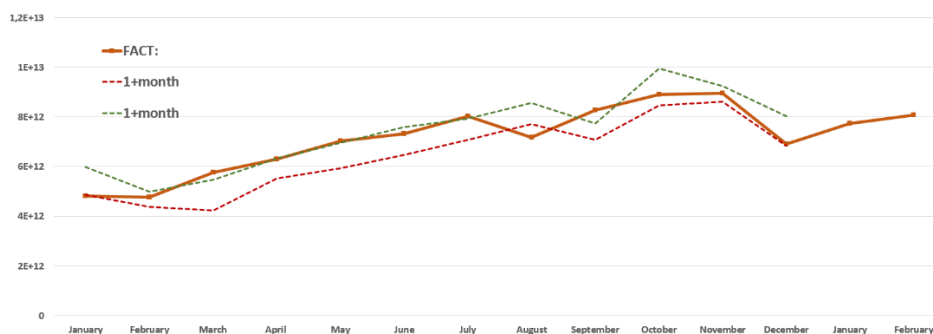


Рис. 5. Сравнение робастной интервальной модели с фактическими данными

Как видно из рис. 5, несмотря на всю грубость модели, фактические данные остатка на ЕКС в ходе эксперимента вышли бы за нижнюю границу интервала лишь однажды в августе 2020 года.

В реальной СППР ФК для таргетирования ЕКС обрабатывается более 40 000 временных рядов отдельных объектов – остатков клиентов. Для построения робастных прогнозных интервалов большинства из них требуется участие эксперта, т.к. в этих эконометрических рядах зачастую встречаются структурные сдвиги и выбросы, которые тяжело идентифицируются моделями. Предложена методика снижения размерности структурно-сложной системы обеспечивающая компромисс между трудоемкостью, скоростью и актуальностью прогнозирования временных рядов. На примере ФК размерность системы снижается до анализа 50 объектов и их прогнозных интервалов.

Сформулирована гипотеза о том, что временные ряды расхода ресурса каждым различным объектом представляют собой различные реализации одного и того же случайного процесса. Как правило, применение теории случайных процессов для моделирования временных рядов имеет несколько существенных проблем, одна из которых: невозможно повторить испытания. Выходом из этой ситуации является базовое предположение о стационарности случайного процесса в широком смысле. При этом еще более сильное предположение, что наблюдаемый временной ряд соответствует среднему процессу (математическому ожиданию в каждой точке). Естественно, в практике другого варианта часто нет, и для коротких временных рядов (посуточный, почасовой, помесечный прогноз) надежный прогноз построить просто математически невозможно, т. к. мало данных, даже если визуально наблюдаются сезонные колебания.

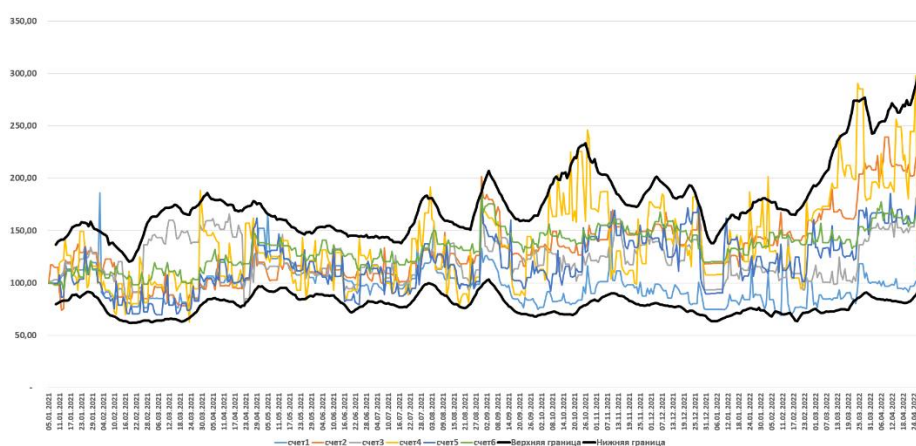


Рис. 6. Реализации одного случайного процесса (реальные остатки на 6 разных счетах)

Каждому входному ряду исходных данных присваивается 2 признака – средняя доля в итоговом прогнозе P (мера «значимости») и коэффициент вариации ряда V (мера «прогнозируемости») [11].

Этап 1. Расчет показателей этих для каждого из рядов.

Результат выполнения этапа можно визуализировать (рис. 7) в координатах (V, P) . Чем выше и правее находится точка, характеризующая ряд, тем он более значимый и такие ряды не будут объединены в классы.

Этап 2. Построение четырех групп объектов, потребляющих ресурсы на основании характеристик рядов.

Разбиение на группы происходит путем разбиения плоскости в координатах (V, P) на четыре квадранта (см. рис. 7), так, чтобы, «значимых» и «прогнозируемых» рядов меньше чем «значимых» и «не прогнозируемых», а «значимых» в целом значительно меньше чем «незначимых».

Далее манипуляции с объектами каждой из четырех групп осуществляются отдельно.

Этап 3. Группировка и предобработка.

3.1 «Незначимые» и «непрогнозируемые» ряды в левом нижнем квадранте агрегируем в один случайный процесс. Получившийся итоговый ряд имеет большую «значимость» и некоторую усредненную «прогнозируемость». Далее работаем с ним как с отдельным случайным процессом – строим робастный прогнозный интервал как для одного объекта структурно сложной системы.

3.2 «Незначимые» и «прогнозируемые» ряды в правом нижнем квадранте агрегируем в один случайный процесс. Получившийся итоговый ряд имеет большую «значимость» и некоторую усредненную «прогнозируемость». Далее работаем с ним как с отдельным случайным процессом, отдельным объектом.

3.3 «Значимые» и «непрогнозируемые» ряды в левом верхнем квадранте требуют уже участия эксперта. Все эти факторы можно учесть в модели в виде экспертного мнения.

Если «убрать» из исходных рядов выбросы и сдвиги и снова провести их оценку по параметрам «значимости» и «прогнозируемости», то большая часть из них переместится в правый верхний квадрант.

3.4 А оставшиеся после 3.3 «значимые» и «непрогнозируемые» ряды можно агрегировать в один случайный процесс, один объект, как в шагах 3.1 и 3.2.

3.5 Со «значимыми» и «прогнозируемыми» рядами в верхнем правом квадранте работаем как с отдельными объектами структурно-сложной системы.

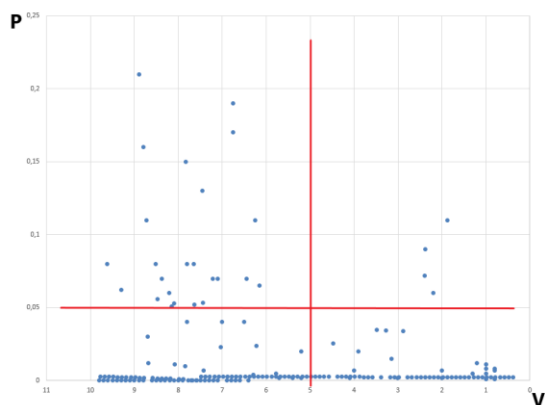


Рис. 7. Все ряды, поступившие в ССПР ФК

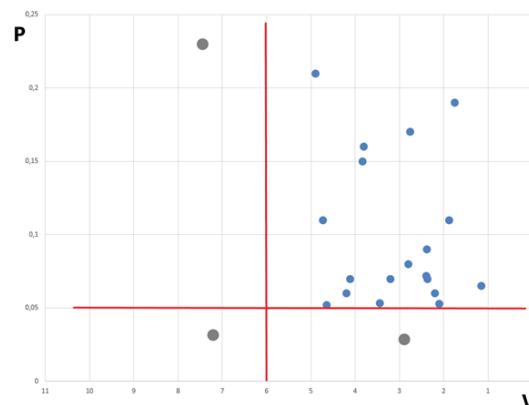


Рис. 8. Итоговые входные ряды для прогнозирования моделями

Приведенная методика позволяет свести анализ 40 000 временных рядов ССПР ФК (рис. 8) к анализу менее чем 50 рядов для обеспечения адекватной трудоемкости и актуальности прогнозирования.

3. Архитектура ССПР

Для реализации промышленной ССПР для решения практических задач робастного прогнозирования динамики действительно больших структурно-сложных систем были также разработаны требования и структура информационно-аналитического обеспечения (ИАО) ССПР.

Требования к ИАО ССПР:

- Необходимо аккумулировать все ИТ-системы в единый контур, для обеспечения возможности доступного оперирования их транзакционными данными.
- Необходимо построить инфраструктуру работы с большими данными, включающую в себя:
 - о интеграционные взаимодействия с открытыми и специализированными источниками данных;
 - о общие элементы, такие как озеро данных, лаборатория моделей машинного обучения, инструменты управления качеством данных на всех этапах их жизненного цикла;
 - о инструменты вывода и визуализации данных – дашборды, мобильные приложения, витрины данных.
- Определить место ССПР в архитектуре бизнес-процессов и в архитектурном ИТ ландшафте.

Логическая структура ССПР ФК представляет собой программно-математический комплекс, формирующий аналитические материалы для обеспечения поддержки принятия решений. На рис. 9 показан порядок обработки данных. Представленный подход обеспечивает требуемую робастность построенных моделей.

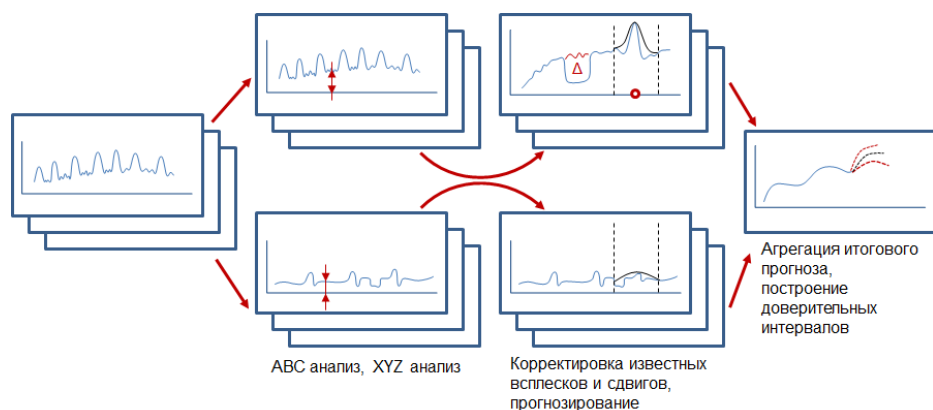


Рис. 9. Структура подготовки аналитических материалов для принятия решений

Структура ИАО СППР при внедрении в ФК приведена на рис. 10.

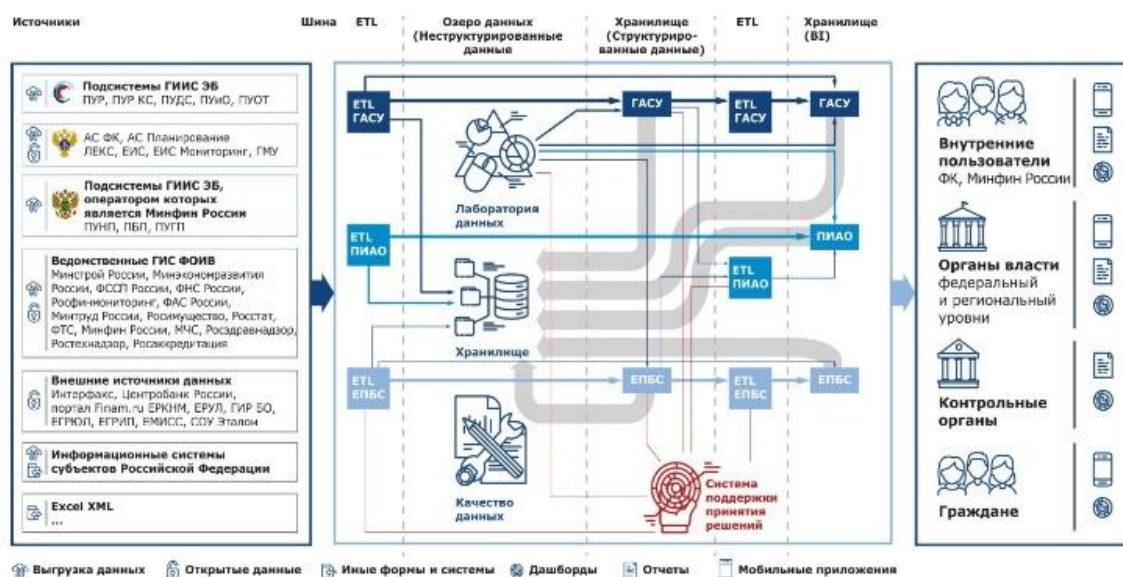


Рис. 10. ИТ-инфраструктура информационно-аналитического обеспечения СППР

4. Заключение

В настоящей работе, на примере СППР Федерального казначейства, рассмотрен комплексный подход к разработке методов построения робастных интервальных моделей и реализации инструментов прогнозирования в СППР в больших структурно-сложных системах. Приведенные результаты демонстрируют эффективность разработанных методов и перспективность их использования в информационно-аналитическом обеспечении крупномасштабных цифровых платформ и систем.

Литература

1. *Quan H., Srinivasan D., Khosravi A.* 'Short-term load and wind power forecasting using neural network-based prediction intervals // IEEE Trans. Neural Netw. Learn. Syst. — 2014. — Vol., 25, N. 2. — P. 303–315.
2. *Khosravi A., Nahavandi S., Creighton D., Atiya A.F.* 'Lower upper bound estimation method for construction of neural network-based prediction intervals // IEEE Trans. Neural Netw. Learn. Syst. — 2011. — Vol. 22, N. 3. — P. 337–346.
3. *Taormina R., Chau K. W.* ANN-based interval forecasting of streamflow discharges using the LUBE method and MOFIPS // Engineering Applications of Artificial Intelligence. — 2015. — Vol. 45. — P. 429-440.
4. *Nikulchev E., Chervyakov A.* Development of Trading Strategies Using Time Series Based on Robust Interval Forecasts // Computation. — 2023. — Vol. 116 N. 5. — P. 99.
5. *Zhou M., Wang B., Guo S., Watada J.* Multi-objective prediction intervals for wind power forecast based on deep neural networks // Information Sciences. — 2021. — Vol. 550. — P. 207-220.

6. *Cao Z., Wan C., Zhang Z., Li F., Song Y.* Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting // *IEEE Transactions on Power Systems.* — 2019. — Vol.35, N. 3. — P. 1881-1897
7. *Червяков А.А., Никульчев Е.В.* Робастное интервальное прогнозирование временных рядов // *International Journal of Open Information Technologies.* — 2023. — Т. 11, №. 4. — С. 122-128.
8. *Никульчев Е.В., Червяков А.А.* Построение робастных интервальных моделей прогнозирования динамики структурно-сложной системы // *Труды НГТУ им. Р.Е. Алексеева.* — 2023. — № 1. — С. 33-41.
9. *Nikulchev E., Chervyakov A.* Prediction Intervals: A Geometric View // *Symmetry.* — 2023. — Vol. 15. N. 4. — P. 781.
10. *Албычев А.С., Червяков А.А., Никульчев Е.В.* Структура системы поддержки принятия решений при таргетировании остатка денежных средств на едином казначейском счете Федерального казначейства // *International Journal of Open Information Technologies.* 2022. Т. 10. №. 12. С. 116-122
11. *Червяков А.А.* Предобработка временных рядов в системе поддержки принятия решений управления портфелем // *Информационный обмен в междисциплинарных исследованиях. Сб. тр. II Всер. науч.-практ. конф. с межд. участием.* — Рязань: Академия ФСИН России, 2023. С. 199–203 (РИНЦ)