WLOG SOLUTIONS

Case Studies in advanced analytics with R

Wit Jakuczun, WLOG Solutions

2017-05-17

Data Community

SQL Day 2017

Atos

Profisee®

Zerto

SILVER SPONSORS

PROMISE GRUPA APN PROMISE

ELITMIND

ITMAGINATION Imagine it. Done.

THEOBALD SOFTWARE

BRONZE SPONSORS

it KONTRAKT®

JCommerce

IT Krajowy Rejestr Długów

KSIĄŻKI IT Z CAŁEGO ŚWIATA NOVATECH www.novatech.com.pl

PEPCO

dbWatch DATABASE CONTROL

querona

Data Scientist as a Service

UBS

VOLVO Volvo Group IT

STRATEGIC PARTNER

Microsoft

~1995

R project has been started

2000

Version 1.0 released

2006

I delivered my first project using R

(~100 models for gas demand forecasting)

2007

Revolution Analytics was founded.

2015

Microsoft acquired Revolution Analytics.

2017

R is (still) one of the most popular language for data science.

May 2017	May 2016	Change	Programming Language	Ratings	Change
1	1		Java	14.639%	-6.32%
2	2		C	7.002%	-6.22%
3	3		C++	4.751%	-1.95%
4	5	↑	Python	3.548%	-0.24%
5	4	↓	C#	3.457%	-1.02%
6	10	⇈	Visual Basic .NET	3.391%	+1.07%
7	7		JavaScript	3.071%	+0.73%
8	12	⇈	Assembly language	2.859%	+0.98%
9	6	↓	PHP	2.693%	-0.30%
10	9	↓	Perl	2.602%	+0.28%
11	8	↓	Ruby	2.429%	+0.09%
12	13	↑	Visual Basic	2.347%	+0.52%
13	15	↑	Swift	2.274%	+0.68%
14	16	↑	R	2.192%	+0.86%
15	14	↓	Objective-C	2.101%	+0.56%
16	42	⇈	Go	2.080%	+1.83%
17	18	↑	MATLAB	2.063%	+0.78%
18	11	⇊	Delphi/Object Pascal	2.038%	+0.03%
19	19		PL/SQL	1.676%	+0.47%
20	22	↑	Scratch	1.668%	+0.74%

Tiobe Index, 2017

What is so special about R?

Low barriers to start using it	Huge base of community Packages (CRAN/Github)	Easy to build a team knowing R

```
problem <- model %>%
    mzn_problem() %>%
    mzn_data_file(system.file("extdata", "examples", "sbprgeost.dzn",
                              package = "minizincR")) %>%
    mzn_out_expr(var_name = "1")  %>%
    mzn_out_expr(var_name = "roff",
                 idxs = c("ROFF", "1..2")) %>%
    mzn_out_expr(var_name = "kind",
                 idxs = c("BLOCK")) %>%
    mzn_out_expr(var_name = "coord",
                 idxs = c("BLOCK", "1..2"))

solver <- fzn_gecode(args = NULL)

solution <- solver %>%
              solve(problem)

print(solution)
```

```
gg <- local({
 k <- function(y) f(y)
 f <- function(x) if(x) x*k(x-1) else 1
})
gg(3)
```

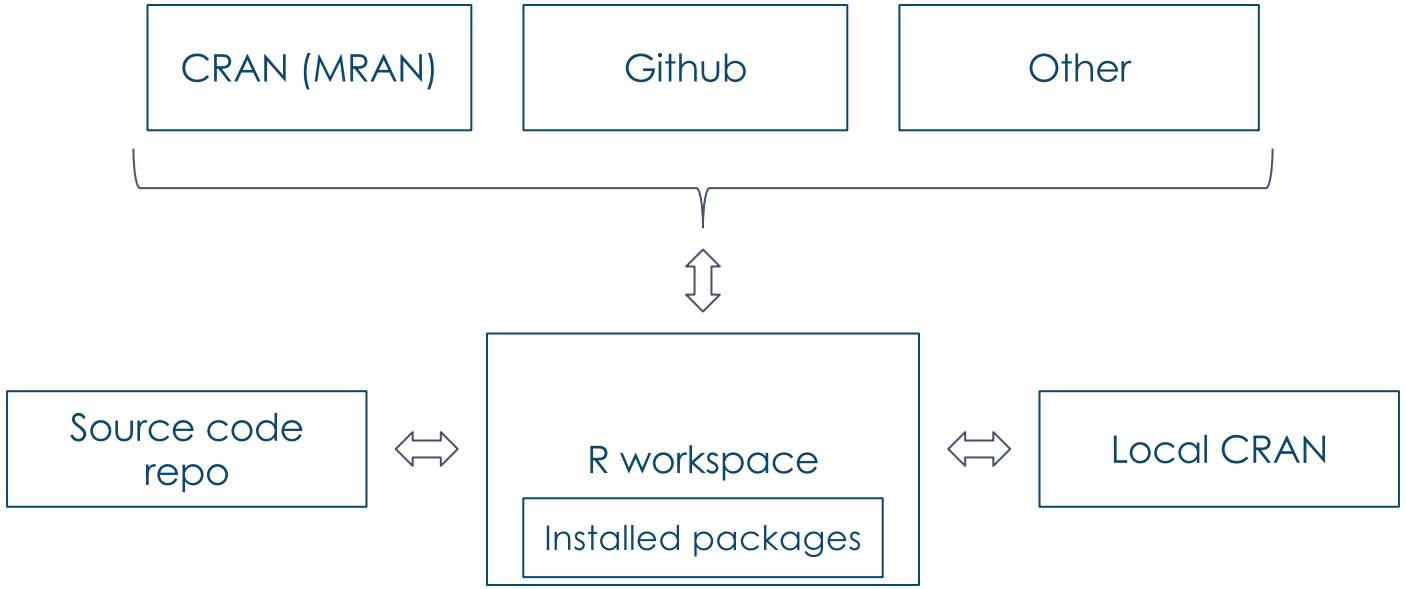
However, **out-of-the-box R does not** provide many essential features required in **large-scale production deployment**!

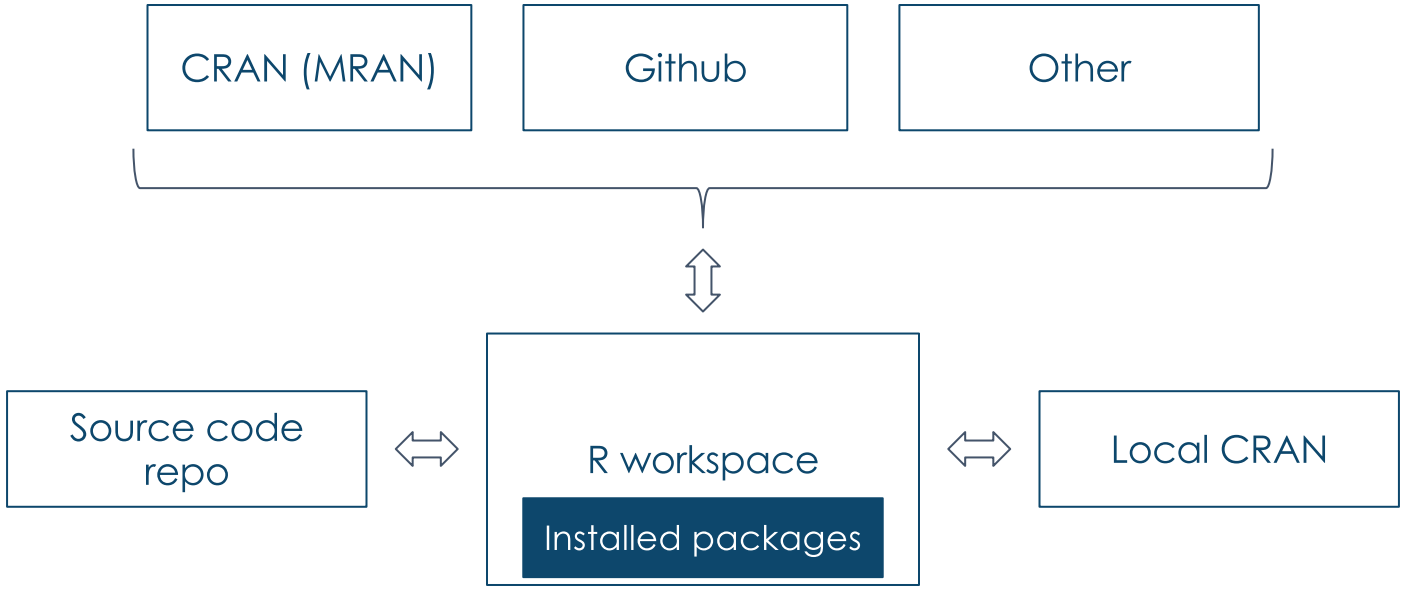
"If you are **using R** and you think **you're in hell**, this is a map for you."

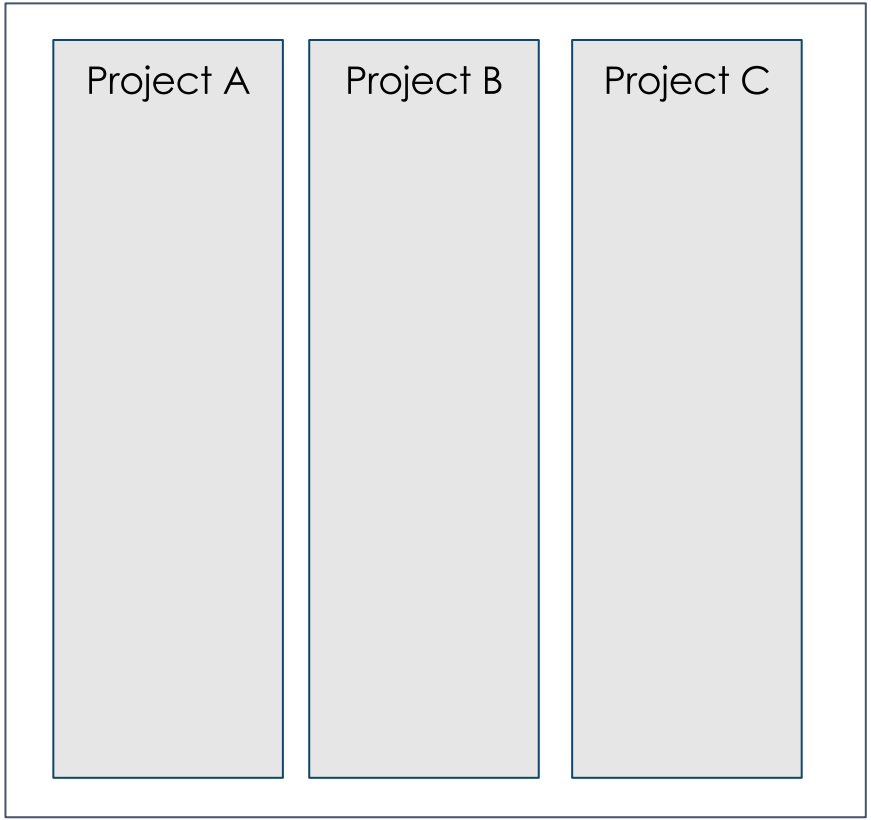
Patrick Burns, "R Inferno", 2011

WLOG R Suite™

How do we deliver R based solutions?


CRAN (MRAN)
Github
Other
Source code repo
R workspace
Installed packages
Local CRAN


CRAN (MRAN)
Github
Other
Source code repo
R workspace
Installed packages
Local CRAN



Production server

- Project independence
- Configuration and version management
- Version control
- Continuous integration&deployment

Configuration management

Source code versioning

WLOG R Suite™

Continuous integration & deployment

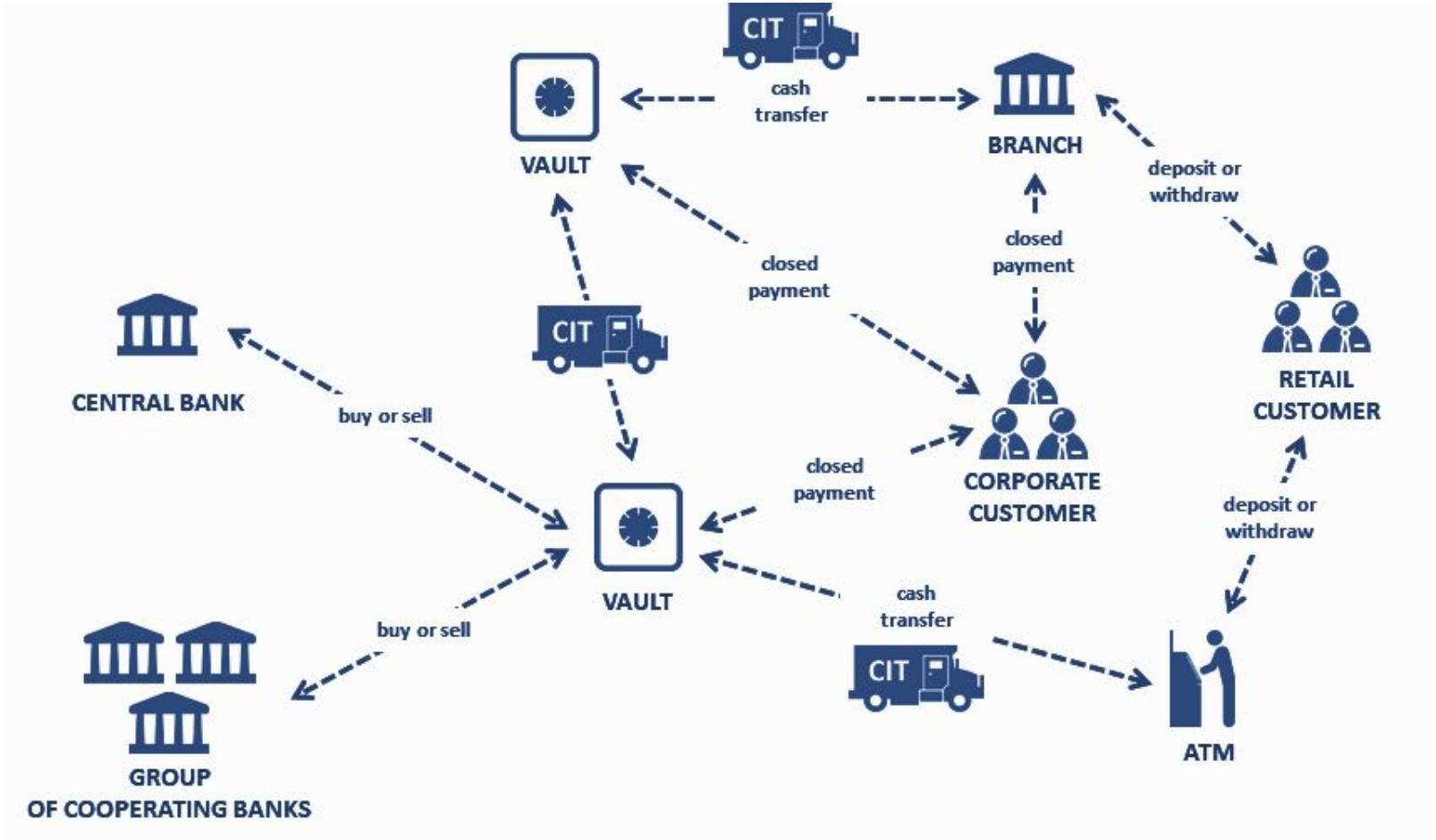
Seasoned R packages

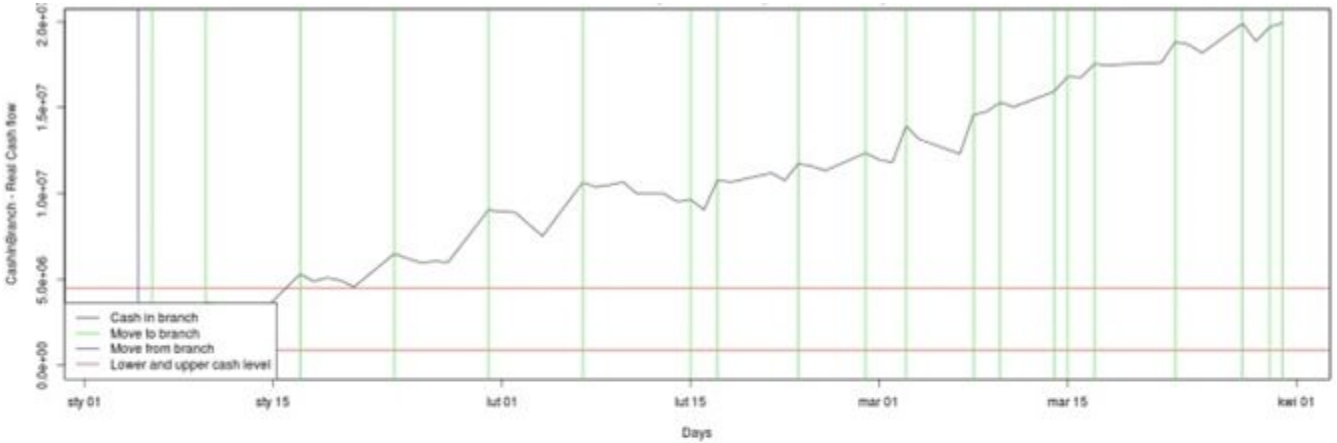
Case Studies in advanced analytics with R

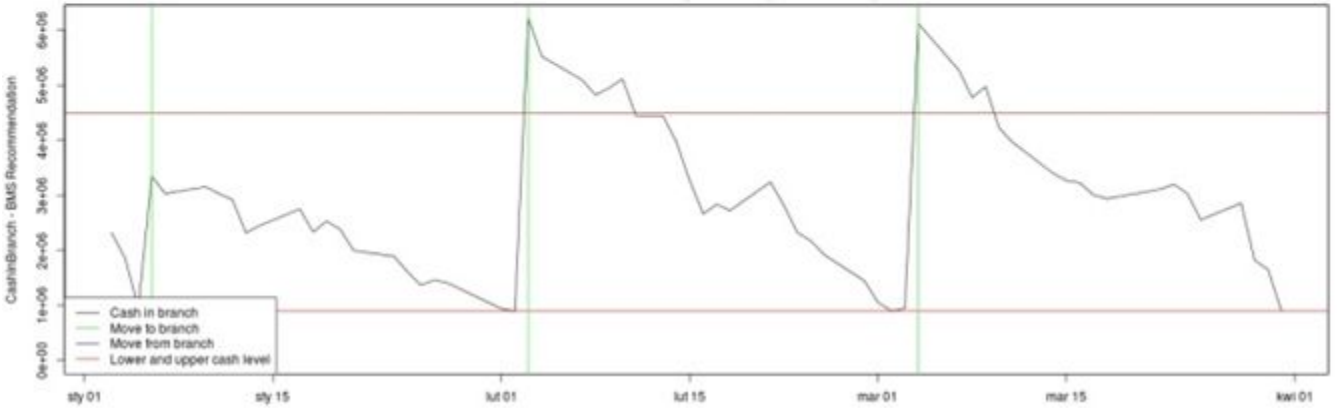
Cash optimisation in Deutsche Bank

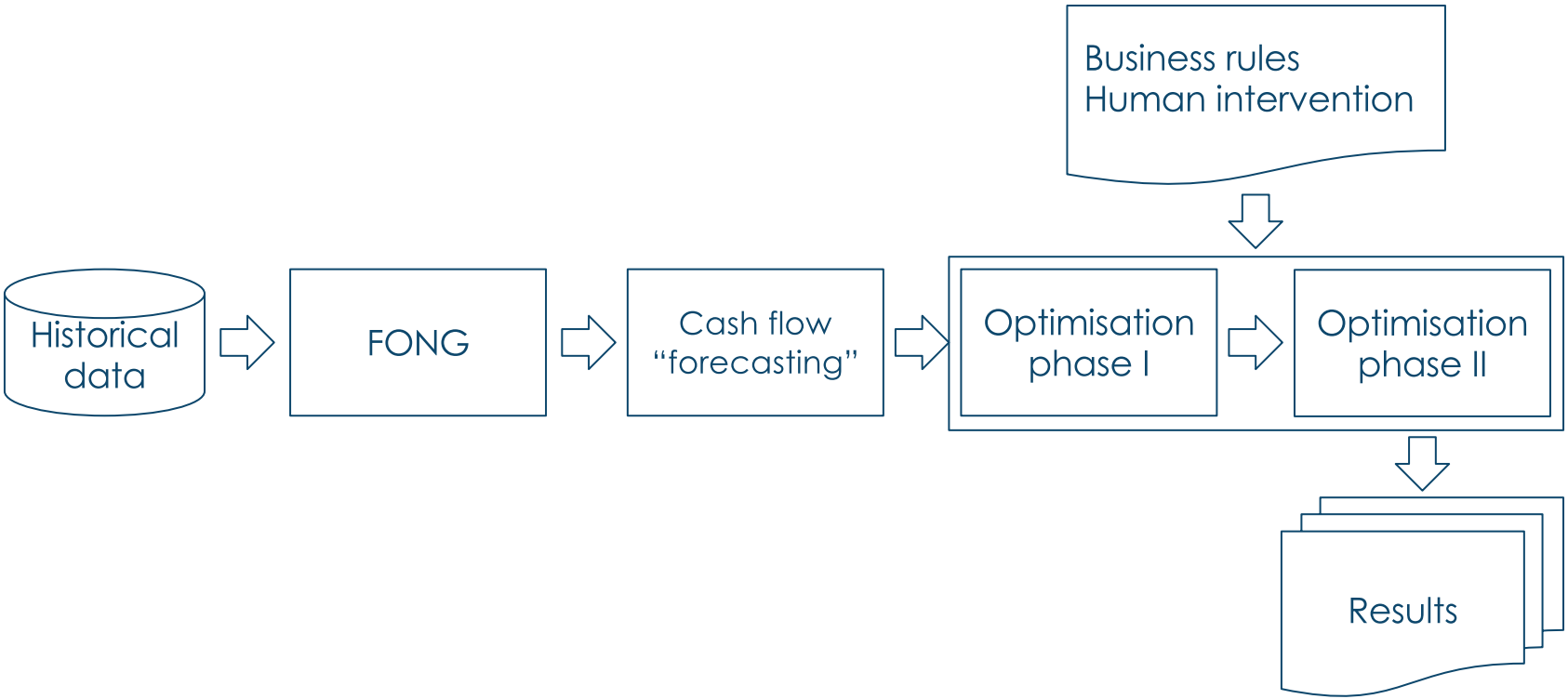
How much, how and when should I transfer between branches and vaults?

Minimizing costs subject to (complex) business rules


CIT
cash
transfer
VAULT
BRANCH
deposit or
withdraw
closed
payment
closed
payment
CIT
CENTRAL BANK
buy or sell
RETAIL
CUSTOMER
closed
payment
CORPORATE
CUSTOMER
deposit or
withdraw
VAULT
cash
transfer
buy or sell
CIT
ATM
GROUP
OF COOPERATING BANKS


CashInBranch - Real Cash flow
0.0e+00
5.0e+06
1.0e+07
1.5e+07
2.0e+07
Cash in branch
Move to branch
Move from branch
Lower and upper cash level
sty 01
sty 15
lut 01
lut 15
mar 01
mar 15
kwi 01
Days


CashInBranch - BMS Recommendation
0e+00
1e+06
2e+06
3e+06
4e+06
5e+06
6e+06
Cash in branch
Move to branch
Move from branch
Lower and upper cash level
sty 01
sty 15
lut 01
lut 15
mar 01
mar 15
kwi 01


Business rules
Human intervention
Historical data
FONG
Cash flow "forecasting"
Optimisation phase I
Optimisation phase II
Results

Summary

- Our first large R project
- Around 20K LOC
- Webservice in C#

- **R core is very stable**
- Logic in packages
- No *print* allowed
- Project folder structure
- Configuration management
- (Almost) No REPL

Midterm model for energy prices forecasting

What can be the energy price next year?

Alternative (to the forward market) spot price forecast.

Historical prices
Wind speed
Temperature
Power availability

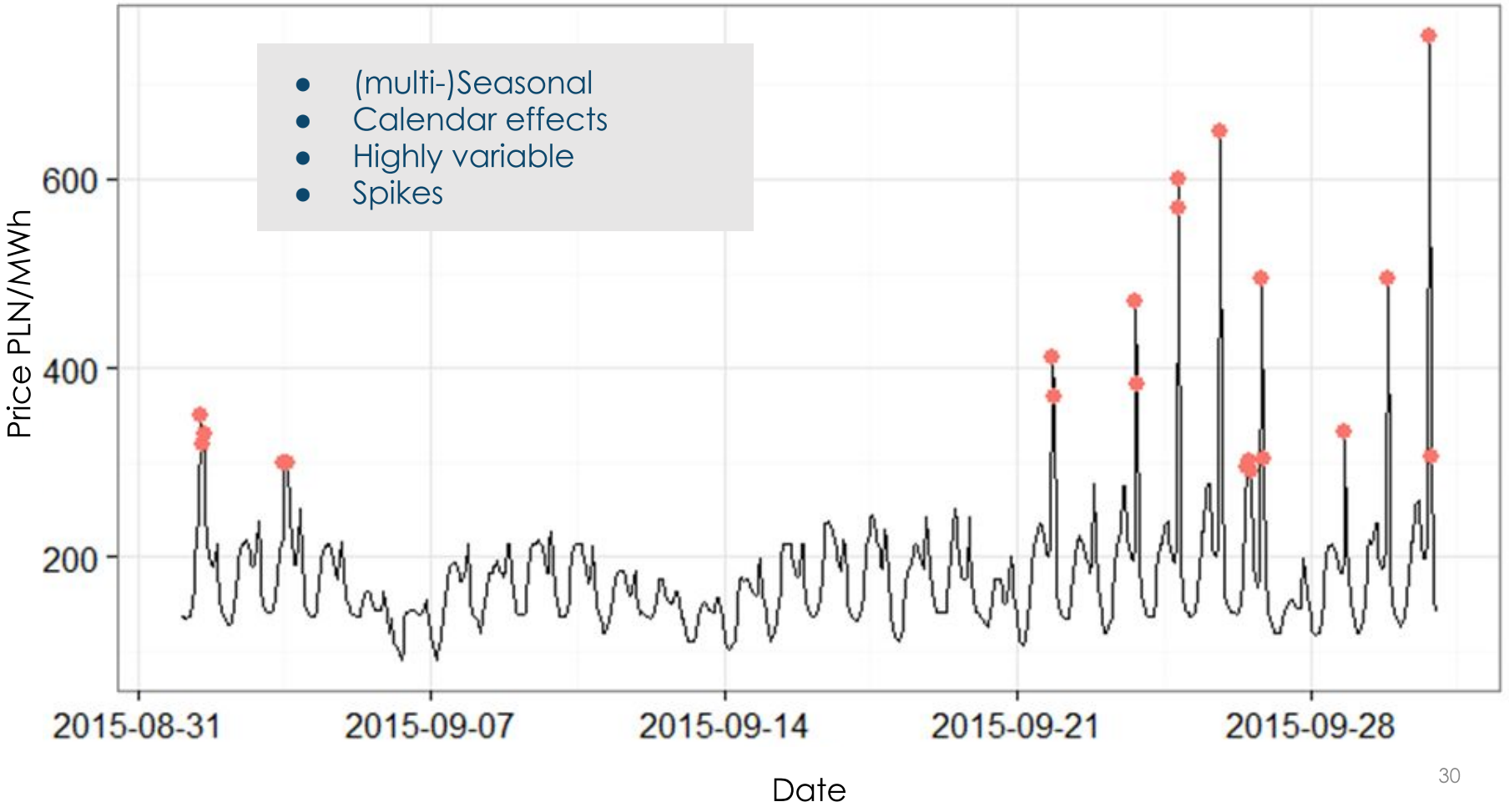
Weather forecast?

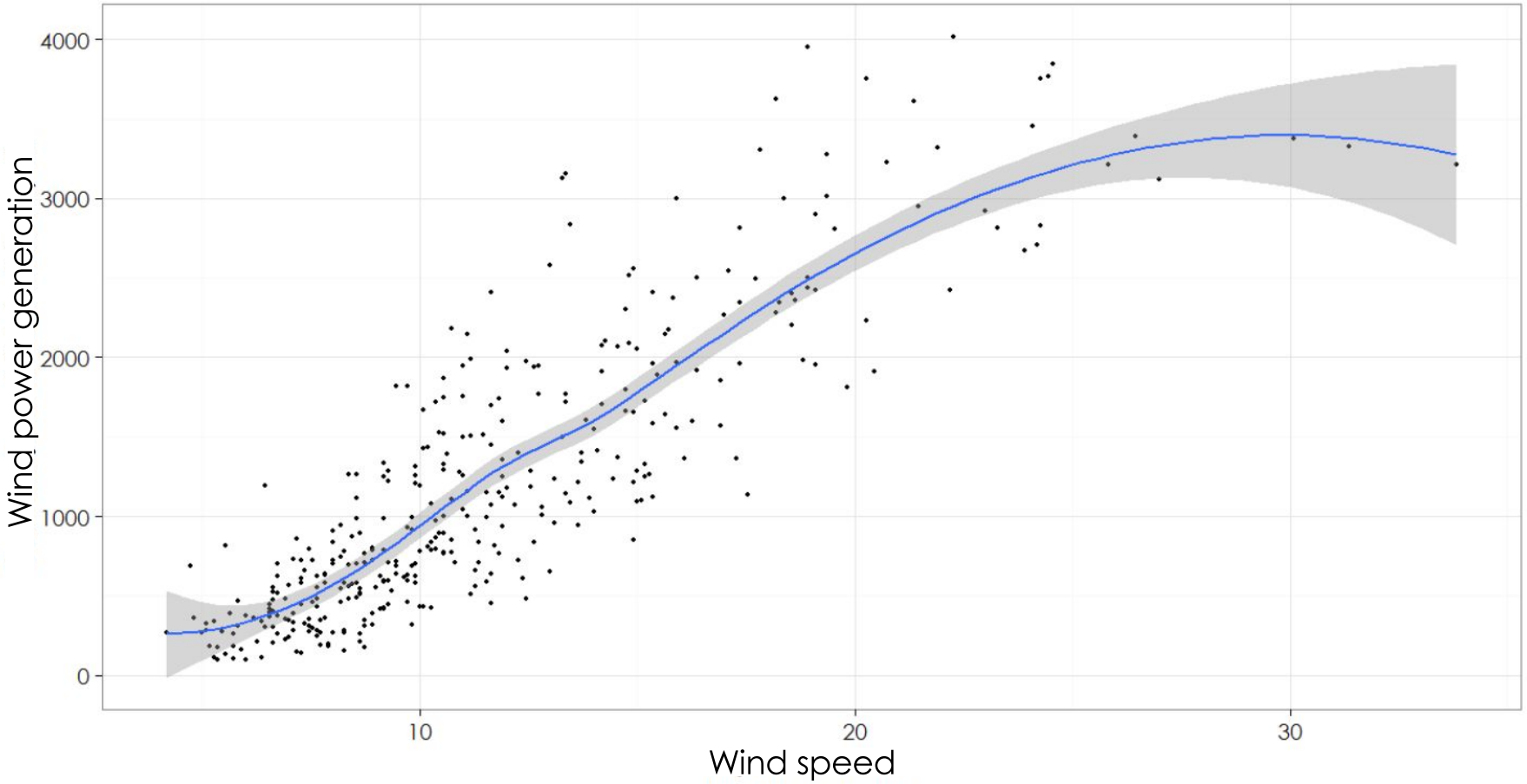
Weather forecast?
Power availability

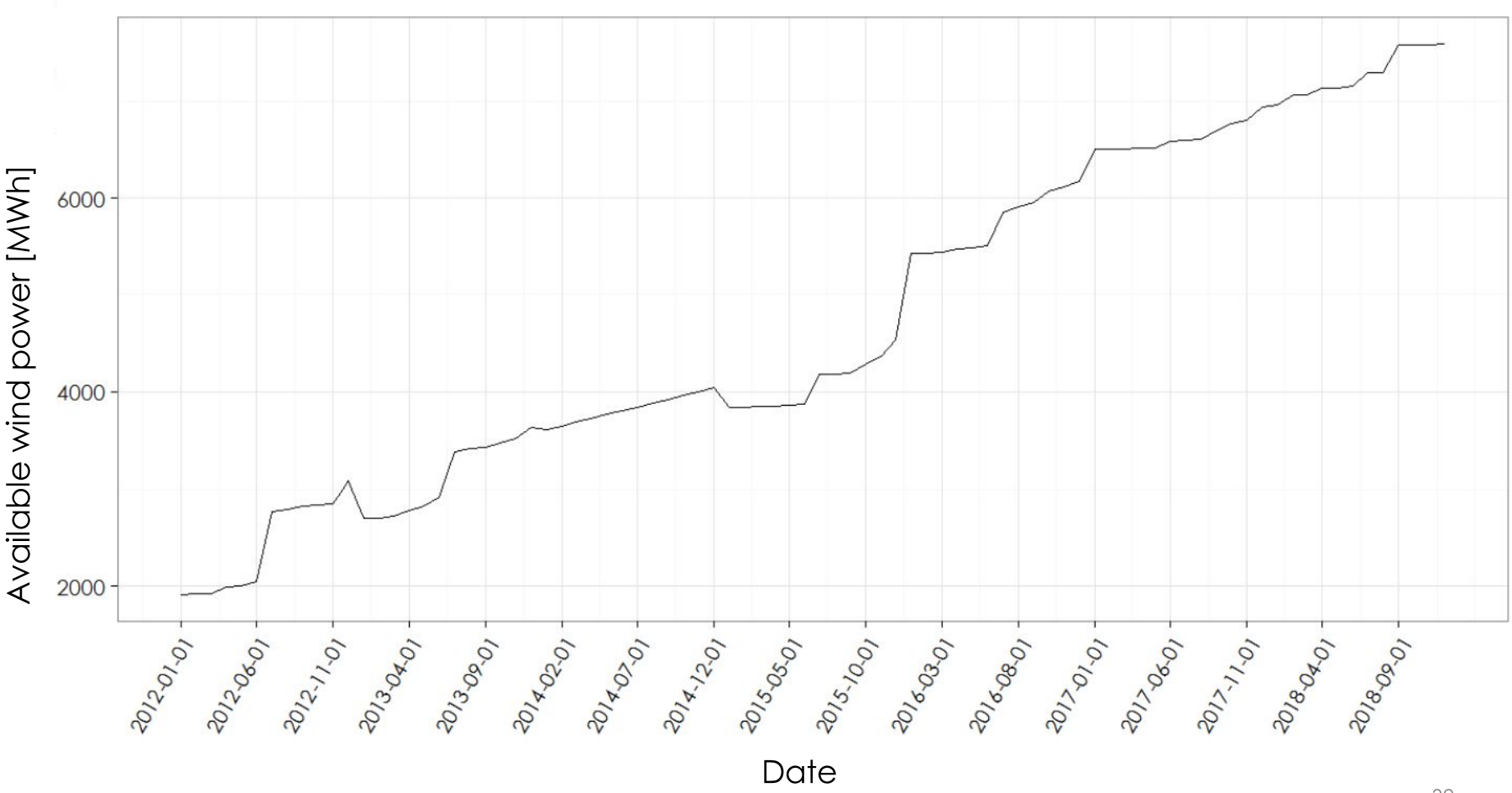
Dane o rynku

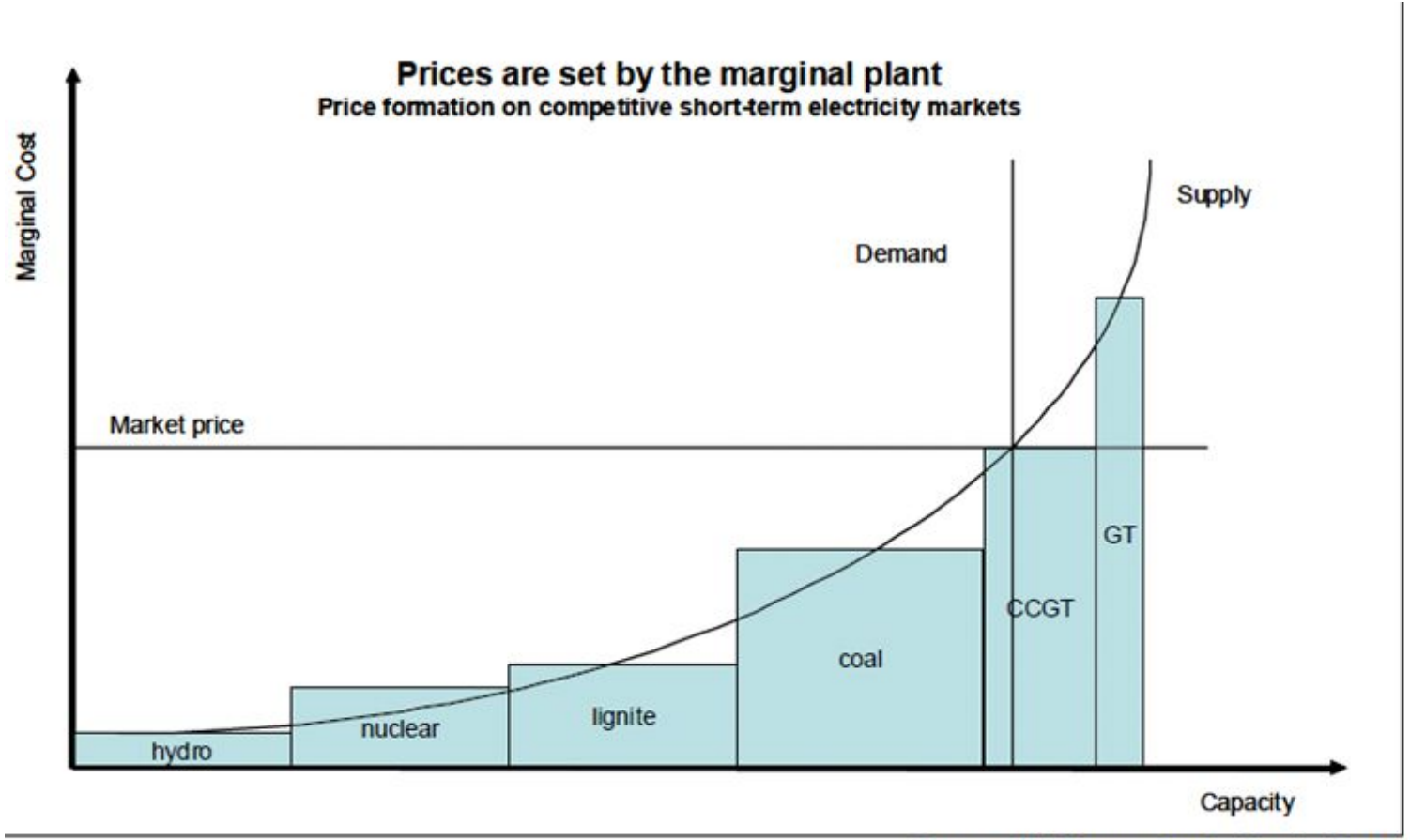
Polskie Sieci Elektroenergetyczne
Szukaj...
O nas
Obszary działalności
Dane systemowe
Biuro prasowe
Przetargi
Dokumenty
Kontakt
Nowa Strategia PSE
2017 - 2019
Aktualności
Komunikaty OSP
Przetargi
2017-04-26 / Komunikat w sprawie rozpoczęcia procesu jednostronnego przetargu miesięcznego na zdolności przesyłowe połączenia PSE S.A. i NEK UKRENERGO na CZERWIEC 2017 r.
Komunikat Operatora Systemu Przesyłowego w sprawie rozpoczęcia procesu jednostronnego przetargu miesięcznego na zdolności przesyłowe połączenia PSE S.A. i NEK UKRENERGO na CZERWIEC 2017 r. (Okres Rezerwacji rozpoczynający się od dnia 1 czerwca 2017 r. i kończący się z dniem 30 czerwca 2017 r.) więcej...
2017-04-11 / Komunikat w sprawie ogłoszenia wyników jednostronnego przetargu miesięcznego na zdolności przesyłowe połączenia PSE S.A. i NEK UKRENERGO na MAJ 2017 r.
Komunikat Operatora Systemu Przesyłowego z dnia 11 kwietnia 2017 r. w sprawie ogłoszenia wyników jednostronnego przetargu miesięcznego na zdolności przesyłowe połączenia PSE S.A. i NEK UKRENERGO na MAJ 2017 r. (Okres Rezerwacji rozpoczynający się od dnia 1 maja 2017 r. i kończący się z dniem 31 maja 2017 r.) więcej...
2017-04-04 / Komunikat w sprawie ogłoszenia jednostronnego przetargu miesięcznego na zdolności przesyłowe połączenia międzysystemowego PSE i NEK UKRENERGO na MAJ 2017 r.
Redukcja zapotrzebowania na polecenie OSP
Bierz aktywny udział w bilansowaniu KSE
Zapotrzebowanie mocy KSE w dniu 2017-04-26
24000
22000
20000
18000
16000
14000

TGE
Szukaj w serwisie
OK
Strony Grupy Kapitałowej GPW
STRONA GŁÓWNA | KONTAKT | MAPA STRONY | ENGLISH VERSION |
O giełdzie
Dane giełdowe
Market Coupling
Członkostwo
Rejestry
Konferencje
Pressroom
Rynek Instrumentów Finansowych
RRM
KOMUNIKATY OPERACYJNE (KO)
Bieżące
Archiwum
27.04.2017 10:03 Zdolnosci przesylowe zostały opublikowane
INFORMACJE
AKTUALNOŚCI
ARCHIWUM
RSS
13-04-17 Konferencja Naukowapt. „20 latustawy – Prawo energetyczne"
12 kwietnia 2017 r. w Sali Pod Kopułą Ministerstwa Rozwoju w Warszawie odbyła się Międzynarodowa Konferencja Naukowa pt. „20 lat ustawy – Prawo energetyczne".
Towarowa Giełda Energii była Partnerem Konferencji. Spotkanie dało możliwość wymiany doświadczeń i poglądów związanych z energetyką i rolą, jaką ma do odegrania krajowy regulator w zrównoważonym rozwoju gospodarczym Polski.
więcej
04-04-17 Podsumowanie obrotów na rynkach TGE w marcu 2017 r.
Marcowe wolumeny obrotu energią elektryczną i gazem ziemnym były na TGE niższe względem roku ubiegłego – o 27,8 proc. dla energii (przy wolumenie 7 738 675 MWh) oraz o 39,9 proc. dla gazu (7 347 094 MWh).
Miesięczny wolumen obrotu prawami majątkowymi dla energii elektrycznej był w marcu najwyższy od dwóch lat i wyniósł 8 120 795 MWh (wzrost r/r o 21,2 proc.).
więcej
21-03-17 Zmiana czasu zimowego na czasletni
W związku ze zmianą czasu zimowego na czas letni Towarowa Giełda Energii SA informuje:
WYNIKI SESJI
RDN
RCB
RTT
RDNg
RDBg
RTTg
RTTpm
RUB
Rynek Dnia Następnego
PLATYNOWE MEGAWATY
III EDYCJA KONKURSU DLA MEDIÓW
ZGŁOSZENIA
20.03.2017 - 20.04.2017
WIĘCEJ
gpi
dane rynku energii elektrycznej
GIEŁDOWA PLATFORMA INFORMACYJNA

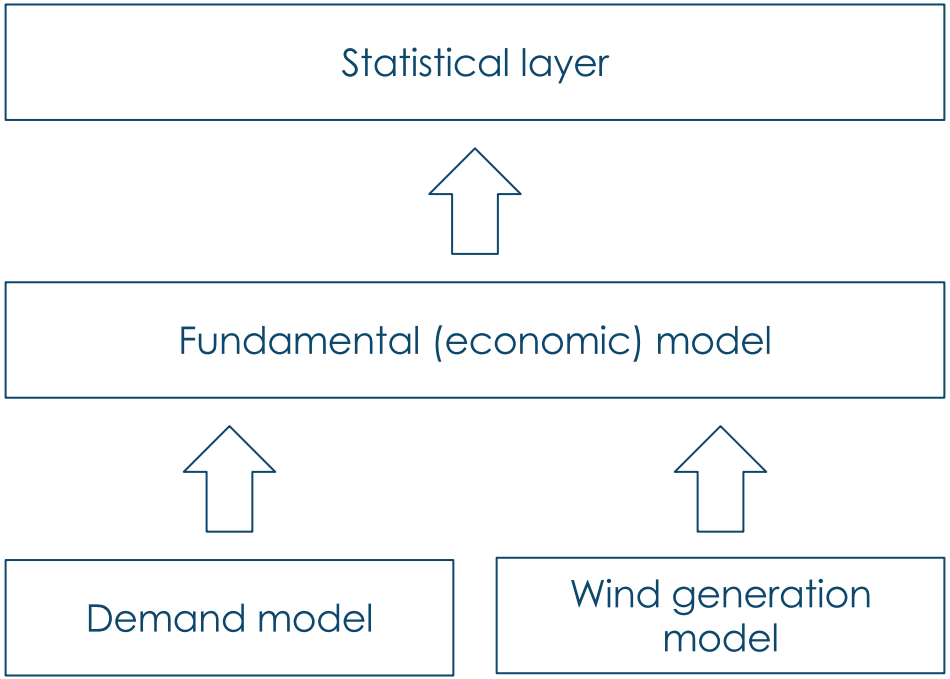

(multi-)Seasonal
Calendar effects
Highly variable
Spikes
Price PLN/MWh
600
400
200
2015-08-31
2015-09-07
2015-09-14
2015-09-21
2015-09-28
Date

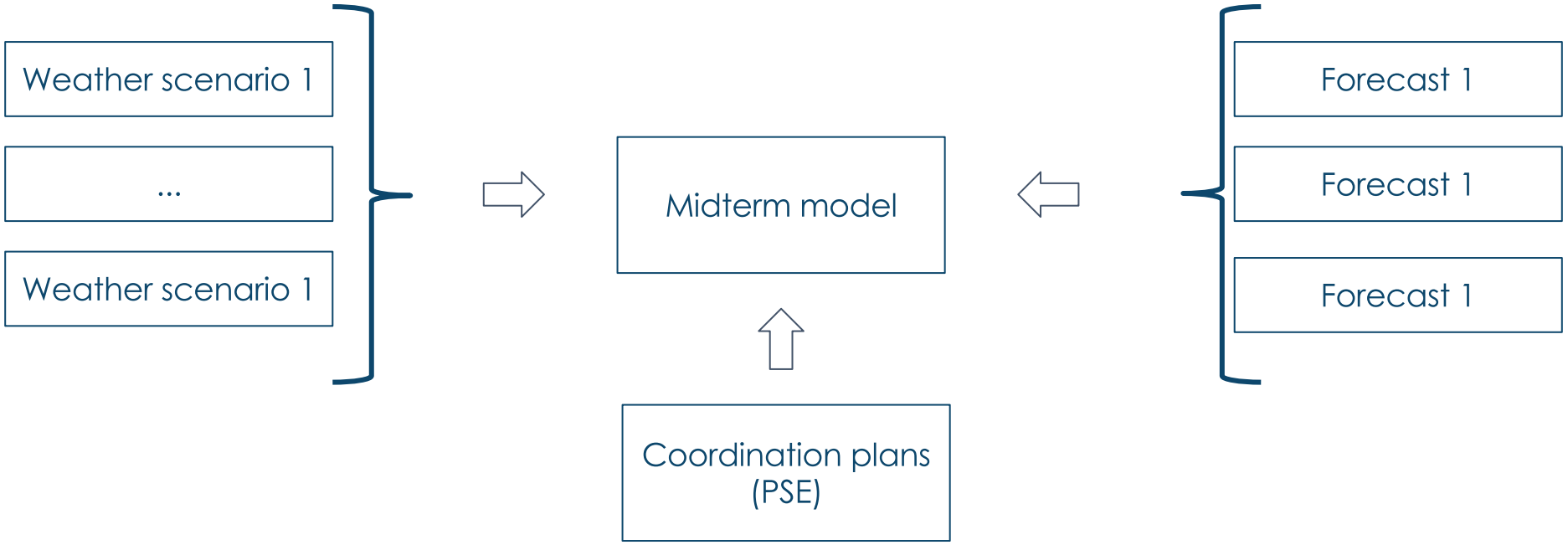

Wind power generation
4000
3000
2000
1000
0
10
20
30
Wind speed


Available wind power [MWh]
6000
4000
2000
2012-01-01
2012-06-01
2012-11-01
2013-04-01
2013-09-01
2014-02-01
2014-07-01
2014-12-01
2015-05-01
2015-10-01
2016-03-01
2016-08-01
2017-01-01
2017-06-01
2017-11-01
2018-04-01
2018-09-01
Date

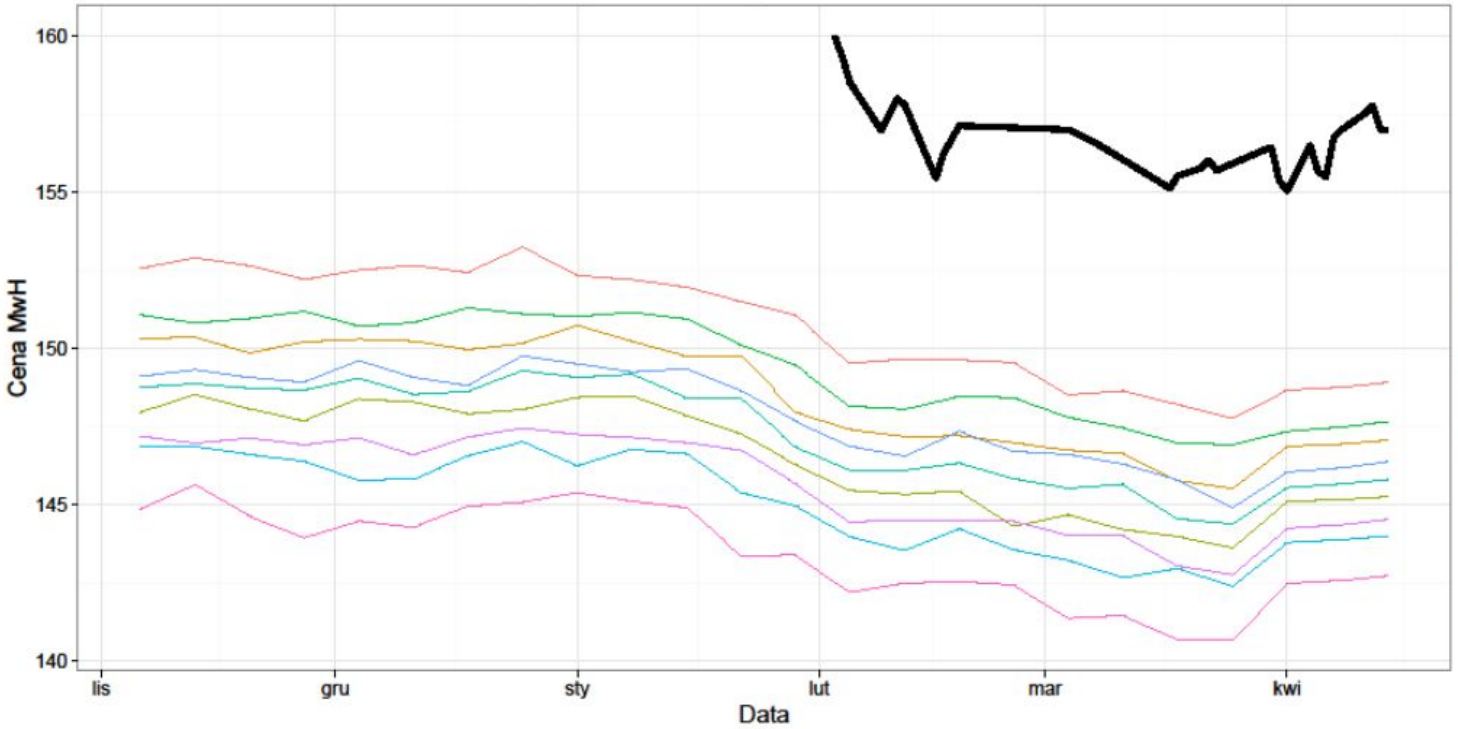


Source: EU Energy Sector Inquiry Report


Statistical layer
Fundamental (economic) model
Demand model
Wind generation model


Weather scenario 1
...
Weather scenario 1
Midterm model
Coordination plans (PSE)
Forecast 1
Forecast 1
Forecast 1

M5 Base 2016 forward contract price and SPOT price scenarios



Scenario
S1 low temp, light wind
S2 low temp, ave wind
S3 low temp, strong wind
S4 ave temp, light wind
S5 ave temp, ave wind
S6 ave temp, strong wind
S7 high temp, light S2 wind
S8 high temp, ave wind
S9 high temp, strong wind

Summary

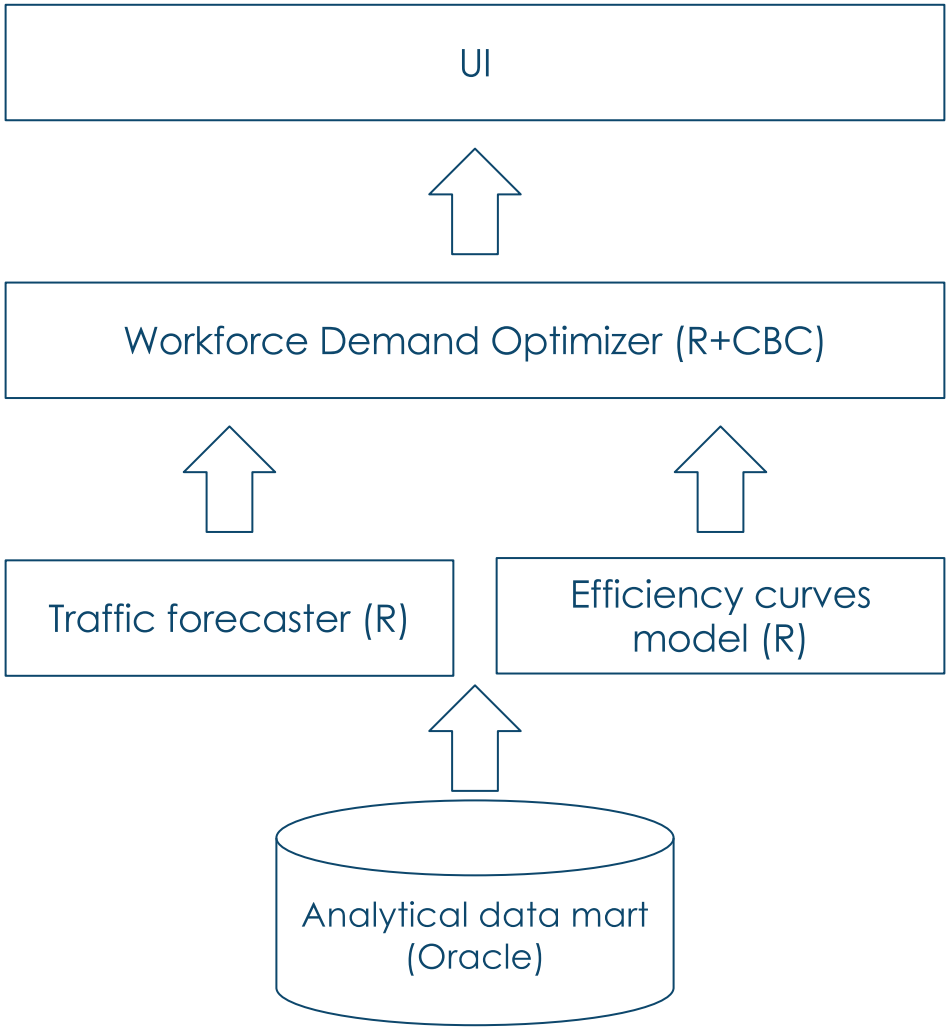
- Around 5K LOC
  - All Energy Pack is around 30K LOC
- Excel based UI

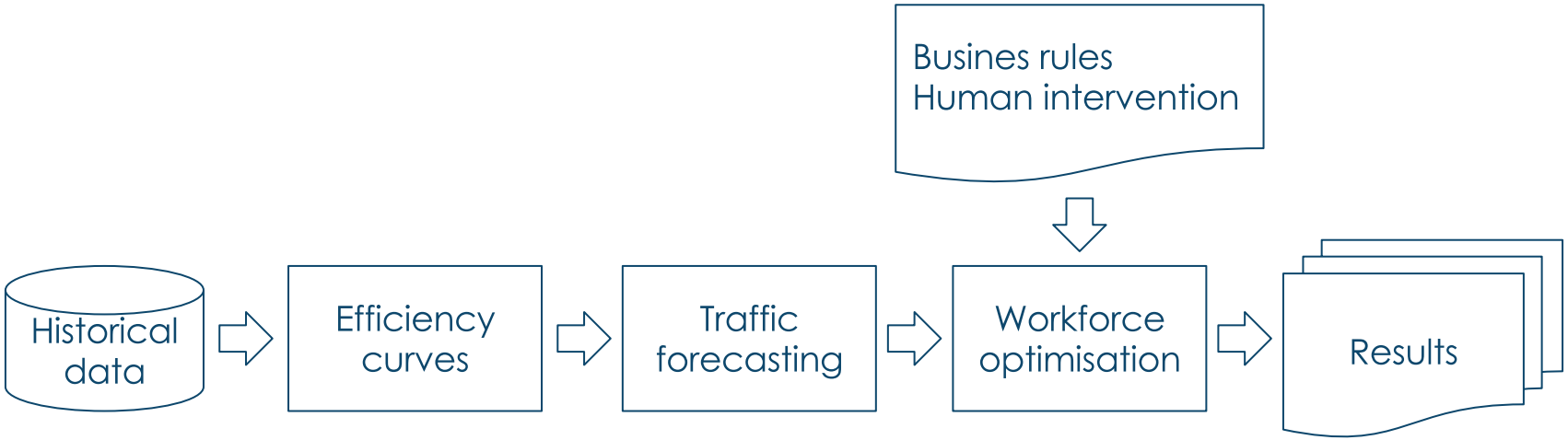
- logging
- data.table
- RCurl
- doParallel
- knitr

Workforce Demand Optimisation

How many people should be in the location to generate optimal margin?

Shop	Warehouse	Call center


UI
Workforce Demand Optimizer (R+CBC)
Traffic forecaster (R)
Efficiency curves model (R)
Analytical data mart (Oracle)


Busines rules
Human intervention
Historical data
Efficiency curves
Traffic forecasting
Workforce optimisation
Results

Summary

- Around 15K LOC
- Integrated via DB
- First version of WLOG R Suite™

- logging
- data.table
- RCurl
- doParallel
- knitr

What is so special about R?

Glue language	Fast prototyping in large scale	Easy to build a team knowing R

WLOG Solutions

We are strategic partner in R based large scale analytics.

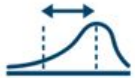
Since 2005 with large scale R deployments with up-to 20-30K LOC.

Analytical experience

Customer analytics

Demand forecasting

Pricing analytics


Risk modelling

Performance optimisation

Supply chain optimisation

Understand why we are best for being Your partner in R for large scale analytics.

Industry knowledge

Banking

Insurance

Logistics

Pharmaceutical

Retail

Telecommunication

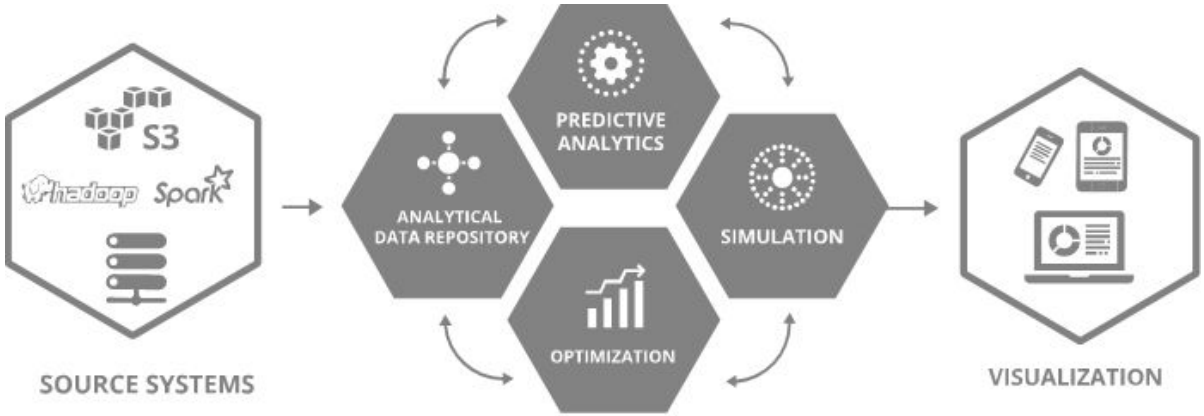
Utilities

Hire us for an analytical project

We offer implementation and maintenance of large-scale R data science solutions linking predictive modelling, optimization and simulation components.

Analytical data repository → **Predictive analytics** + **Optimization** + **Simulation** → **Visualization**

Understand our analytical stack



Elastic

- We can attack any business problem.

Available

- Fast prototyping for both models and UI

Scalable

- More power adding computing nodes

Efficient

- High quality in short time

Wit Jakuczun, PhD

wit.jakuczun@wlogsolutions.com
http://www.wlogsolutions.com
https://github.com/WLOGSolutions