



Avancerad dataanalys: En datadriven möjlighet för mer tillförlitliga tillrinningsprognoser

9 december 2020

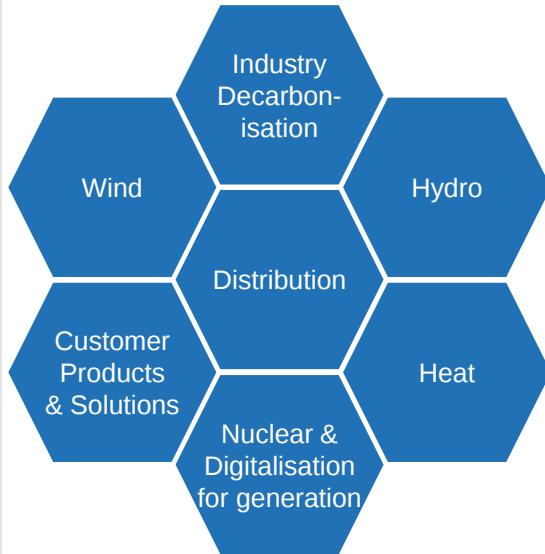
**Mats Billstein, Mikael Klingvall,
Fredrik Dillner**

Disposition

1. **Introduktion till Vattenfall R&D**
2. **Syfte och mål med projektet**
3. **Begrepp datadriven analys**
4. **Långtidsprognos av inflöden**
5. **Korttidsprognos av inflöden**
6. **Vårflods ankomst**
7. **Sammanfattning**

Portfolios and sections

7 R&D Portfolios

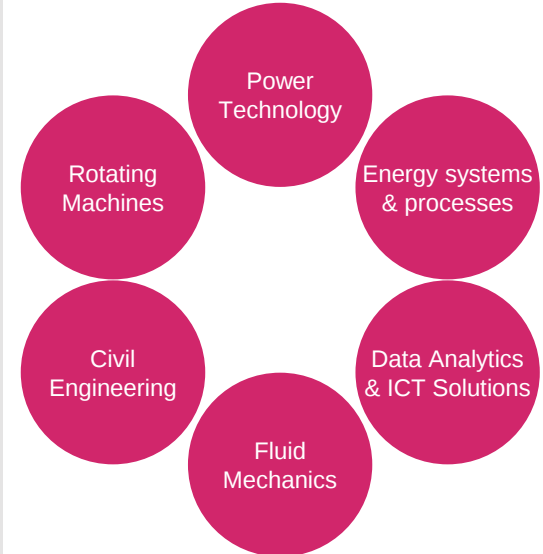


Staff: 125 employees

Strategic R&D:
Development,
Demonstration,
Deployment

Tactical R&D:
Problem solving,
qualified testing and
validation for internal
and external customers

6 Depts / Sections



Vattenfall's value chain



Production

- Production from
- Hydro
- Nuclear
- Coal
- Natural gas
- Wind
- Solar
- Biomass
- Waste
- Actively phasing out fossil-based production



Electricity distribution

- Guarantees secure supply via well-functioning distribution networks and smart network solutions
- Enables customers to feed self-generated electricity into the grid ("prosumers")



Sales of electricity, heat and gas

- Sells electricity, heat and gas to consumers and business customers
- Focuses on various price and service models, and gives customers the opportunity to reduce their environmental impact



District heating

- Drives the transformation towards fossil-free heating and cooling solutions together with cities and regions
- One of Europe's largest producers and distributors of district heating



Energy services & decentralised generation

- Offers energy services
- Heat pumps
- Solar panels
- Charging solutions for electric vehicles
- Battery storage
- Network services
- Smart meters
- Provides marketplaces and access to marketplaces where customers can buy and sell electricity

Laboratory resources

- Fluid Mechanics
- Hydropower turbines
- Rotating machines
- Fatigue tests
- ICT-labs
- Concrete technology
- Eco-hydraulics
- 3D-engineering
- Drone tests
- Embankment dam
- Battery lab



Hydro R&D

FLEXIBLE GENERATION

- System analyses
- Ancillary services
- Cost and value



DAM SAFETY & INFRASTRUCTURE

- Hydraulic eng
- Soil mechanics
- Concrete and rock mechanics



TURBINES & GENERATORS

- Transient turbine test
- Rotordynamics
- Field measurements



ENVIRONMENT

- Fish migration
- Ecohydraulics
- Oil strategy

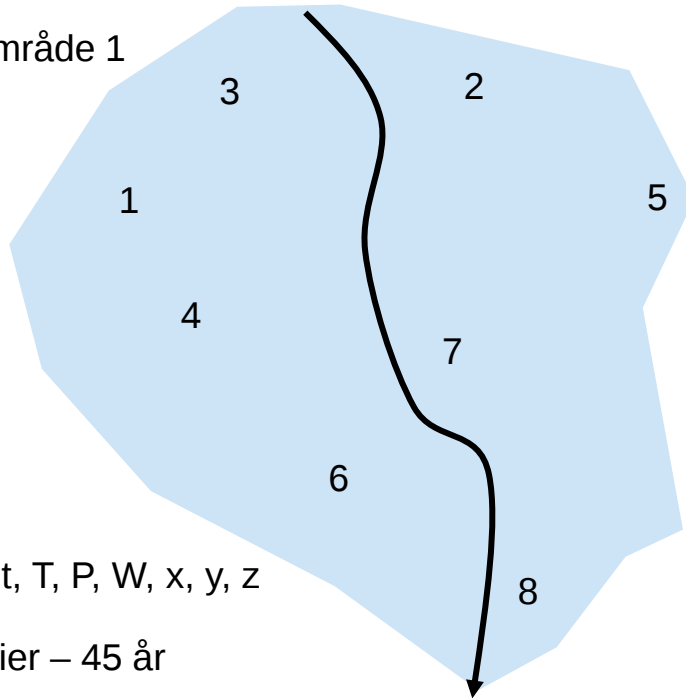


Vattenfalls prognosarbete (= lagerinventering)

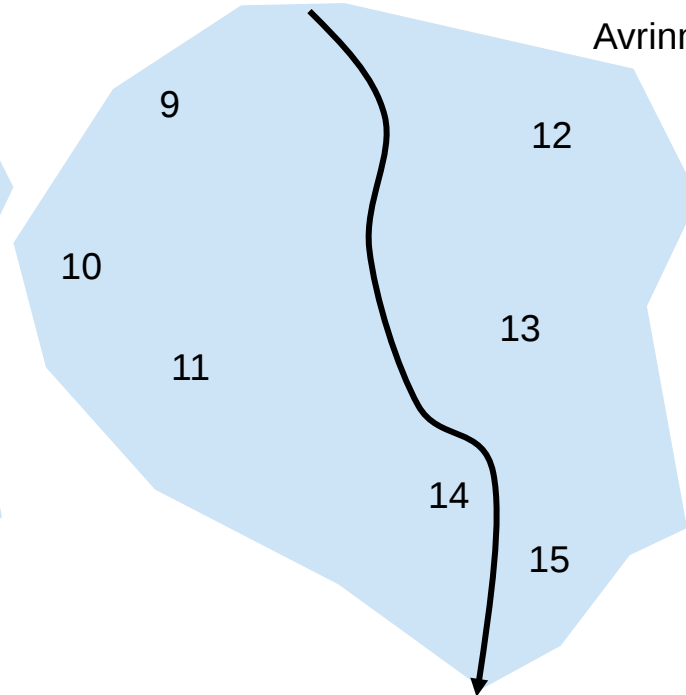
- Susanne och Mikael (och några till)
- Väl beprövade prognosmodeller
- **Bondepraktikan** (anno 1508 i Tyskland och 1662 i Sverige)
 - Om Anders braskar Julen slaskar och tvärt om
 - Mycket rönnbär ger kall och snörik vinter

Bondepraktikan 2.0 (anno 2020)

Avrinningsområde 1



Avrinningsområde 2



Mätstationer: t, T, P, W, x, y, z

Långa tidsserier – 45 år

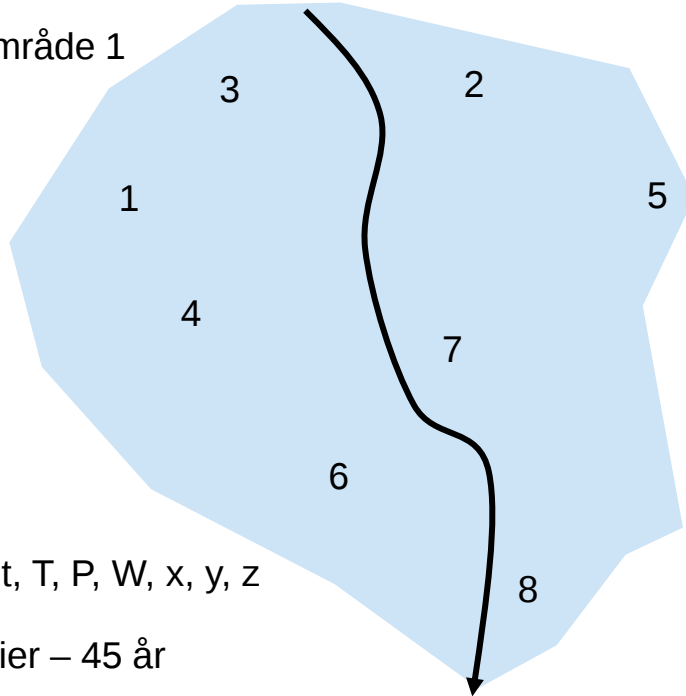
$$Q(t) = f(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots)$$

Känt

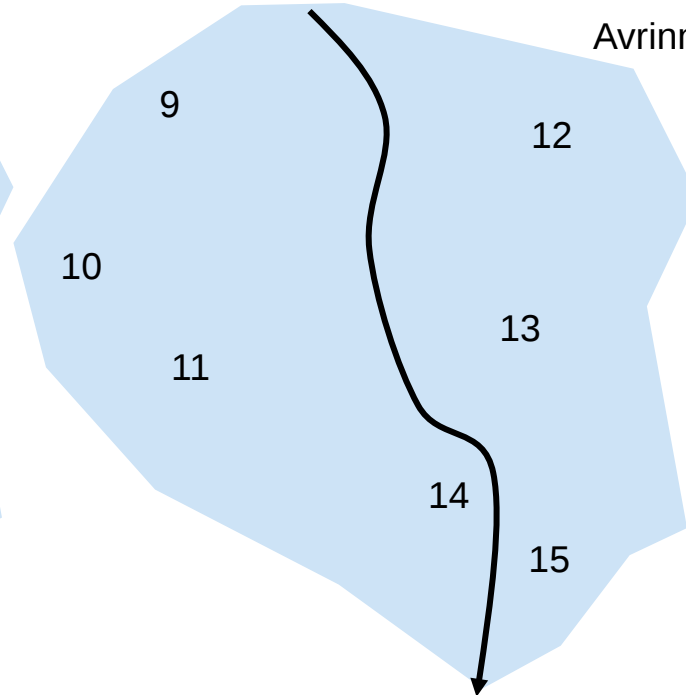
Känt

Bondepraktikan 2.0 (anno 2020)

Avrinningsområde 1



Avrinningsområde 2



Mätstationer: t, T, P, W, x, y, z

Långa tidsserier – 45 år

$$Q(t) = f(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots)$$

Okänd

Syftet och Mål



- Syfte:
 - Utforska historiska data med hjälp av avancerad dataanalys för prognostisering av vårflöden
 - Volym
 - Start
 - Utvärdera hur avancerad dataanalys står sig mot den etablerade HBV-modellen
- Mål:
 - Hitta kompletterande metoder och information som förbättrar tolkningen av befintlig modell
- Förutsättningar:
 - Tillgång till ML-modeller, statistiska metoder och stor datorkraft
 - Begränsad hydrologisk kunskap
 - Projektstart 1 oktober

Motsvarigheten i vårt årsmagasin

På dag 150 som är Veras*
kommer floden att dubblas

* 31/5 är Veras namnsdag

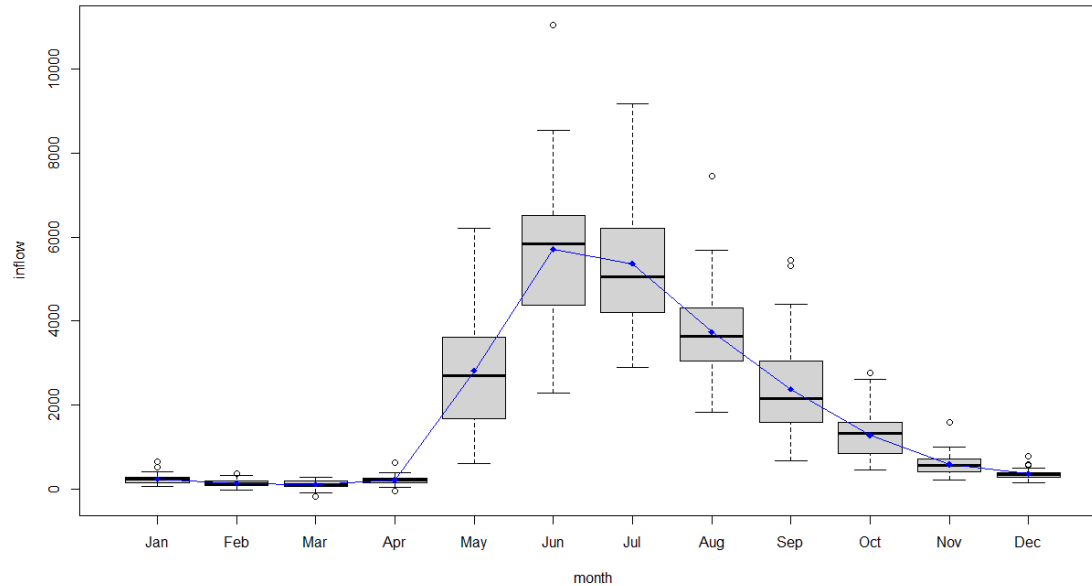
(16 % träffsäkerhet)

Begrepp

- Datadriven modellering
- Korrelation är inte kausalitet
- Överanpassning (eng overfitting)
- White box / Gray box / Black box

Långtidsprognoser av inflöde

- Utvärdera statistiska modeller



Dekomposition

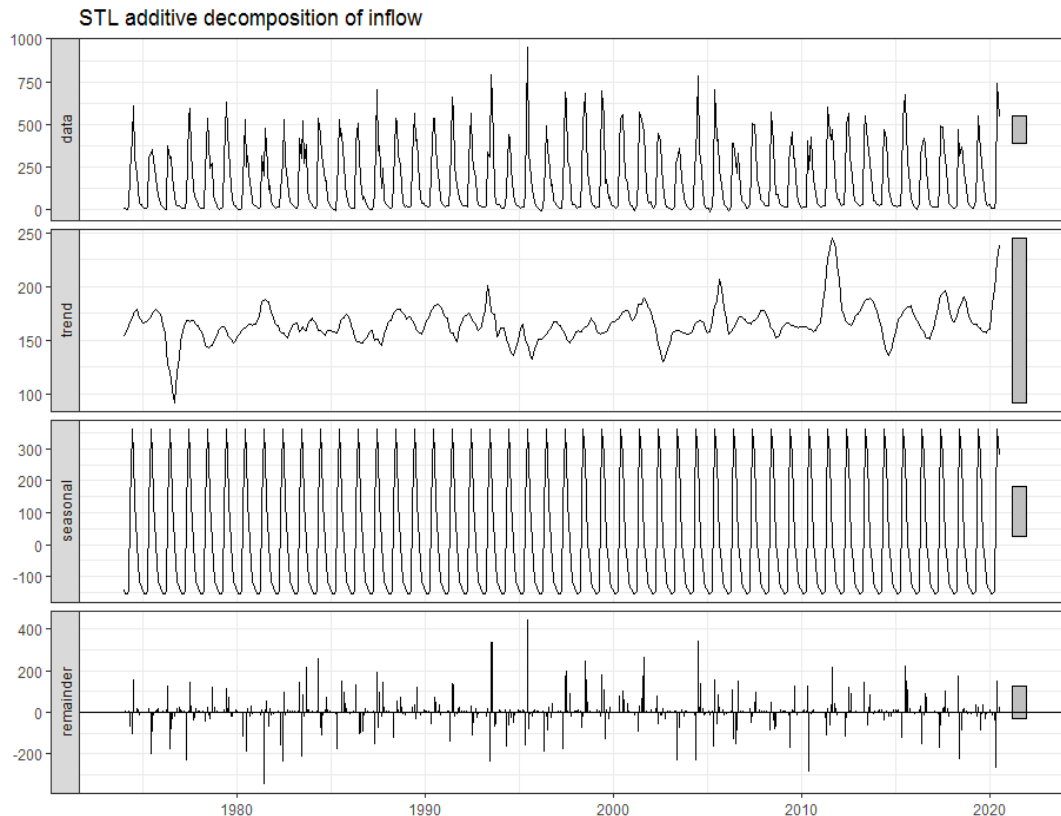
inflöden

- Trendkomponenten T_t
- Säsongskomponenten S_t
- Återstoden/residualen R_t

Additiv dekomposition av
tidsserien

$$Inflow_t = T_t + S_t + R_t$$

- Liten trend stark säsongs-
komponent



Tidsserier: modeller och prognoser

- Dataset

90% träning (1974 ~ 2015) 10% test (2016 ~ 2020)

- Modeller & metoder

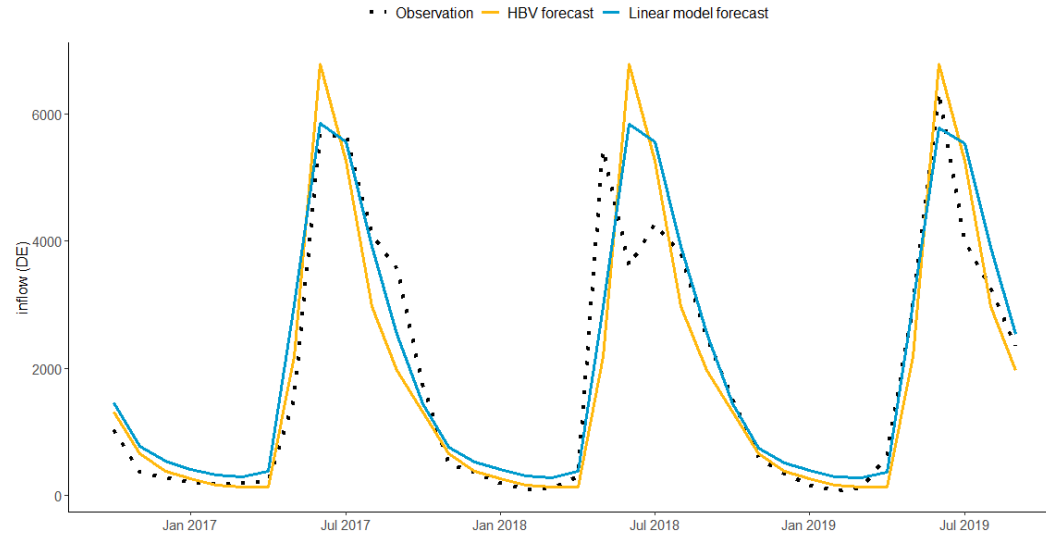
0. Benchmark: seasonal naive method

1. Linjär modell: trend säsong i en linjär modell

2. Säsong ARIMA(p,d,q) modell: $\text{ARIMA}(0,0,0) \times (1,1,1)_{12}$

Prognosjämförelse 2016 ~ 2019

(preliminär)



Modell	Prognos		Modell prestanda
	RMSE	Förbättring mot HBV (%)	R^2 value
Benchmark (navie)	1335.7	-42.6	0.64
HBV model	936.8	0	-
Linear model	758.5	19	0.849
ARIMA	757.8	19.1	0.85

Slutsats långtidsprognoser

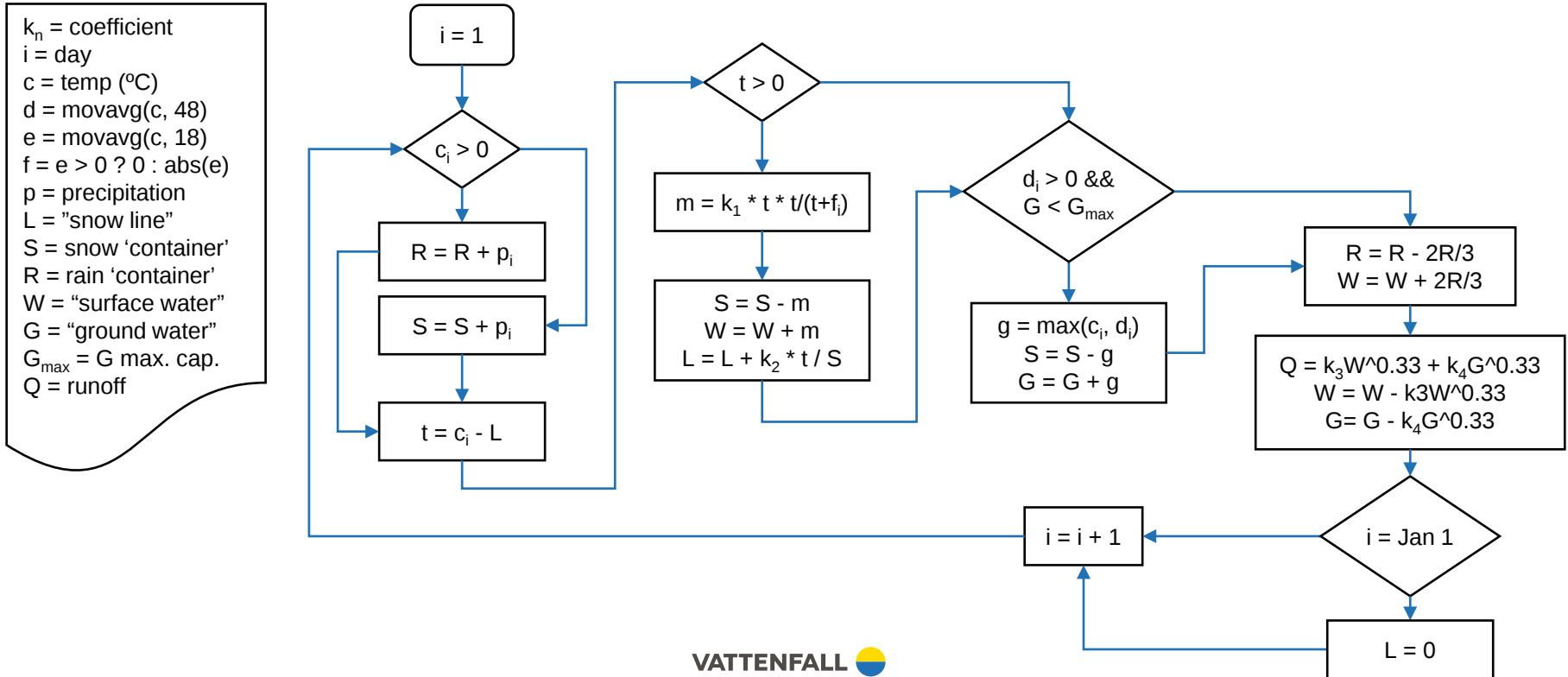
- Linjär modell avseende säsong och trend är en bra kandidat

$$Inflow_{t+h|t} = a + b \cdot Trend_t + c \cdot Seasonality_t + noisy_t$$

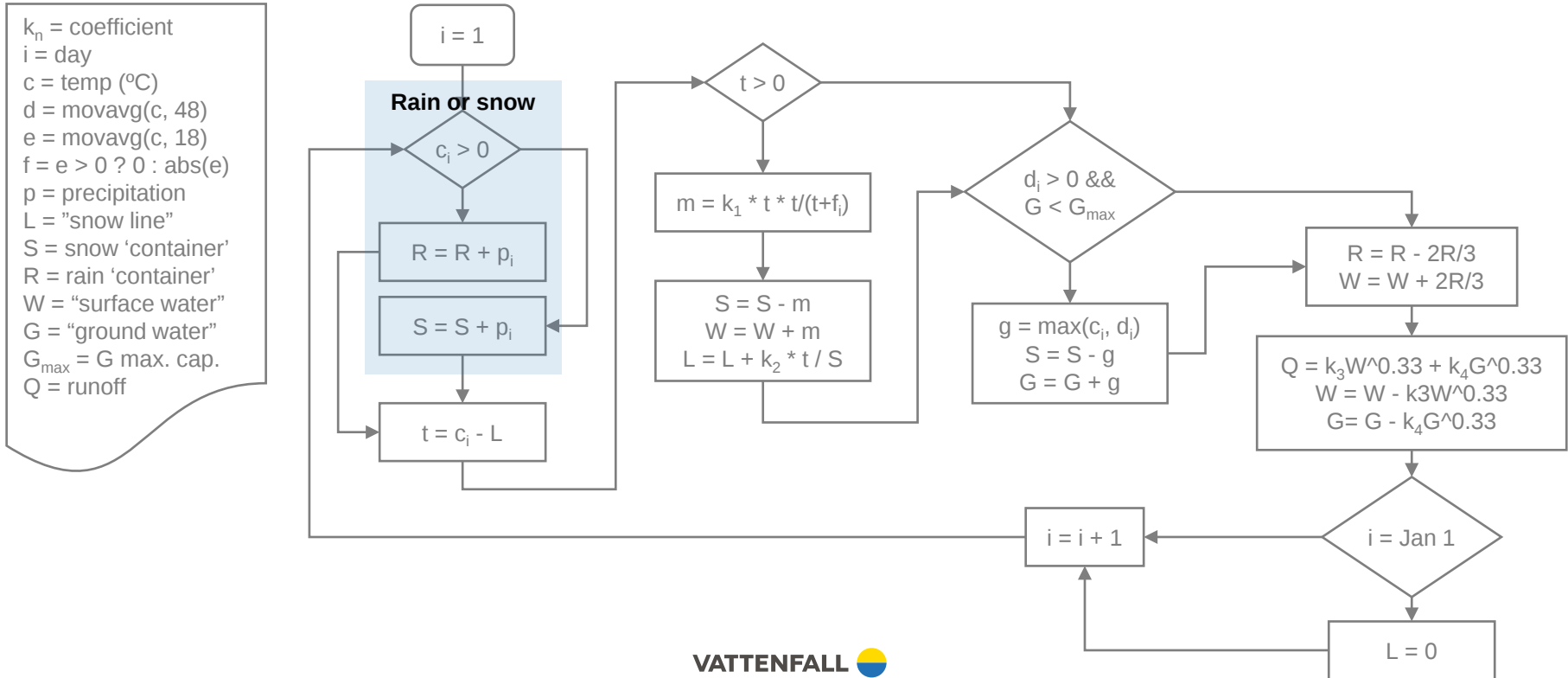
Korttidsprognos av inflöde

- Designmål: en enkel inflödesmodell (för prediktion av vårflodsstart)
- Syfte: hitta de mest prediktiva komponenterna i datasetet
- Datadriven analys (vi är som sagt inte hydrologer...)

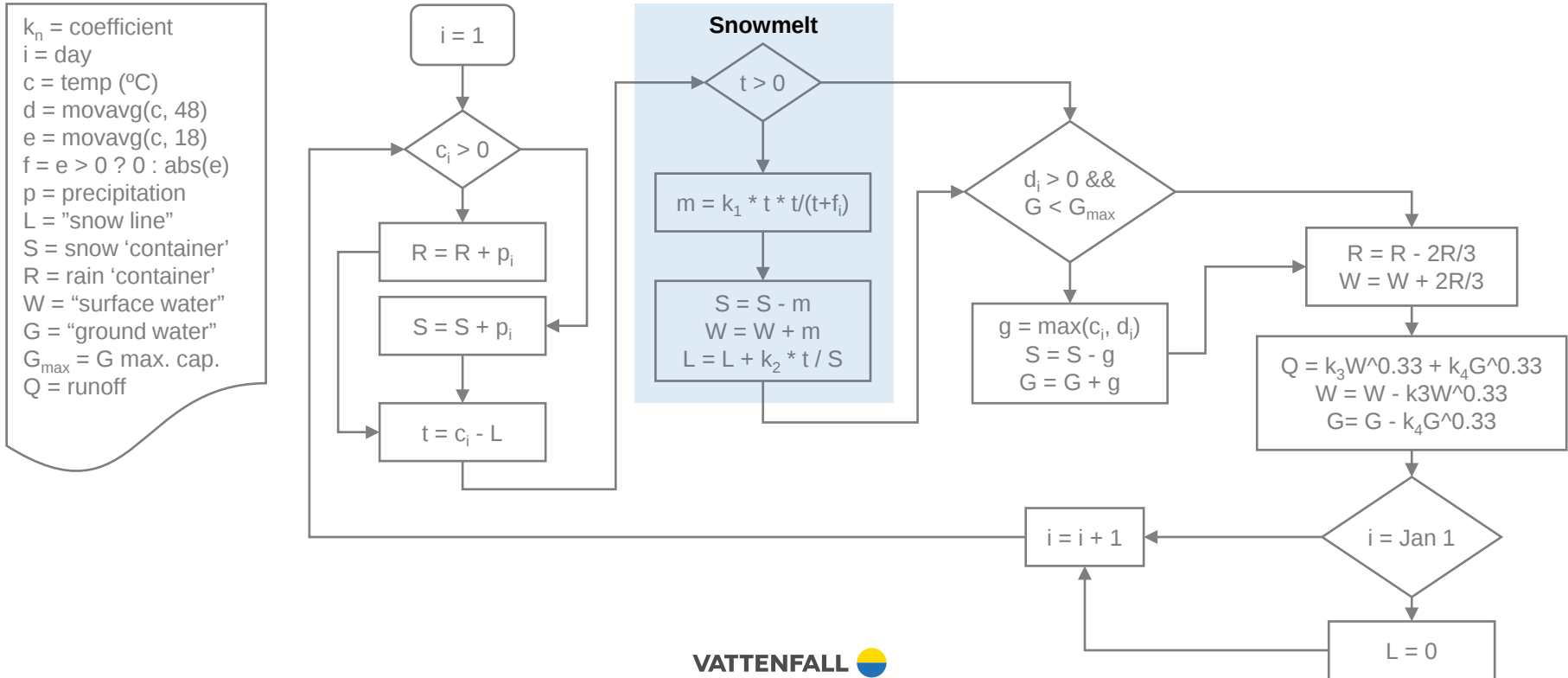
Runoff algorithm v0.4



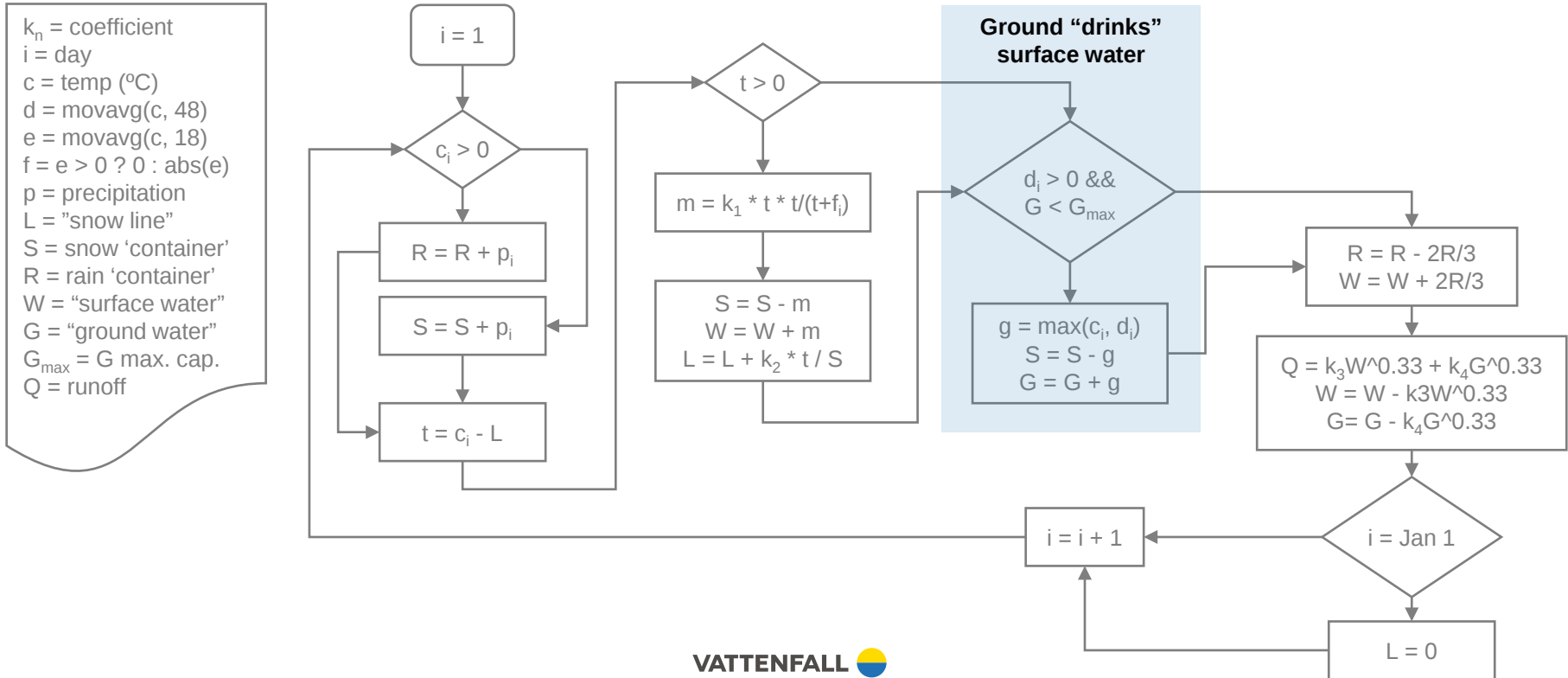
Runoff algorithm v0.4



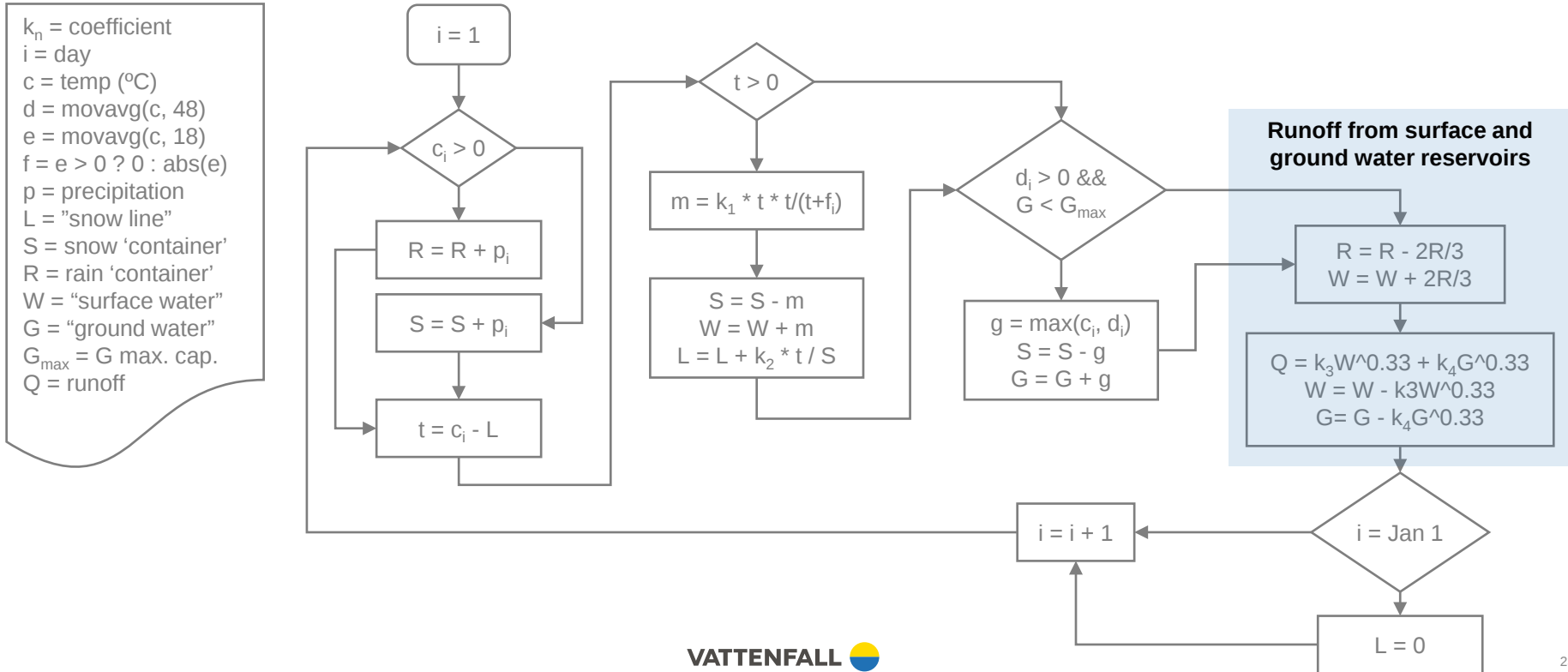
Runoff algorithm v0.4



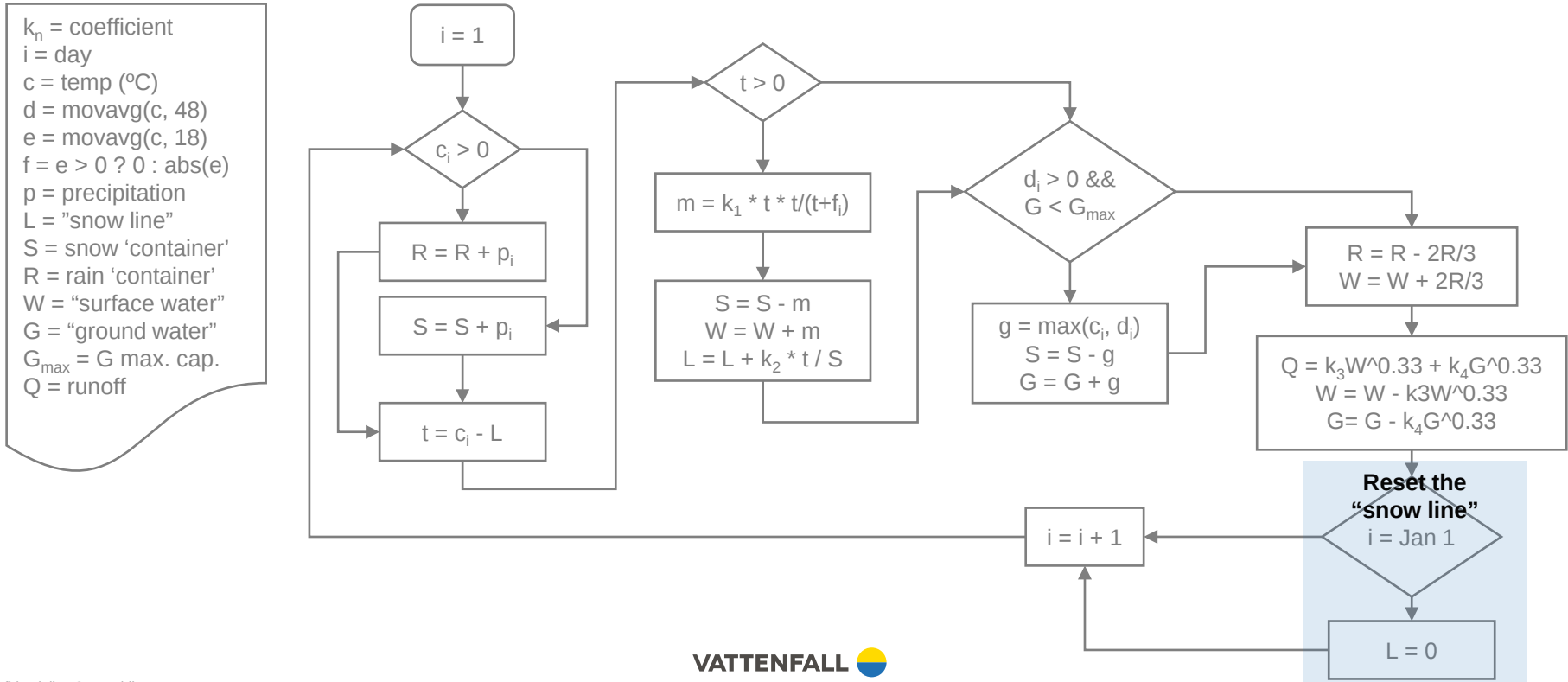
Runoff algorithm v0.4



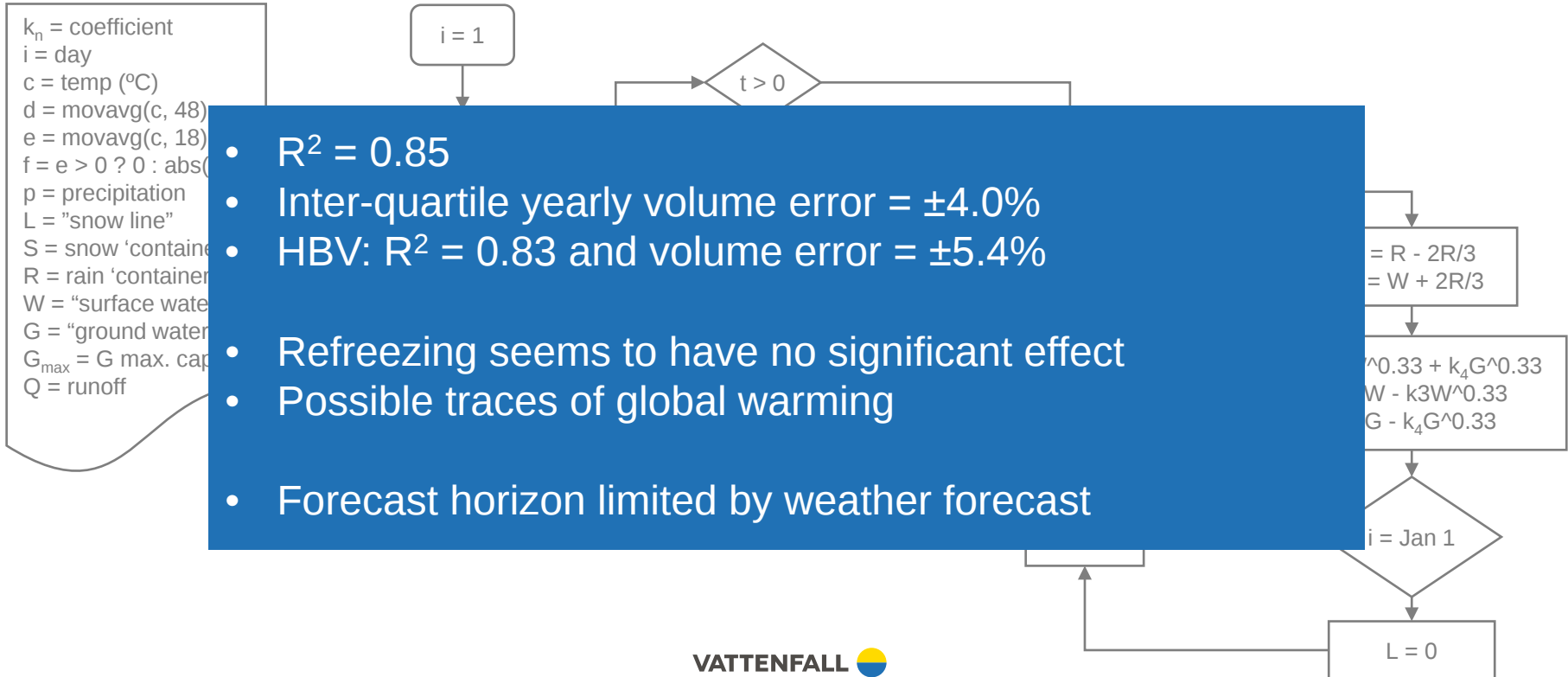
Runoff algorithm v0.4



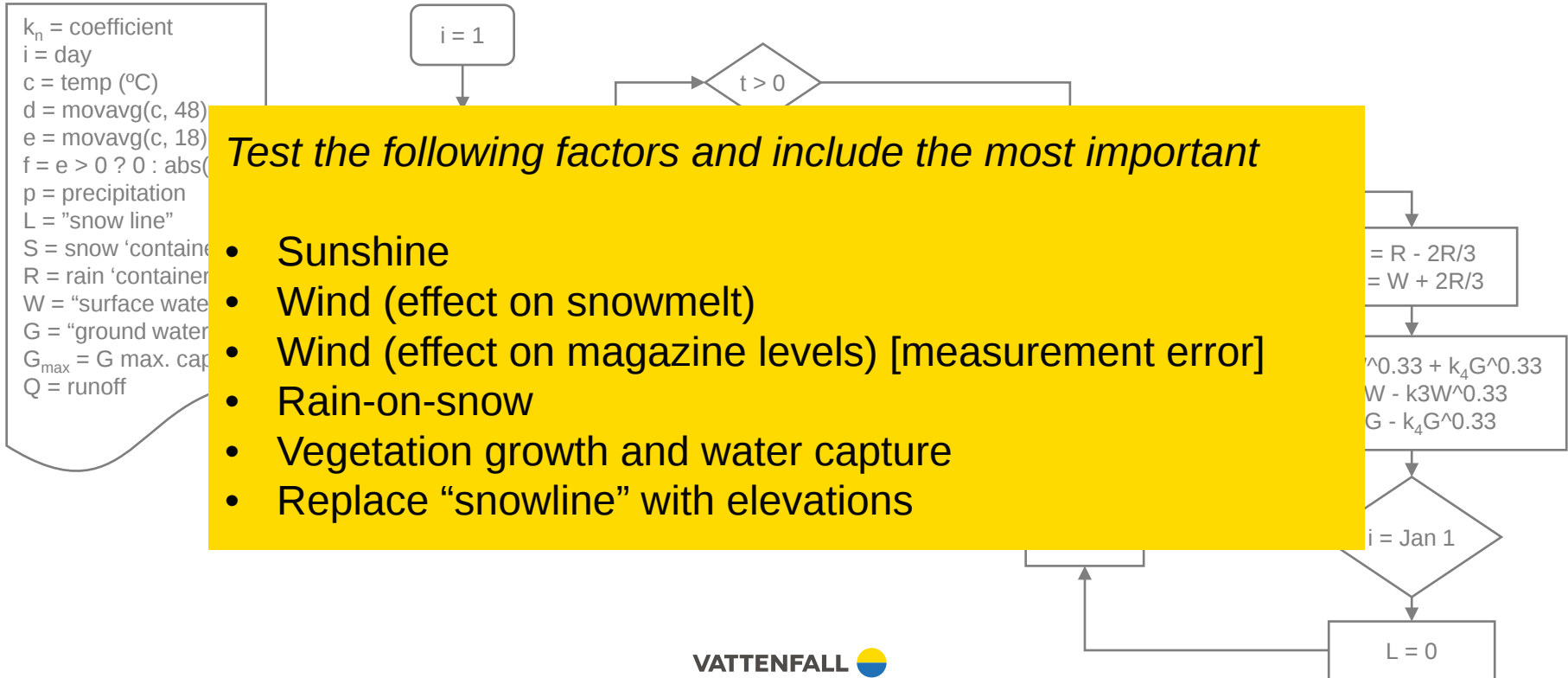
Runoff algorithm v0.4



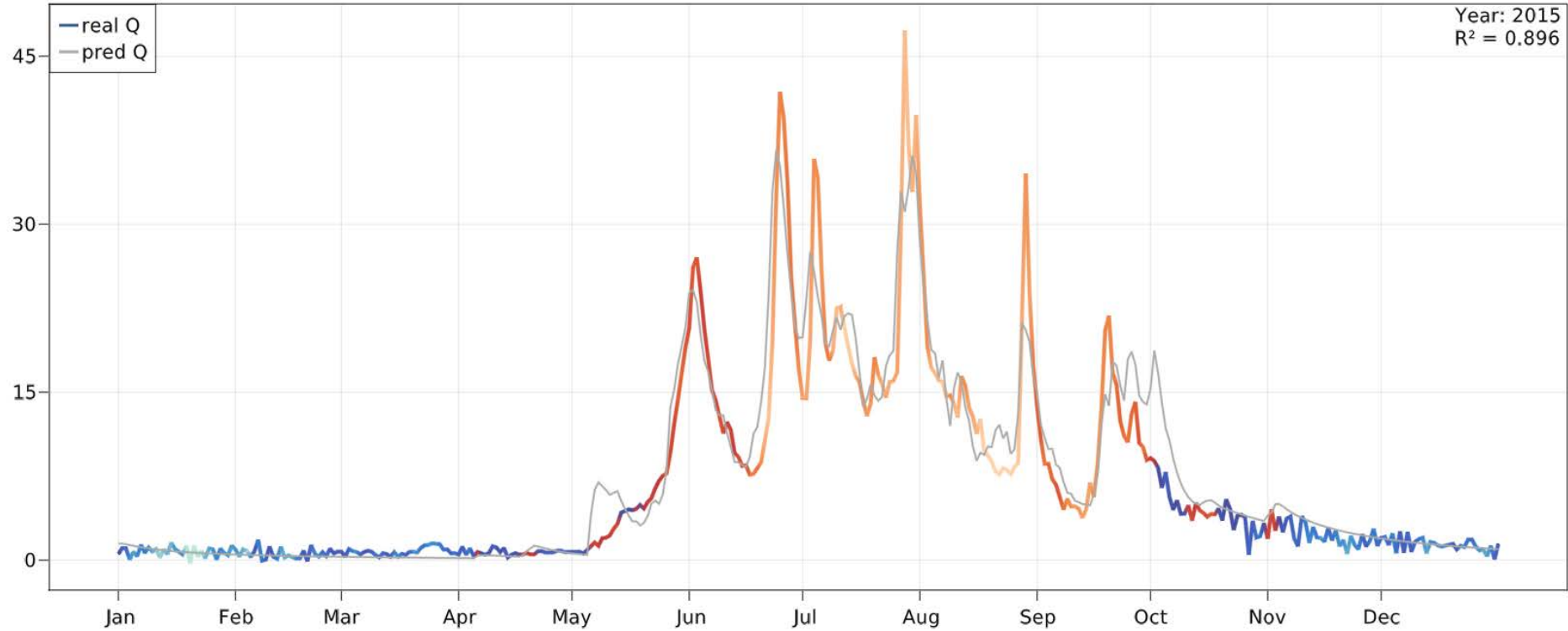
Runoff algorithm v0.4



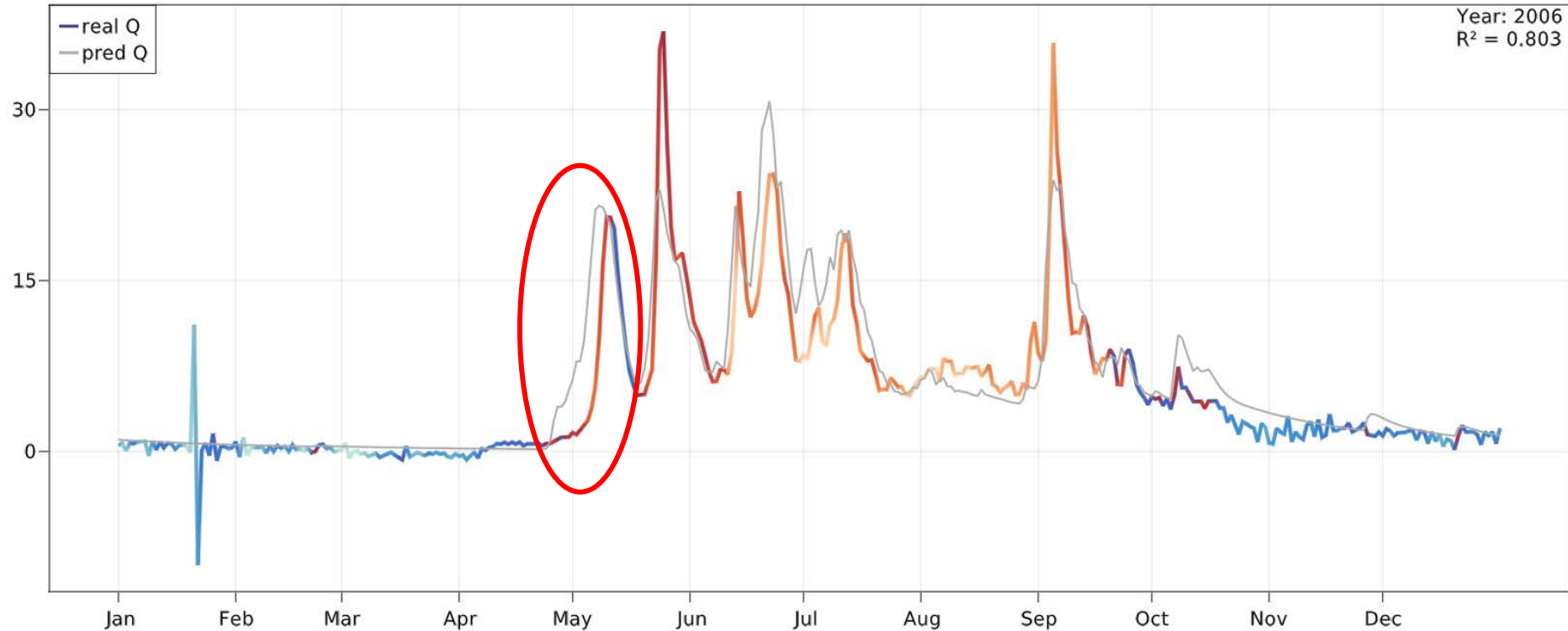
Runoff algorithm – next version



Example: 2015



Example: 2006



Vårflodsprognos

- När infaller vårfloden ?



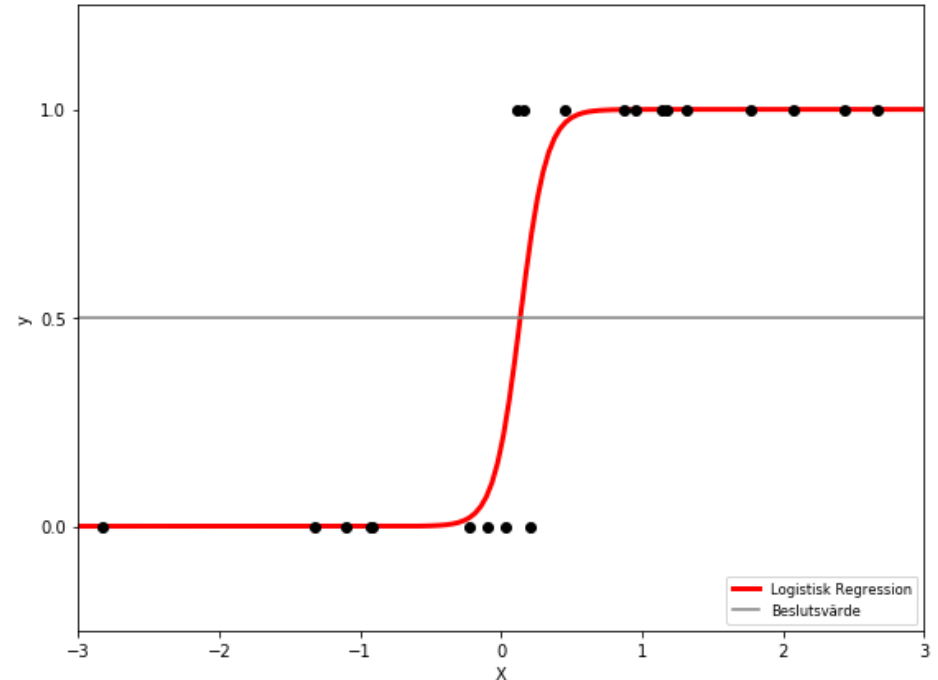
Logistisk regression

Logistisk funktionen, en variabel:

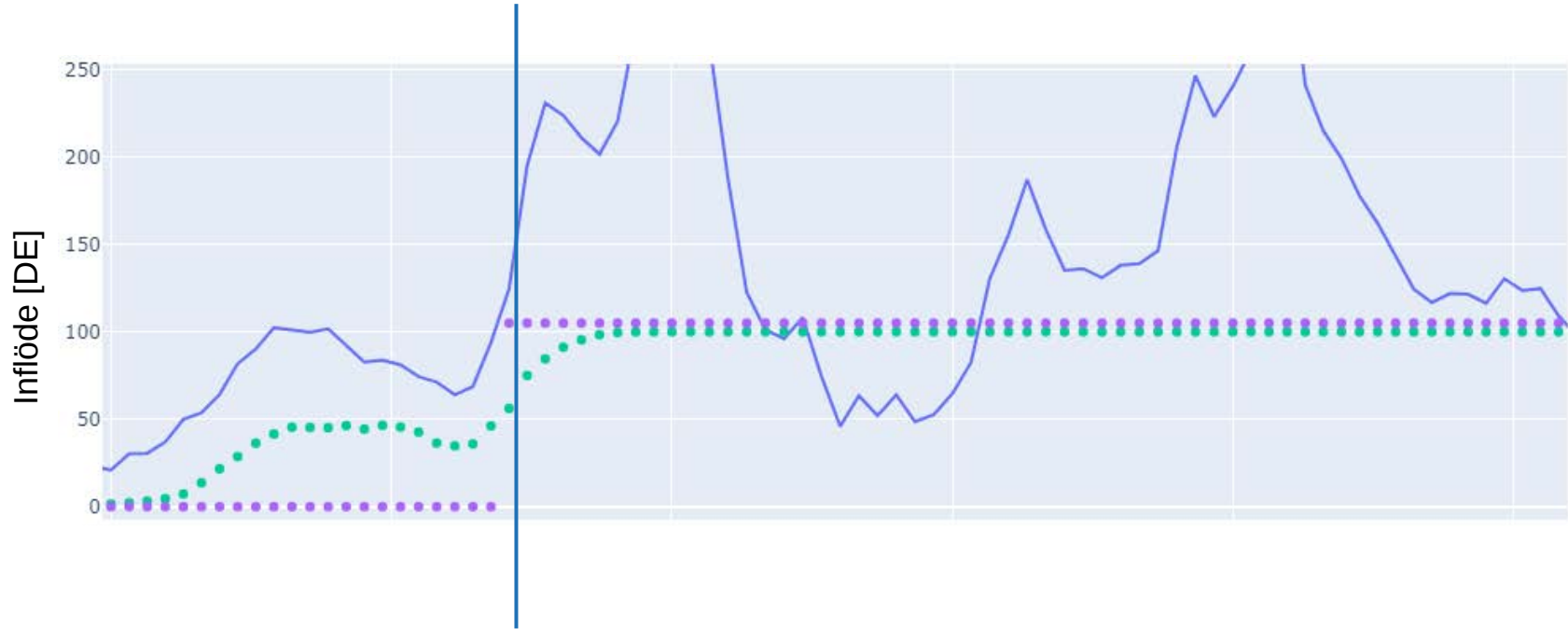
$$f(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Logistisk funktionen, flera variabler:

$$f(x_1, x_2) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$



Logistisk regression



Modell och Träffsäkerhet

Variabler:

Temperaturmedelvärde senaste nio dagarna

Summa Graddagar(0) sedan 1 april

	HBV	Log.Reg.
Rätt dag	11%	26%
+/- en dag	51%	60%
+/- två dagar	64%	69%
Prognos		33%

Sammanfattning

- Avancerad dataanalys kan komplettera beslutsunderlaget
- Mer arbete kvarstår
- Vattenfall planerar att fortsätta utvecklingsarbetet att prognostisera vårflöden (volym och start) med hjälp av avancerad dataanalys

Bondepraktikan 2.0 (anno 2020)

Om medeltempen är över 4.5 grader
så börjar vi återfylla våra lager*

** i 7 dagar*

44% träffsäkerhet på +/- en dag

64% träffsäkerhet på +/- två dagar



**Tack för
uppmärksamheten**