Hydraulic Engineering in Petabyte Age

Machine Learning for a Flow Resistance Problem

or

Man vs. Machine

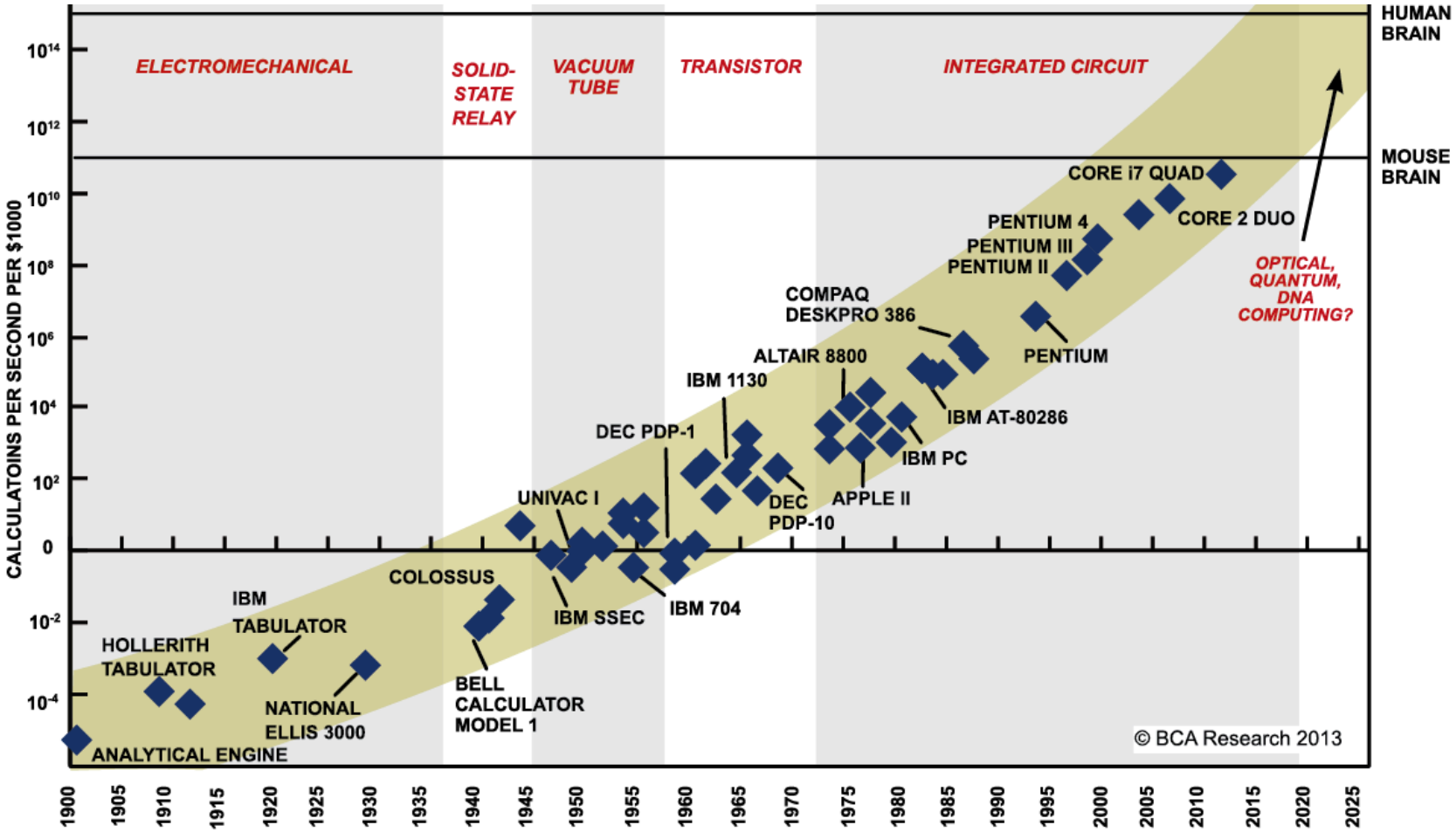
Prof. Vladan Babovic

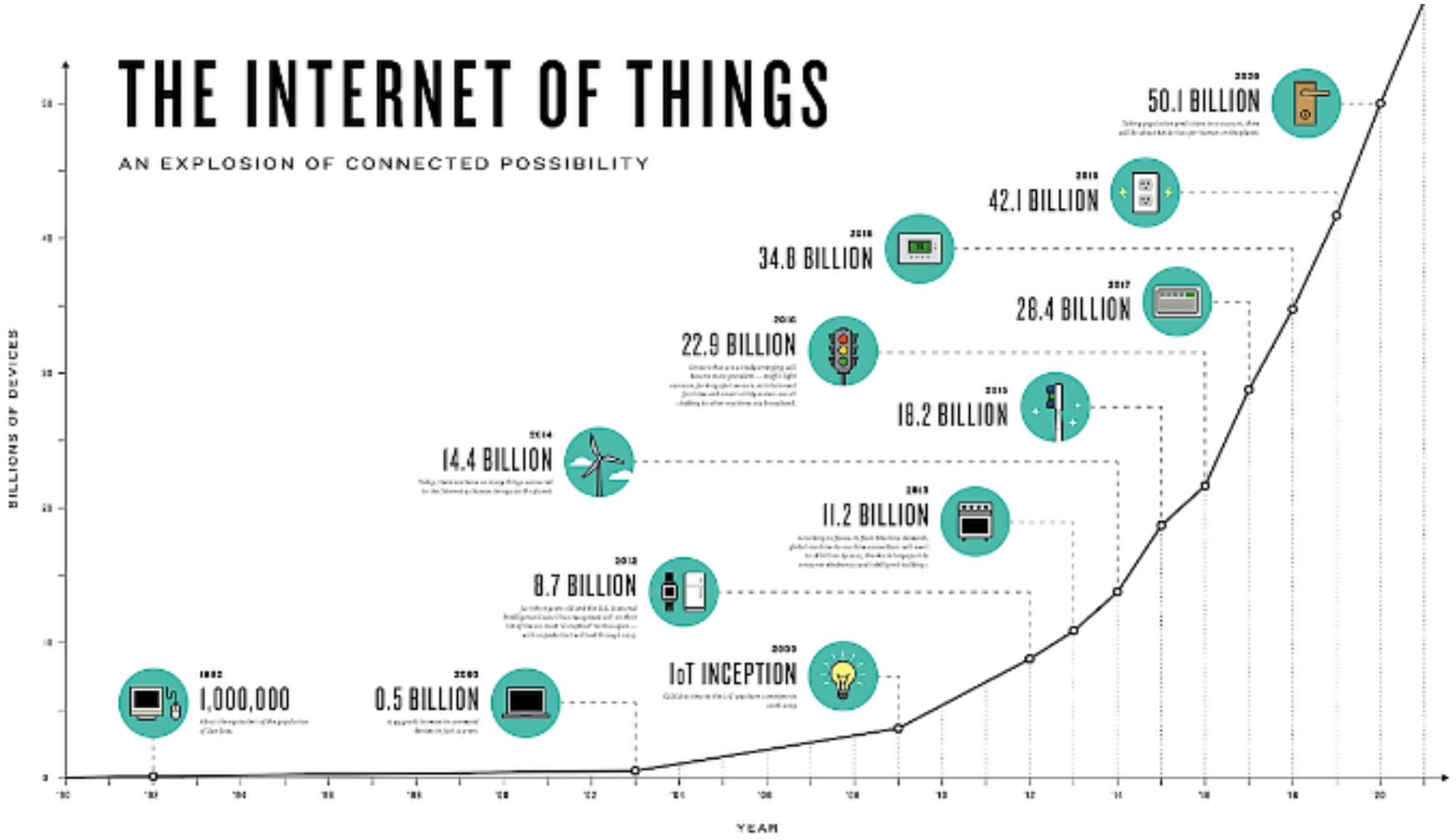
National University of Singapore

We Live in the Second Machine Age

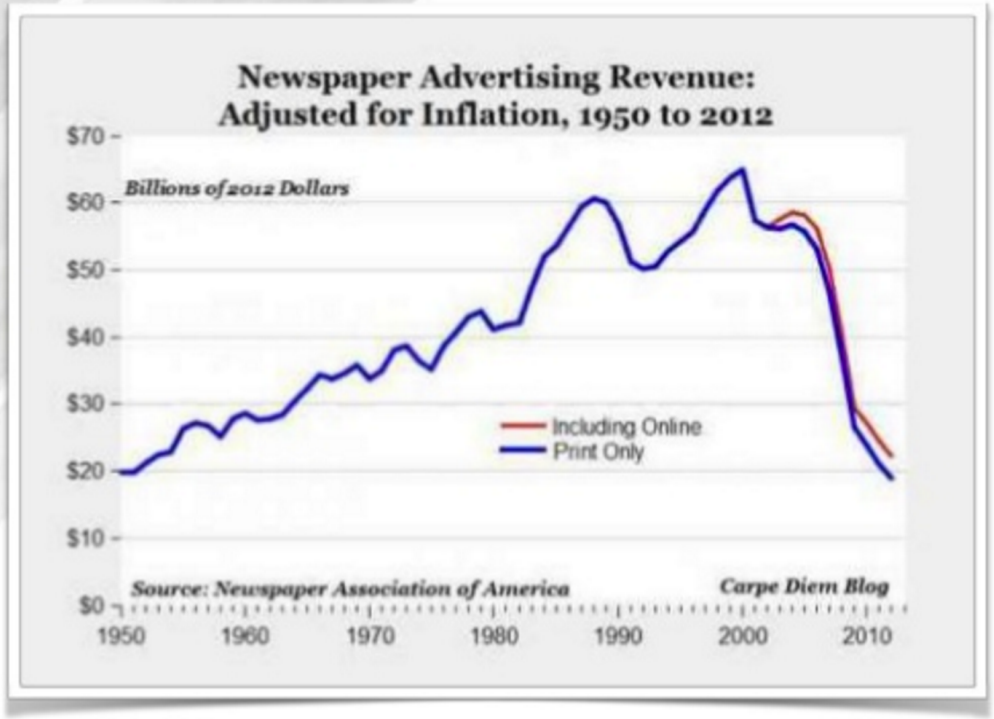
our profession is being disrupted

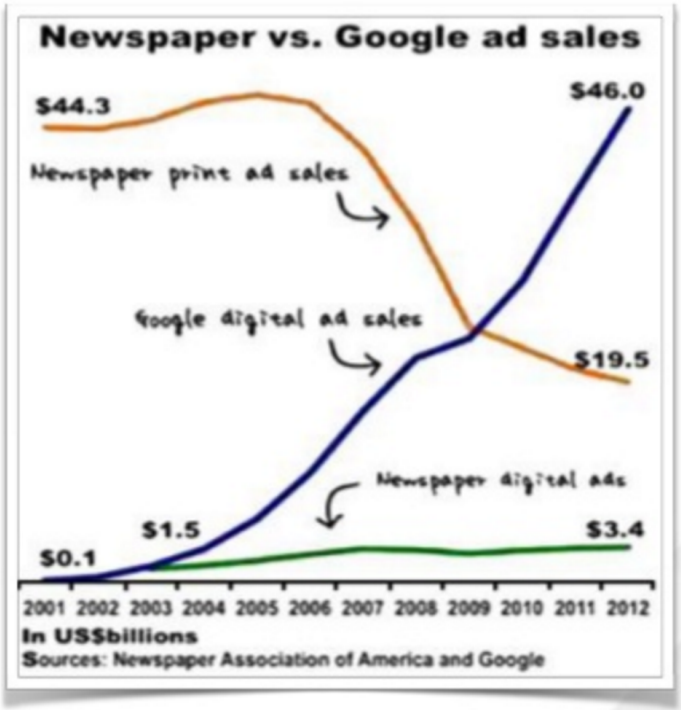
Moore's Law


CALCULATOINS PER SECOND PER $1000
10^{14}
10^{12}
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
0
10^{-2}
10^{-4}
ELECTROMECHANICAL
SOLID-STATE RELAY
VACUUM TUBE
TRANSISTOR
INTEGRATED CIRCUIT
HUMAN BRAIN
MOUSE BRAIN
CORE i7 QUAD
CORE 2 DUO
PENTIUM 4
PENTIUM III
PENTIUM II
OPTICAL, QUANTUM, DNA COMPUTING?
COMPAQ DESKPRO 386
PENTIUM
ALTAIR 8800
IBM 1130
IBM AT-80286
DEC PDP-1
IBM PC
UNIVAC I
DEC PDP-10
APPLE II
COLOSSUS
IBM SSEC
IBM 704
IBM TABULATOR
HOLLERITH TABULATOR
NATIONAL ELLIS 3000
BELL CALCULATOR MODEL 1
ANALYTICAL ENGINE
© BCA Research 2013
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025


THE INTERNET OF THINGS
AN EXPLOSION OF CONNECTED POSSIBILITY
BILLIONS OF DEVICES
YEAR
1992
1,000,000
2003
0.5 BILLION
2009
IoT INCEPTION
2010
8.7 BILLION
2012
11.2 BILLION
2014
14.4 BILLION
2015
16.2 BILLION
2016
22.9 BILLION
2018
34.8 BILLION
2017
28.4 BILLION
2018
42.1 BILLION
2020
50.1 BILLION

Newspapers


Newspaper Advertising Revenue:
Adjusted for Inflation, 1950 to 2012
Billions of 2012 Dollars
$70
$60
$50
$40
$30
$20
$10
$0
Including Online
Print Only
Source: Newspaper Association of America
Carpe Diem Blog
1950
1960
1970
1980
1990
2000
2010


Newspaper vs. Google ad sales
$44.3
$46.0
Newspaper print ad sales
Google digital ad sales
$19.5
Newspaper digital ads
$1.5
$3.4
$0.1
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
In US$billions
Sources: Newspaper Association of America and Google

LINEAR → EXPONENTIAL

1996

MarketCap: **$28B**

Employees: **140,000**

2012

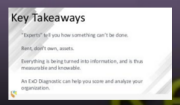
Bankrupt

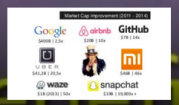
E[illegible]**7,000**

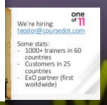
April 2012

MarketCap: **$1B**

Employees: **13**







Redefines: A Kodak Moment

The Second Machine Age and Water Management?

a case study in

vegetation and hydraulic resistance

Transformations .....

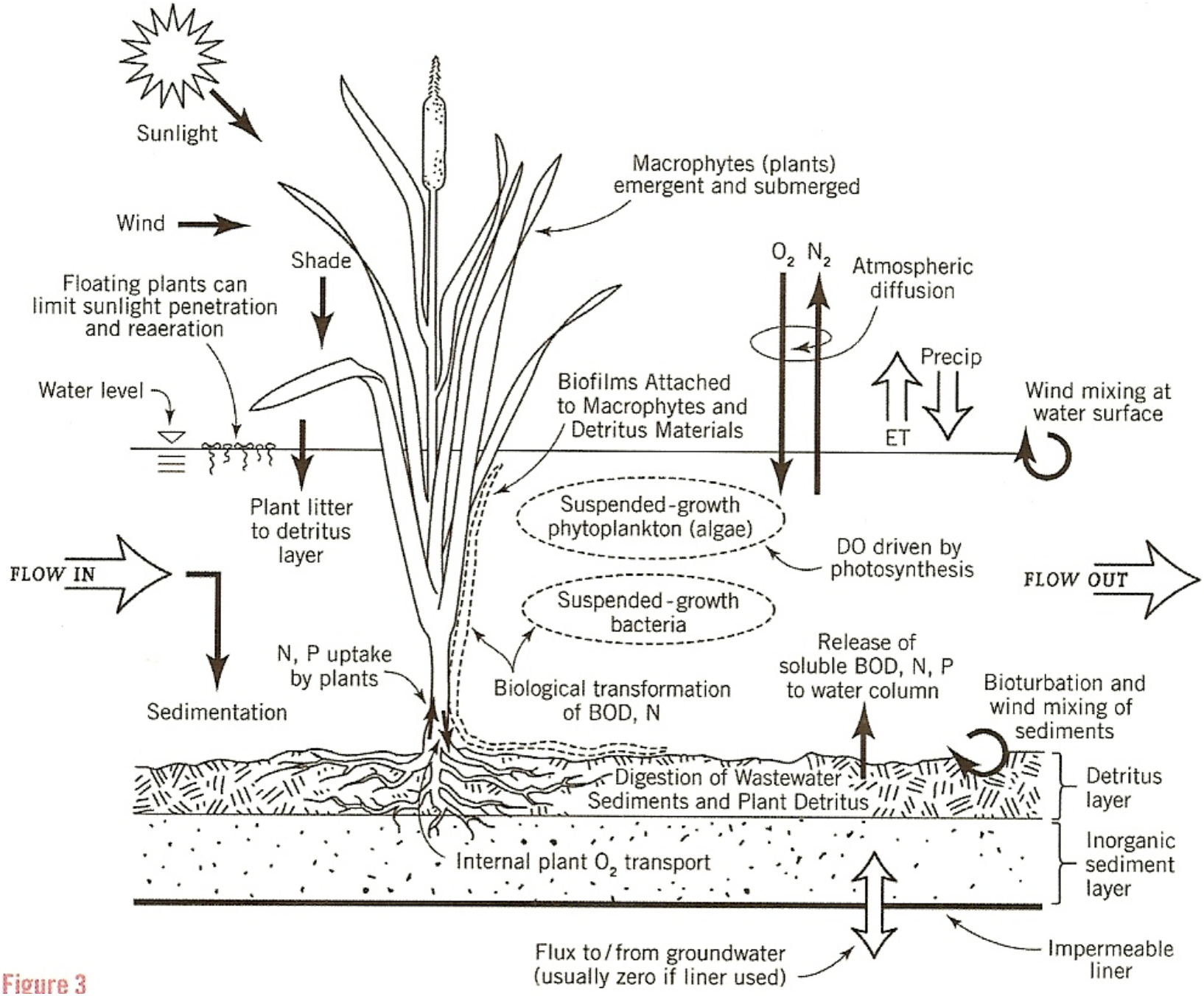

Sunlight
Wind
Shade
Floating plants can limit sunlight penetration and reaeration
Water level
Macrophytes (plants) emergent and submerged
O_2 N_2
Atmospheric diffusion
Precip
ET
Wind mixing at water surface
Biofilms Attached to Macrophytes and Detritus Materials
Plant litter to detritus layer
Suspended-growth phytoplankton (algae)
DO driven by photosynthesis
FLOW IN
FLOW OUT
Suspended-growth bacteria
N, P uptake by plants
Biological transformation of BOD, N
Release of soluble BOD, N, P to water column
Bioturbation and wind mixing of sediments
Sedimentation
Digestion of Wastewater Sediments and Plant Detritus
Detritus layer
Inorganic sediment layer
Internal plant O_2 transport
Flux to/from groundwater (usually zero if liner used)
Impermeable liner

Figure 3

Background

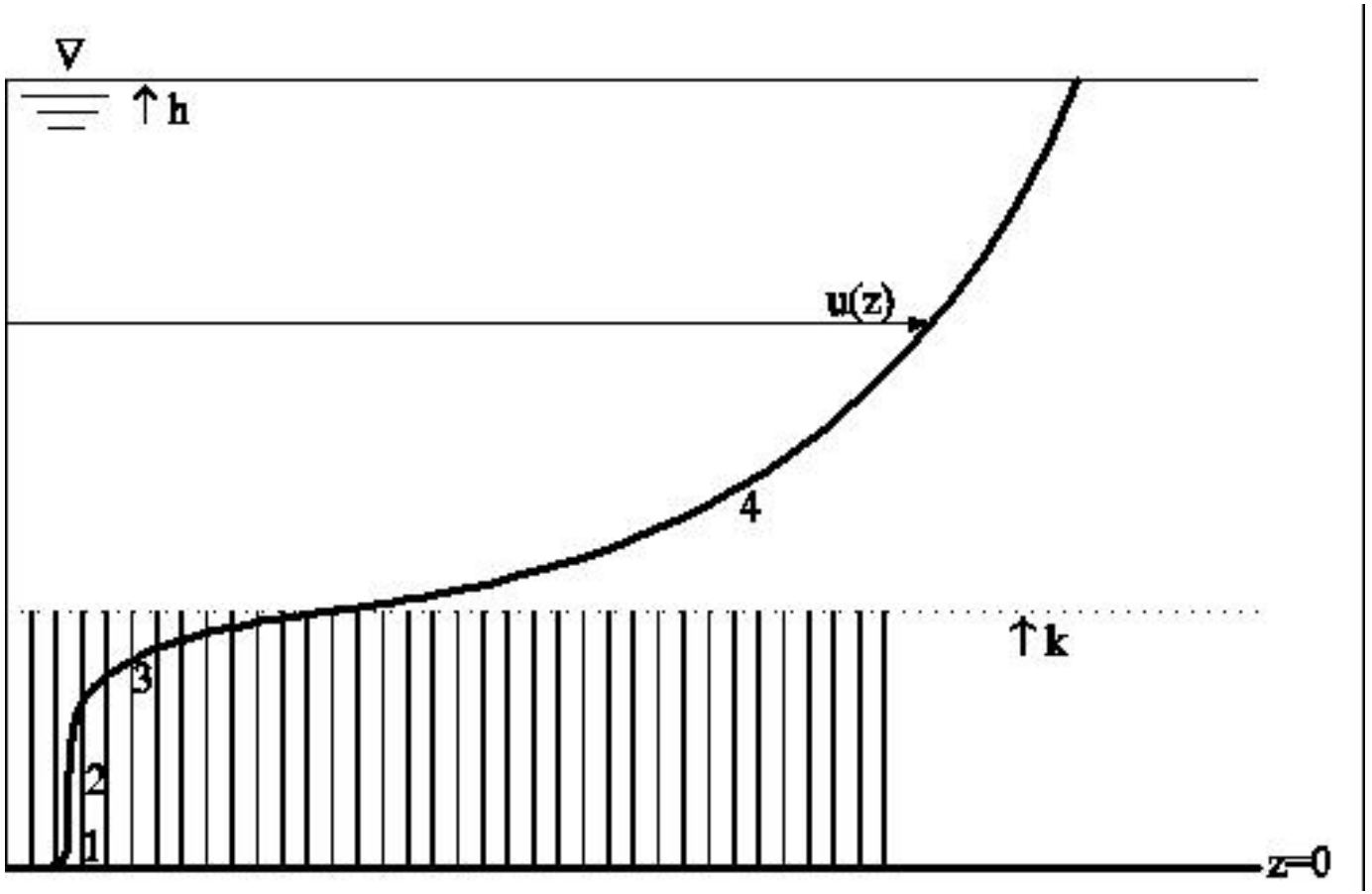
Antoine Chezy (1776) $\bar{u} = C\sqrt{Ri}$

Von Karman $u(z) = \frac{u_*}{\kappa} \ln\left(\frac{z}{z_0}\right)$

Nikuradze (1930) $z_0 = k_N/30$

White-Colebrook formula $C = 18\log\left(\frac{12R}{k_N}\right)$

Vertical velocity profile



Detailed Account of Roughness

1 DV Turbulence Model

Uittenbogaard (2003)

Detailed description

1 DV Turbulence Model

$$\rho_0 \frac{\partial u(z)}{\partial t} + \frac{\partial p}{\partial x} = \frac{\rho_0}{1 - A_p(z)} \frac{\partial}{\partial z} \left((1 - A_p(z)) \left(\nu + \nu_T(z) \right) \frac{\partial u(z)}{\partial z} \right) - \frac{F(z)}{1 - A_p(z)}$$

$$\frac{\partial u}{\partial x} = 0$$

$$\frac{\partial k}{\partial t} = \frac{1}{1 - A_p} \frac{\partial}{\partial z} \left(\left(1 - A_p \right) \left(\nu + \nu_T / \sigma_k \right) \frac{\partial k}{\partial z} \right) + T + P_k - B_k - \varepsilon$$

Detailed description

1 DV Turbulence Model

Drag force $$F(z) = \tfrac{1}{2}\rho_0 C_D(z) D(z) m(z) u(z) \left| u(z) \right|$$

Additional turbulence $$T(z) = F(z) u(z)$$

Turbulence production $$P_k = \nu_T \left(\frac{\partial u}{\partial z} \right)^2$$

Buoyancy $$B_k = -\frac{\nu_T}{\sigma_k} \frac{g}{\rho_0} \frac{\partial \rho}{\partial z}$$

Dissipation rate $$\frac{\partial \varepsilon}{\partial t} = \frac{1}{1 - A_p} \frac{\partial}{\partial z} \left(\left(1 - A_p\right) \left(\nu + \nu_T / \sigma_\varepsilon\right) \frac{\partial \varepsilon}{\partial z} \right) + P_\varepsilon - B_\varepsilon - \varepsilon_\varepsilon + c_{2\varepsilon} \frac{T}{\tau_{eff}}$$

Analytical Formulations

Baptist (2005)

Non-submerged vegetation

- Balance of horizontal momentum:

$$\tau_t = \tau_b + \tau_v$$

From which Chezy rougness is derived:

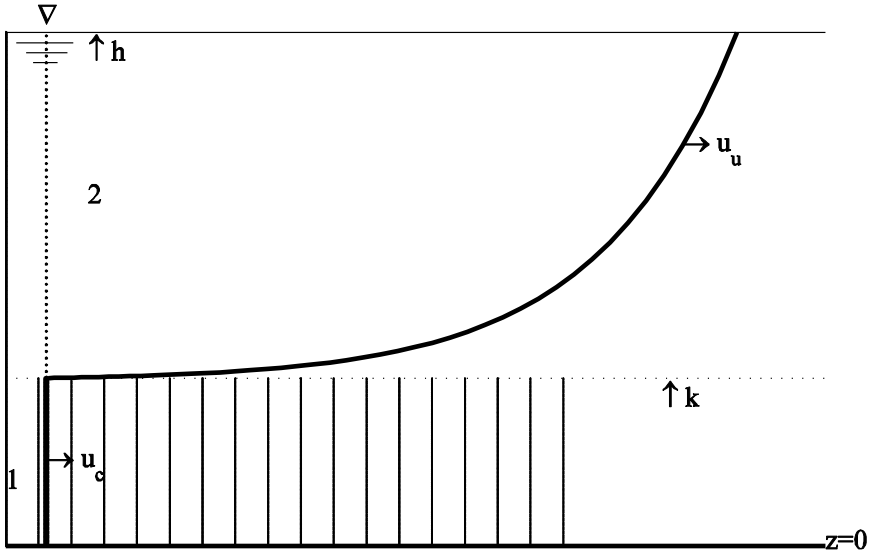
$$C_k = \sqrt{\frac{1}{\frac{1}{C_b^2} + \frac{C_D m D h}{2g}}}$$

or simplified

$$C_k = \sqrt{\frac{2g}{C_D m D h}}$$

Submerged vegetation

Method of Effective Water Depth (Baptist, 2005)

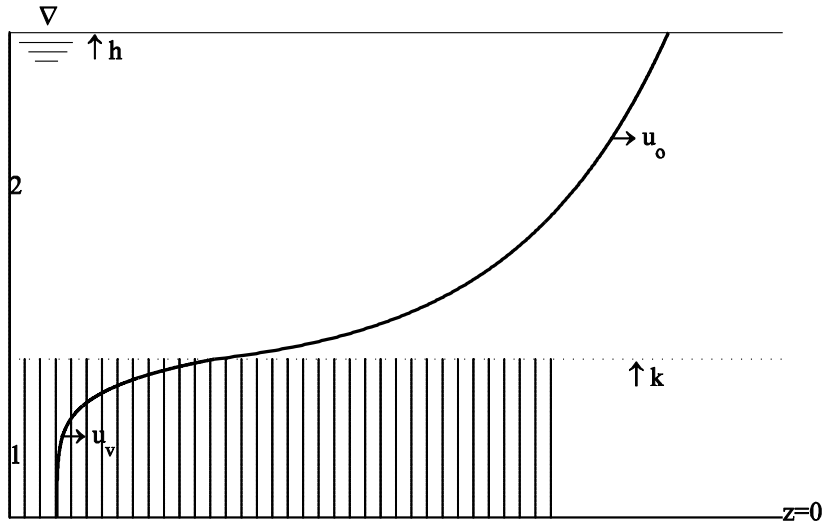


A

$$C_r = \sqrt{\frac{1}{C_b^{-2} + \frac{C_D m D k}{2g}} + \frac{(h-k)^{3/2}\frac{\sqrt{g}}{\kappa}\ln\left(\frac{h-k}{ez_0}\right)}{h^{3/2}}}$$

Submerged vegetation

Analytical Method (Baptist, 2005)

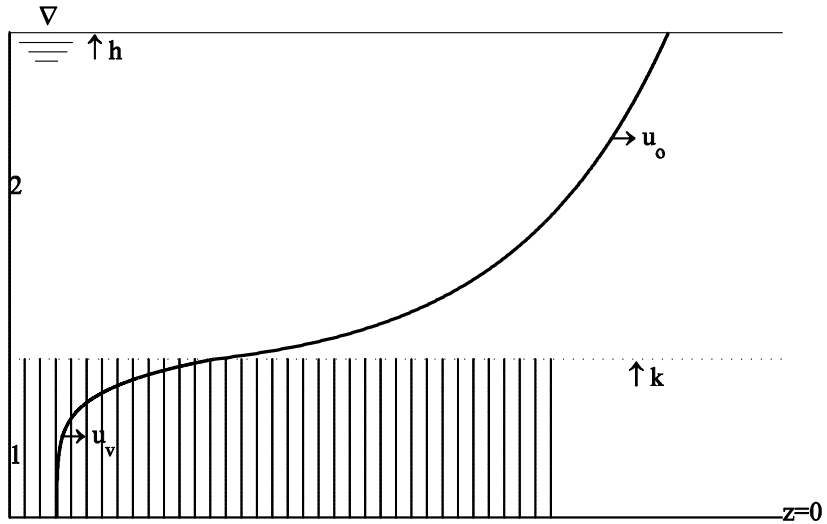


B

$$C_r = \frac{1}{h^{3/2}} \left\{ L \left[2\left(u_{vk} - \sqrt{a_v + u_{v0}^2} \right) + u_{v0} \ln \left(\frac{\left(u_{vk} - u_{v0} \right)\left(\sqrt{a_v + u_{v0}^2} + u_{v0} \right)}{\left(u_{vk} + u_{v0} \right)\left(\sqrt{a_v + u_{v0}^2} - u_{v0} \right)} \right) \right] \right.$$
$$\left. + \frac{\sqrt{g\left(h-k\right)}}{\kappa\left(h-k\right)} \left[\left(h-d\right) \ln\left(\frac{h-d}{z_0} \right) - \left(k-d\right) \ln\left(\frac{k-d}{z_0} \right) - \left(h-k\right) \right] \right\}$$

Submerged vegetation

Analytical Method (Baptist, 2005)



Where:

$$a = \frac{2Lg\left(h-k\right)i}{c_p \mathrm{l} \exp\left(\frac{k}{L}\right)}$$

$$d = k - \int_0^k \frac{\exp\left(\frac{z}{L}\right)}{\exp\left(\frac{k}{L}\right)} \mathrm{d}z = k - L\left(1 - \exp\left(-\frac{k}{L}\right)\right)$$

$$z_0 = \left(k-d\right)\exp\left(-\kappa\sqrt{\frac{2L}{c_p \mathrm{l}}\left(1+\frac{L}{h-k}\right)}\right)$$

$$c_p = \frac{1}{20}\frac{h-k}{\mathrm{l}}$$

Symbolic Regression

Knowledge Discovery in Petabyte Age

The Age of Information Overflow

- Development of technologies for data acquisition, storage and transmission continues at a breakneck pace
- The same cannot be asserted about extraction of useful information and knowledge from data

- *What is to be done with all these data?*

Data Mining for Scientific Discovery

- Philanthropic activity (without major material satisfaction)
- Every scientific discipline holds a vast amount of knowledge that should not be just ignored (black boxes are not tolerated)
- Combination of theory-driven (understanding-rich) and data-driven discovery process

3rd Kepler's Law of Planetary Motion

Tycho Brahe collected observables

- READINGS OF THE BOOK OF NATURE ITSELF

Kepler analysed the data

- MODEL OF THE PROCESS - ECONOMY OF THOUGHT

Evolutionary Algorithms

- ☆Loosely based on Darwin's theory of evolution
- ☆Survival of the fittest
- ☆Genetic operators
  - ☆Reproduction (crossover)
  - ☆Mutation
  - ☆Selection

Evolutionary Criteria

- Criterion of Heredity

  Offspring are similar to their parents: the copying process maintains high fidelity

- Criterion of Variability

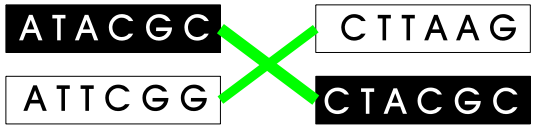
  Offspring are not exactly the same as their parents or each other: the copying process is not perfect

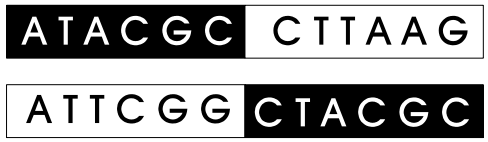
- Criterion of Fecundity

  Variants leave different number of offspring: specific variations have an effect on behaviour, and behaviour has an effect on reproductive success

Evolutionary Algorithms
Crossover

- Select two parental strings according to the fitness criterion
- Randomly select crossover site
- Swap the two sub-strings





Evolutionary Algorithms
Crossover (1)

THISISTHEWESTMESSAGE

THINISTHEBESTMESSAGE

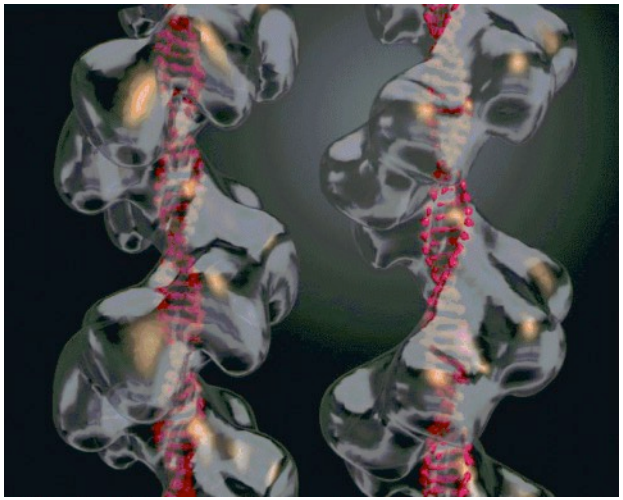
Evolutionary Algorithms
Crossover (2)

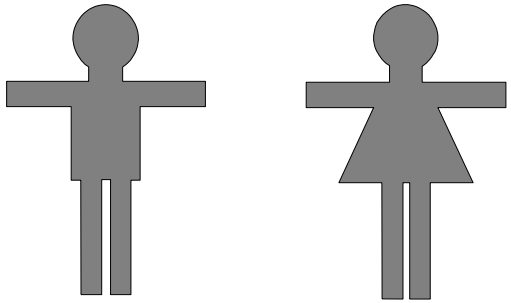
Evolutionary Algorithms
Crossover (3)

THISISTHEBESTMESSAGE

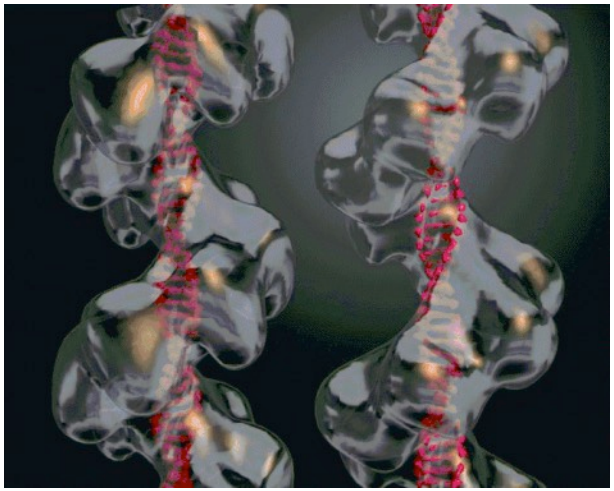
THINISTHEWESTMESSAGE

Natural Evolution





Genetic Programming



$$\frac{\partial x}{\partial t} + \frac{\partial y}{\partial t} + \frac{\partial z}{\partial t} = 0$$

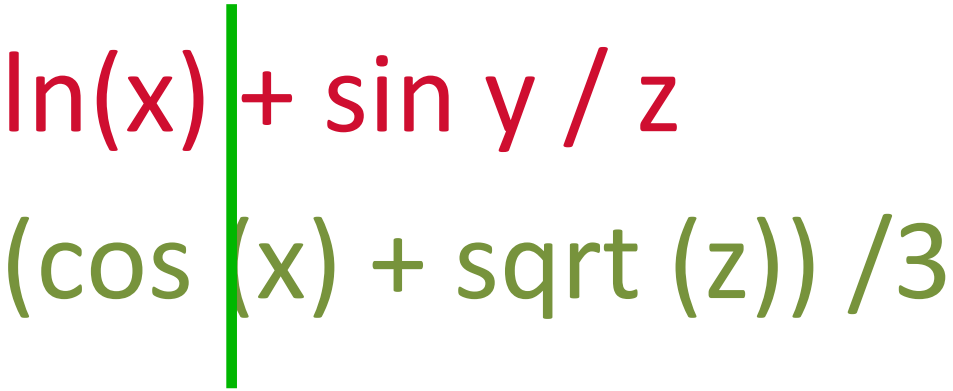
$$\sin\theta + \ln\lambda - \sum_{t=0}^{\rho} x^t$$

Genetic Programming
Crossover (1)

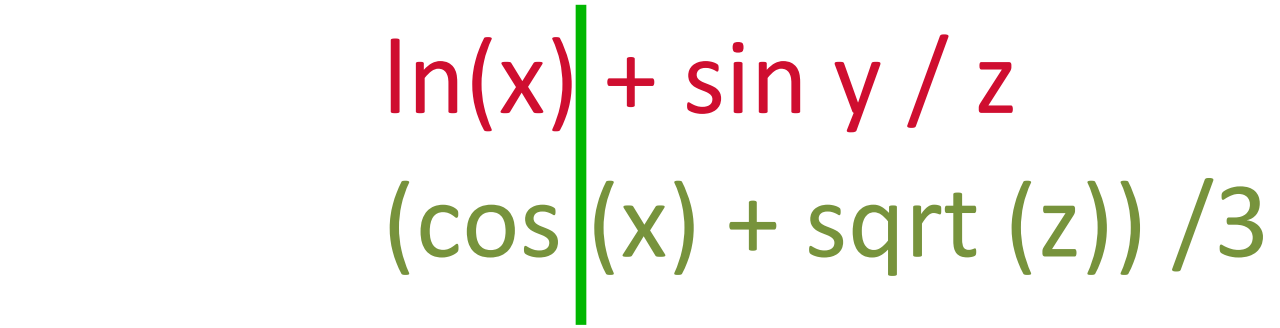
ln(x) + sin y / z

(cos (x) + sqrt (z)) /3

Genetic Programming
Crossover (2)



Genetic Programming
Crossover (3)



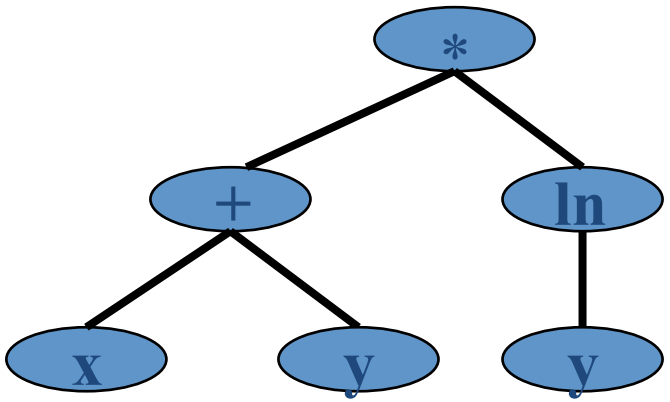
ln(x) (x) + sqrt (z)) /3

(cos + sin y / z

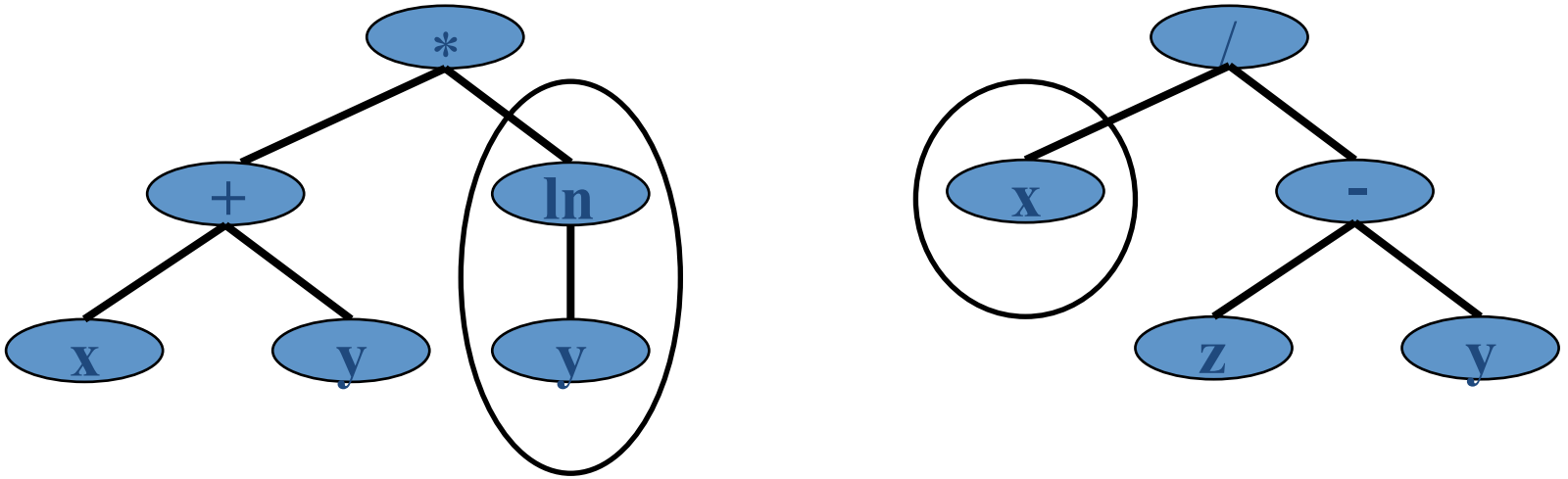
Genetic Programming

- Parse Trees
  - –Function Set
  - –Terminal Set
  - –Constants
- Symbolic Regression
  - –Find a relationship between:
  - –dependent
  - –independent variables

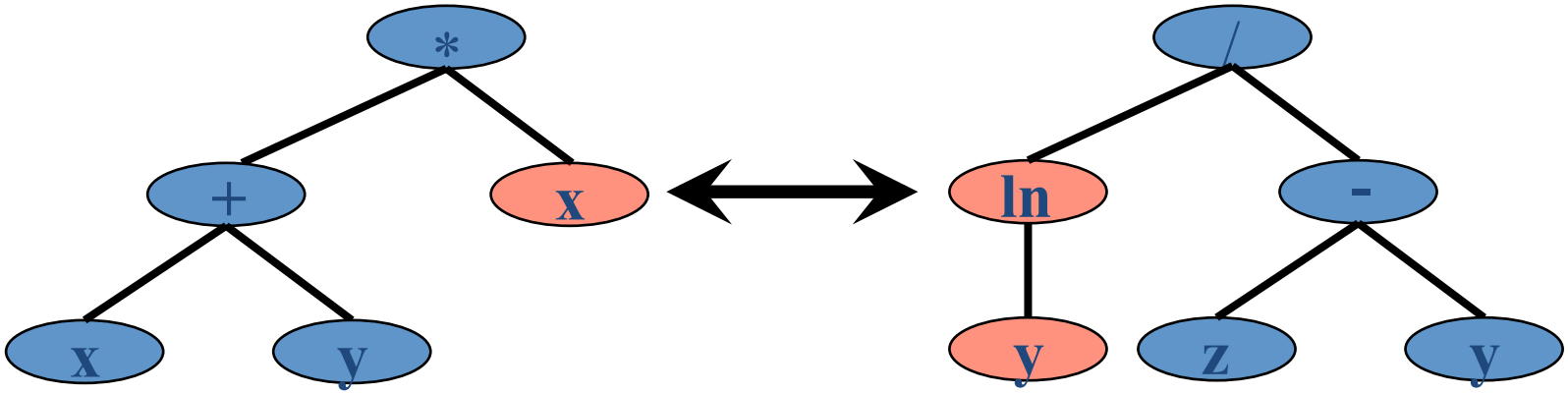
(x + y) ln(y)



Breeding: Crossover



Breeding: Crossover



Evolutionary Algorithms
Mutation

- Randomly select parental string
- Randomly alter one of the symbols

THISISTHEWESTMESSAGE

THISISTHEWESRMESSAGE

Genetic Programming
Mutation

① Branch-mutation

» a complete sub-tree is replaced with another arbitrary sub-tree

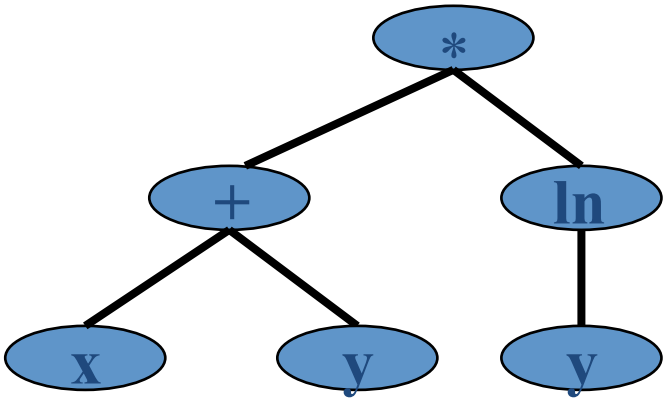
② Node-mutation

» a random change to a single node with another, arbitrary node of a same arity

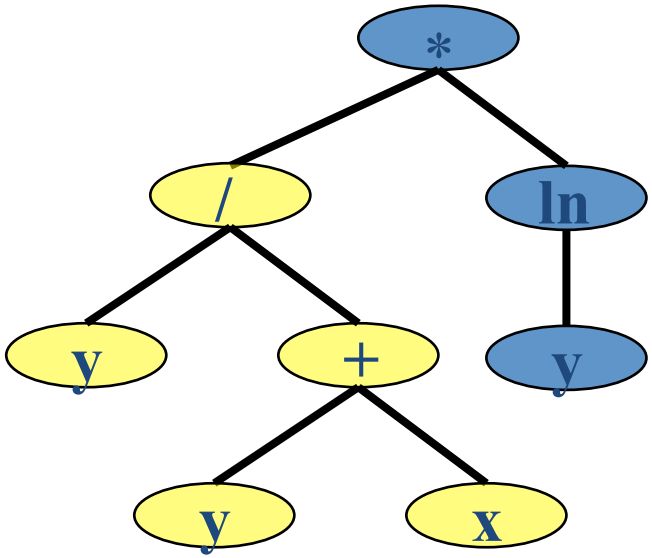
③ Inversion-mutation

» inversion of an order in which operands are ordered in an expression, for example f(x, y, z) may become f(z, x, y)

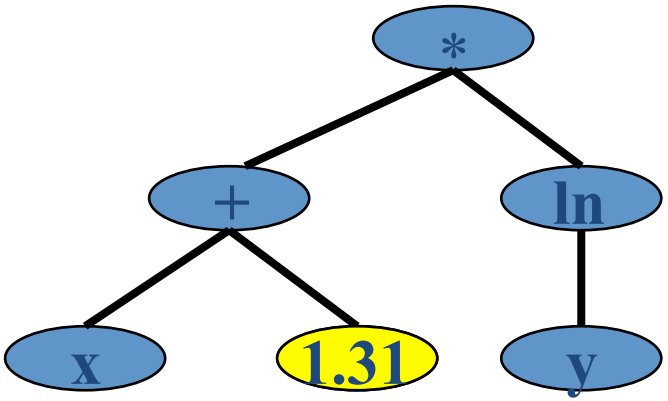
Breeding: Branch Mutation



Breeding: Branch Mutation



Breeding: Constant Mutation



- White noise
- Small uniform steps
- Occasional large jumps
- Real vector optimization (gauss-Newton, evol. Strategies)

Genetic Programming
Selection

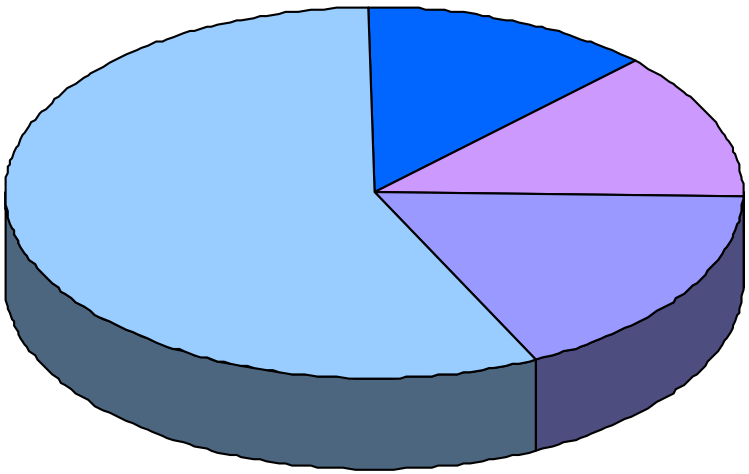
Gives direction to evolutionary search process

Favours more fit entities by selecting them more often

‘Survival of the fittest’

Genetic Programming
Selection

- Biased Roulette Wheel
- Areas are proportional to individual fitnesses
- Selection is random, but still promoting more fit individuals
- Balance between homogenising selection and generation of variation



Genetic Programming Fitness

$z = \frac{x^2}{y}$ $z = x^3 + y^2$ $z = x^2 + y$

x	y	z	Calculated Z		
15.0	45.0	1289.0	5.0	5400.0	2250.0
35.0	32.0	354.0	38.3	43899.0	2249.0
0.0	8.00	51381.0	0.0	64.0	64.00
5.0	51.0	518.0	0.49	2726.0	2626.0
		ERRORS:	21837.0	101181.0	56281.0
		FITNESSES:	6.67E-06	1.66E-07	6.12E-07

‘Fitness’ can be any measure (RMSE, NMSE, Classification Accuracy)

Evolutionary Algorithms pseudo code

1.) Randomly create an initial population

2.) Iteratively perform the following:

(a) evaluate fitness of each individual

(b) select ‘parents’ on the basis of fitness

(c) perform (sexual) recombination

(d) mutate

3.) Best so-far individual

Why Symbolic Regression?

- Approximate a relationship *y = f(x) b*y minimizing error where *f* is built of a number of primitive functions
- Other regression methods require pre-specified functional form
  - Linear regression
  - Nonlinear regression (including ANN and SVM)
  - Coefficient fitting
- GP manipulates symbolic information
  - Hypothesis: GP can lead to *interpretable* equations

Caveats

- When uncontrolled
  - Tree size grows
  - Referred to as bloat
- Possible solutions
  - Parsimony Pressure
  - Short formulations higher fitness than long
  - Length as penalty
  - Length as second objective
  - …

- Is there an approach one can follow in order control this and similar problems?

```
•(((((0.18916 / 0.098799) -
sqrt((((((x_1 / 0.22097) -
sqrt(0.22097)) -
sqrt(sqrt((((((x_1 / 0.22097) -
sqrt(0.019076)) - sqrt(0.28856)) +
x_1) - sqrt(0.22097)))))) + 0.22097)
- sqrt(0.22097)))) + -0.034116) -
sqrt(x_1)) + (((((((((0.017471 -
0.071534) - sqrt((((((x_1 / 0.22097)
- sqrt(0.22097)) - sqrt(0.28856)) +
x_1) - sqrt(0.22097)))) + x_1) -
sqrt(x_1)) + (((((x_1 / 0.22097) -
sqrt(x_1)) - sqrt(x_1)) + x_1) -
sqrt(0.017471))) - sqrt((((((x_1 /
0.22097) - sqrt(0.019076)) -
sqrt(0.28856)) + x_1) -
sqrt(0.22097)))) - sqrt(x_1)) + x_1)
- 0.017471))
```

Genetic Programming in Natural Sciences

Can we utilise already existing knowledge in order to enhance interpretability and avoid bloat?

Genetic Programming

- Programming computers by telling them what to do, instead telling them how to do it
- But, how do we say to computer what we want?
- And what is that we want anyway?

What is that we want from a learning machine?

Find a good fit, do not worry about inconsistencies

- Transform the problem to the one without units of measurements (Buckingham's Pi theorem) and then find a good fit

Find a good fit, do not accept incorrect or inconsistent solutions

Find a good fit with least possible inconsistencies

Find a good fit only

- Standard GP, neural networks, polynomial fits, splines, SVM, RVM...
- Problem is posed in either raw or dimensionless form
- If we are lucky we can interpret GP expressions
- In similar sense one can attempt to interpret other approximations

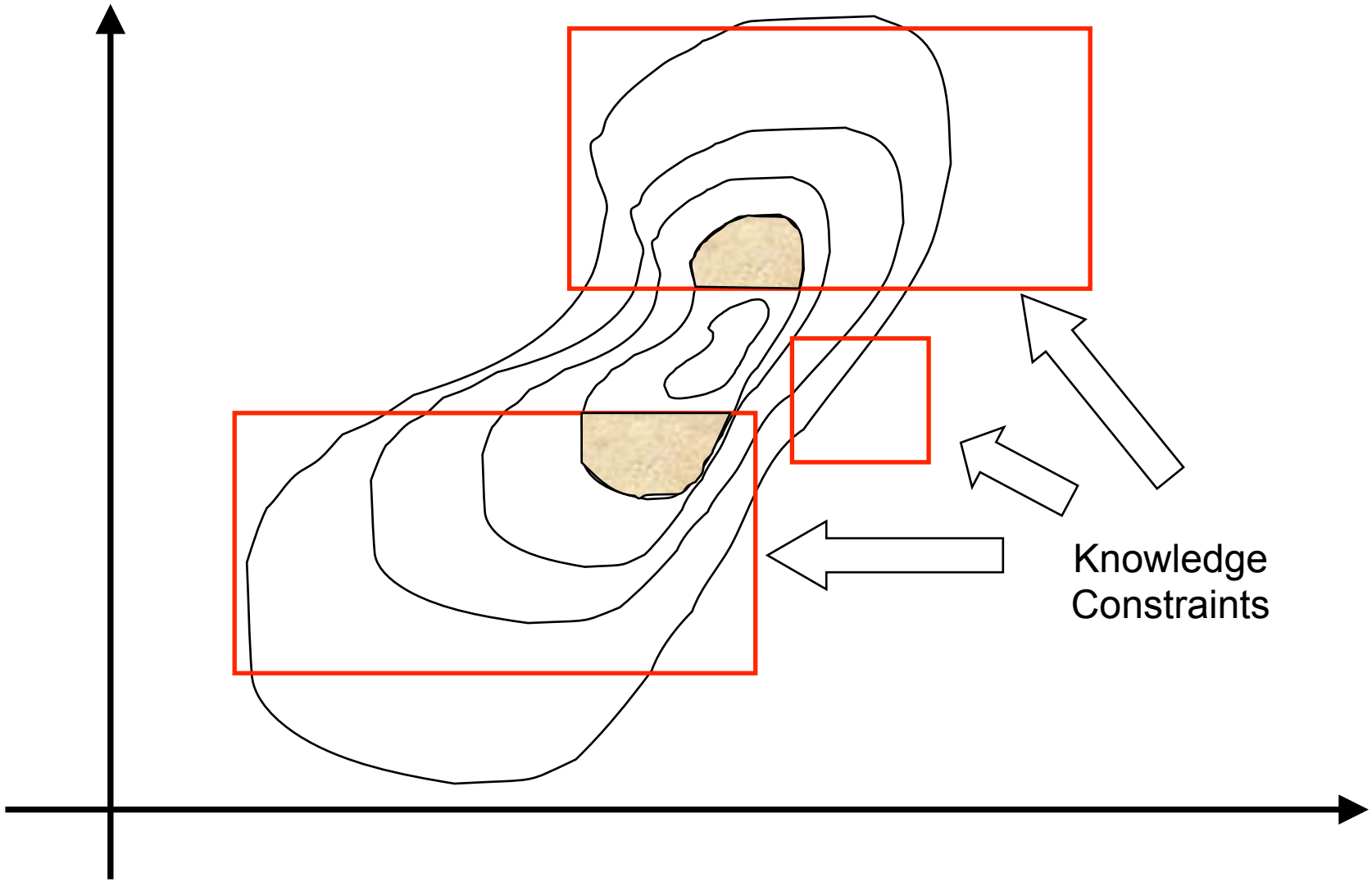
Check for inconsistencies using background knowledge - Bias

- Bias refers to any kind of basis for choosing one generalisation over another, other than strict consistency with the observed training examples
- Bias is based on knowledge about the domain
- Bias can be provided in strong (declarative) and weak (preferential) form

Declarative Bias

- Strongly Typed GP (STGP)
- Declarative bias takes form of constraints and strict adherence with constraints is fully enforced
- Violation of constraints eliminates proposed equation altogether (all or nothing)
- Indispensable for well defined problems
- Dramatically reduces search space
- But, it also breaks down search space resulting in a loss of diversity

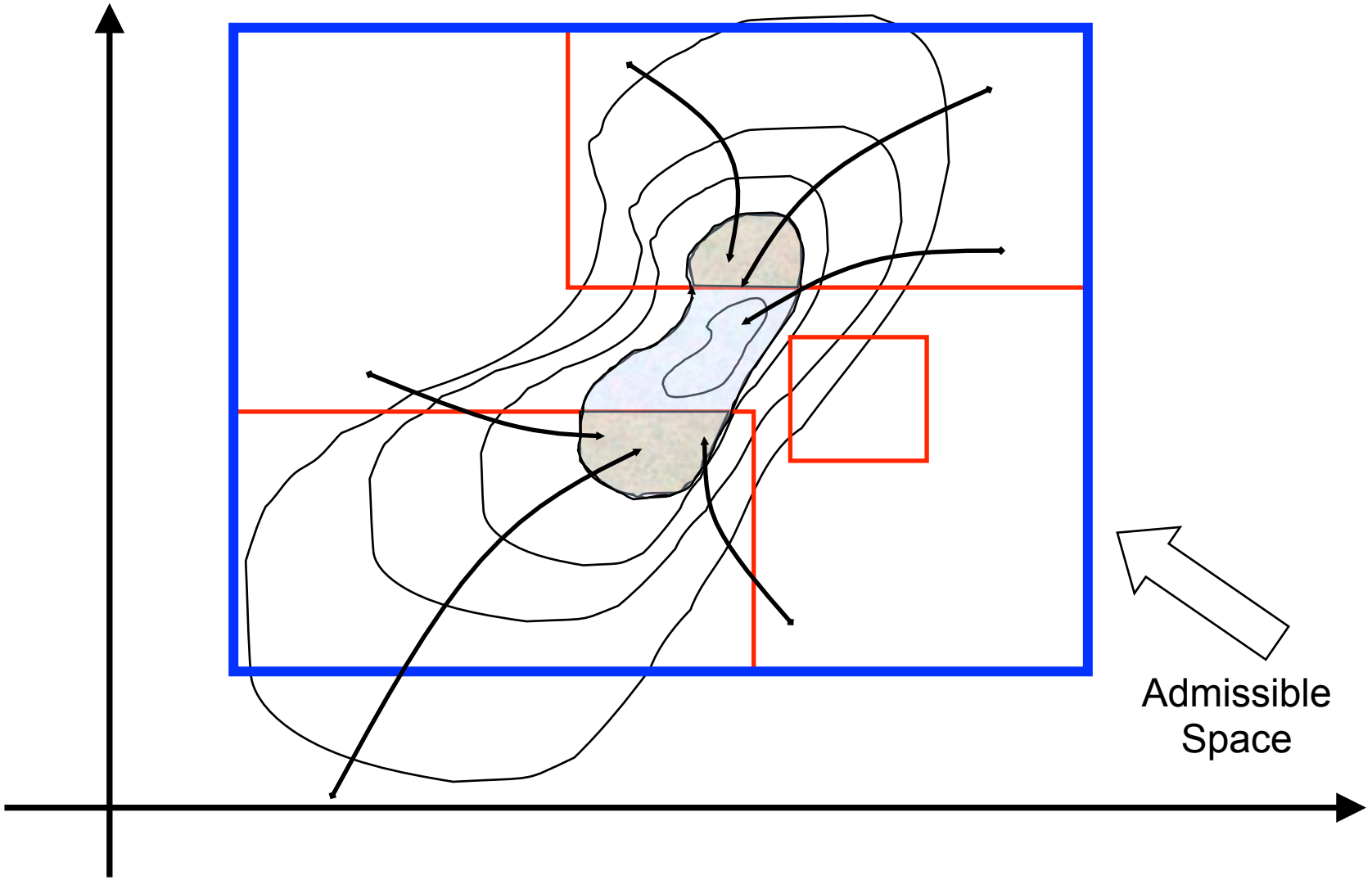
Broken ergodicity



Preferential bias

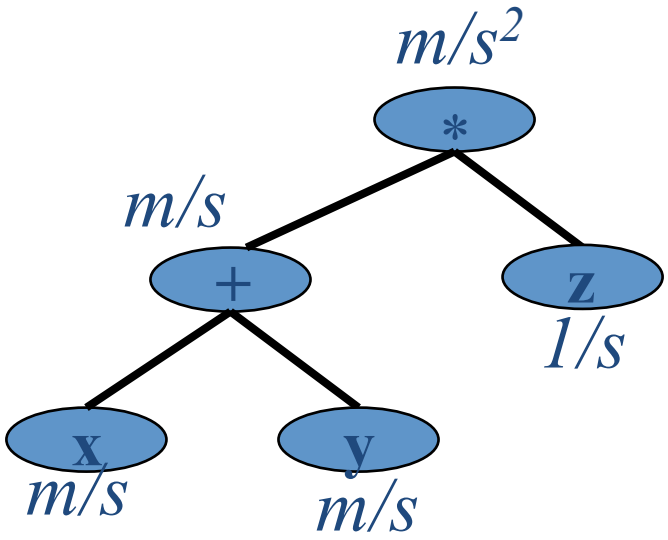
- Introduces knowledge in a form of objective function(s)
- It does not guarantee strict adherence with background knowledge
- It only defines preference towards equations which more closely satisfy requirements imposed by background knowledge

Preferential bias

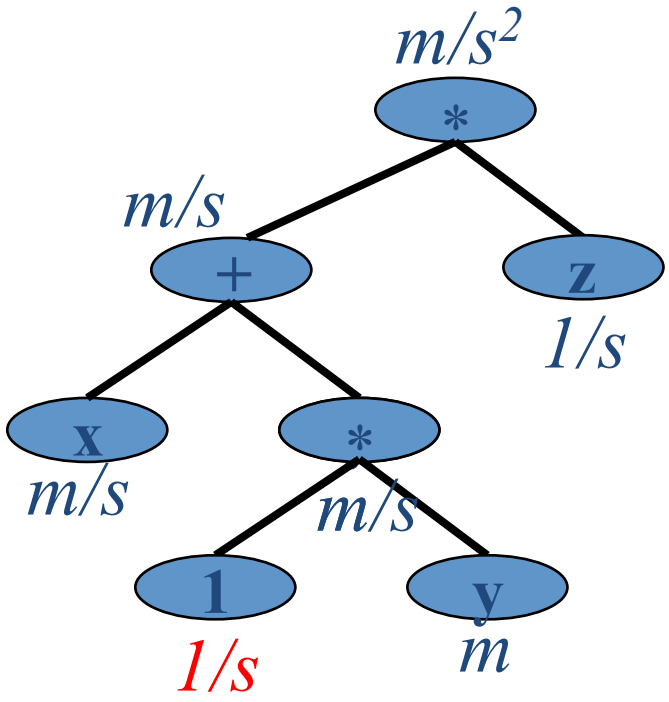


Bias in the form of units of measurements

- Why restrict to dimensionless values?



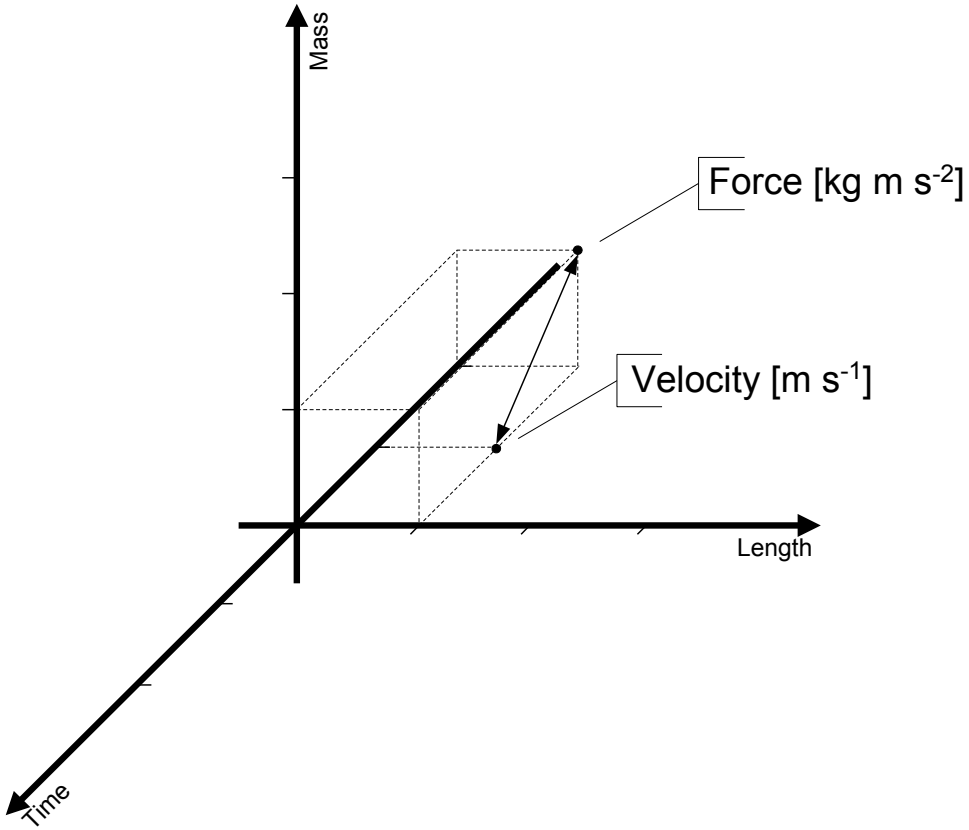
Coercion



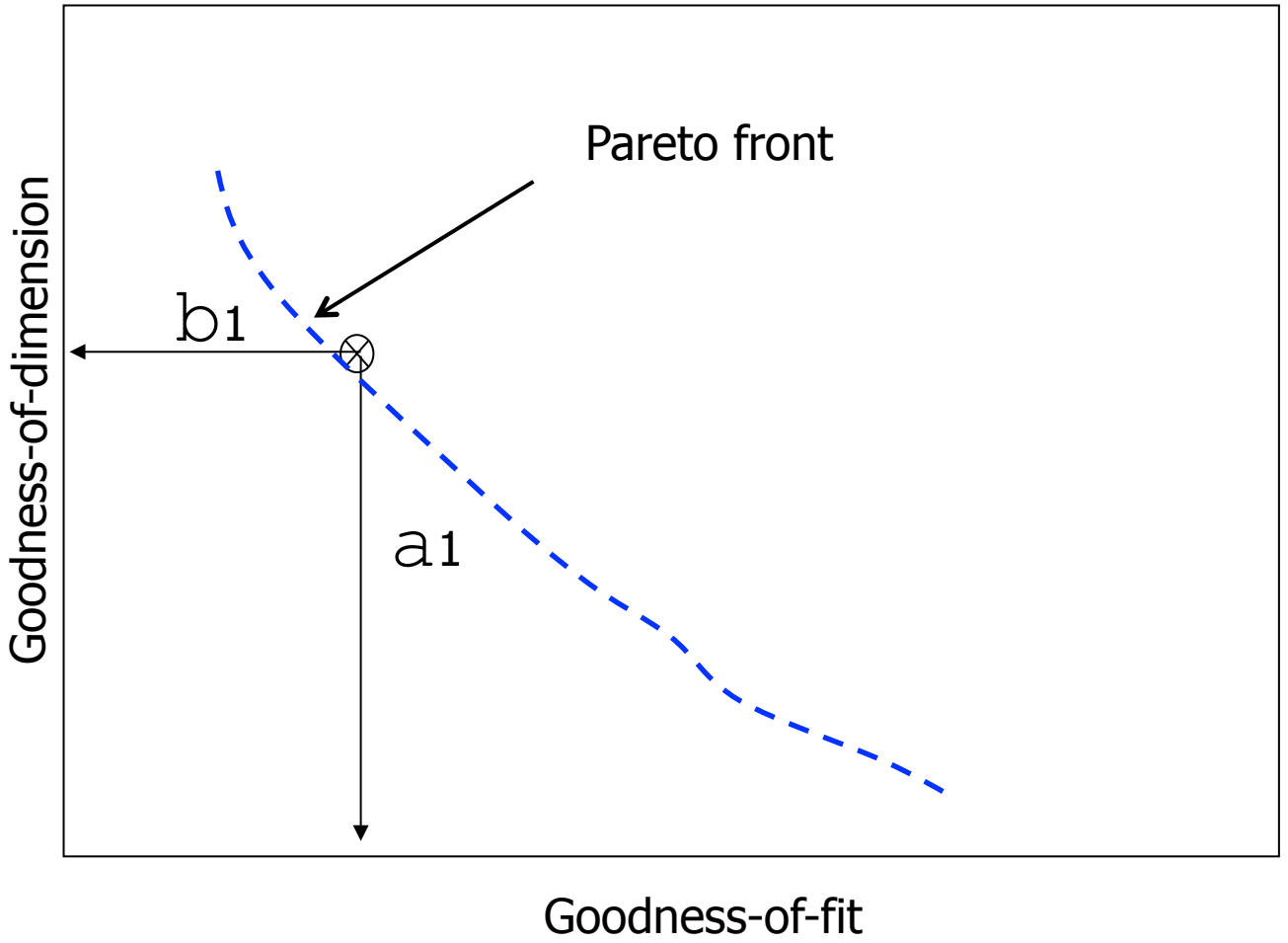
Dimensionally Aware GP

(Babovic & Keijzer 1998)

- Counts number of *coercions* needed
  - –multiplication with unity of needed transformation
- Pareto Balance
  - –Goodness of fit
  - –Goodness of dimension



Dimensionally Aware GP: the idea



What You Get

RMS	CoD	UnitError	FitnessPerNode	Hypothesis
0.044012	0.845497	0	0.230786	cb = (5.56558445e-002 * (((Uf' / (-Uf' - Uf)) * (6.60554916e-002 * -(ws - Uf'))) / ((-(((ws - ((w
0.044042	0.846303	0	0.25133	cb = (-((ws / ((32.0675392 / Uf') * d50)) / ((30.6716137 / (Uf' * (--(Uf' - ws) / (Uf' - -Uf)))) * -U
0.044133	0.846303	0	0.262919	cb = ((-(Uf' * 0.302570283) / (Uf' - -Uf)) * ((ws / ((32.0675392 / ((Uf' - ws) * 0.106477648)) *
0.044143	0.845912	0	0.223538	cb = ((-d50 * ((-ws / ((53.9560928 * d50) * (g / (Uf' - -Uf)))) + ((60.3299637 * ((d50 / Uf') * g
0.044403	0.845827	0	0.230876	cb = ((-d50 * ((-ws / ((53.9560928 * d50) * (g / (Uf' - -Uf)))) + (-ws / Uf'))) * (0.694262862 / -
0.04542	0.835046	0	0.274278	cb = (2.39495747e-003 * ((-13.870719 + -(-(ws / (g * -d50)) * (ws + -Uf'))) / 4.9209795))
0.045739	0.830503	0	0.211226	cb = ((Uf' / d50) * (0.175336733 * (0.368466854 * (0.144605324 * -(0.167134091 * (0.3684

- *Judge* which hypothesis balances the objectives best
- Rearrange terms
- Set up ‘little theory’
- Use results to gain insight

Back to Roughness - Data

990 ‘observations’ generated by the 1 DV model turbulence model

(75% training, 25% validation)

Input	*Dimension*	Description
D	**L**	**Diameter of the stems.**
m	**L^{-2}**	**Number of stems per square meter**
k	**L**	**Vegetation height.**
C_D	**-**	**Drag coefficient of a single stem.**
C_b	**$L^{0.5}/T$**	**Bed Chézy resistance coefficient.**
h	**L**	**Water depth.**

177 flume experiments data points (out of sample validation)

GP Result

$$C_r = \sqrt{\frac{2g}{c_D m D k}} + 2\sqrt{g}\ln\left(\frac{h}{k}\right)$$

Non submerged conditions

C

Logarithmic term

The fun is only starting

$$C_r = \sqrt{\frac{2g}{c_D m D k}} + 2\sqrt{g}\ln\left(\frac{h}{k}\right)$$ D

$$C_r = \sqrt{\frac{1}{\frac{1}{C_b^2} + \frac{1}{2g} C_D m D k}} + \frac{\sqrt{g}}{\kappa}\ln\left(\frac{h}{k}\right)$$

Differential Form

$$C_r = \sqrt{\frac{1}{\frac{1}{C_b^2} + \frac{1}{2g} C_D m D k}} + \frac{\sqrt{g}}{\kappa} \ln\left(\frac{h}{k}\right)$$

$$C_r = C_1 + \frac{\sqrt{g}}{\kappa} \ln\left(\frac{h}{k}\right)$$

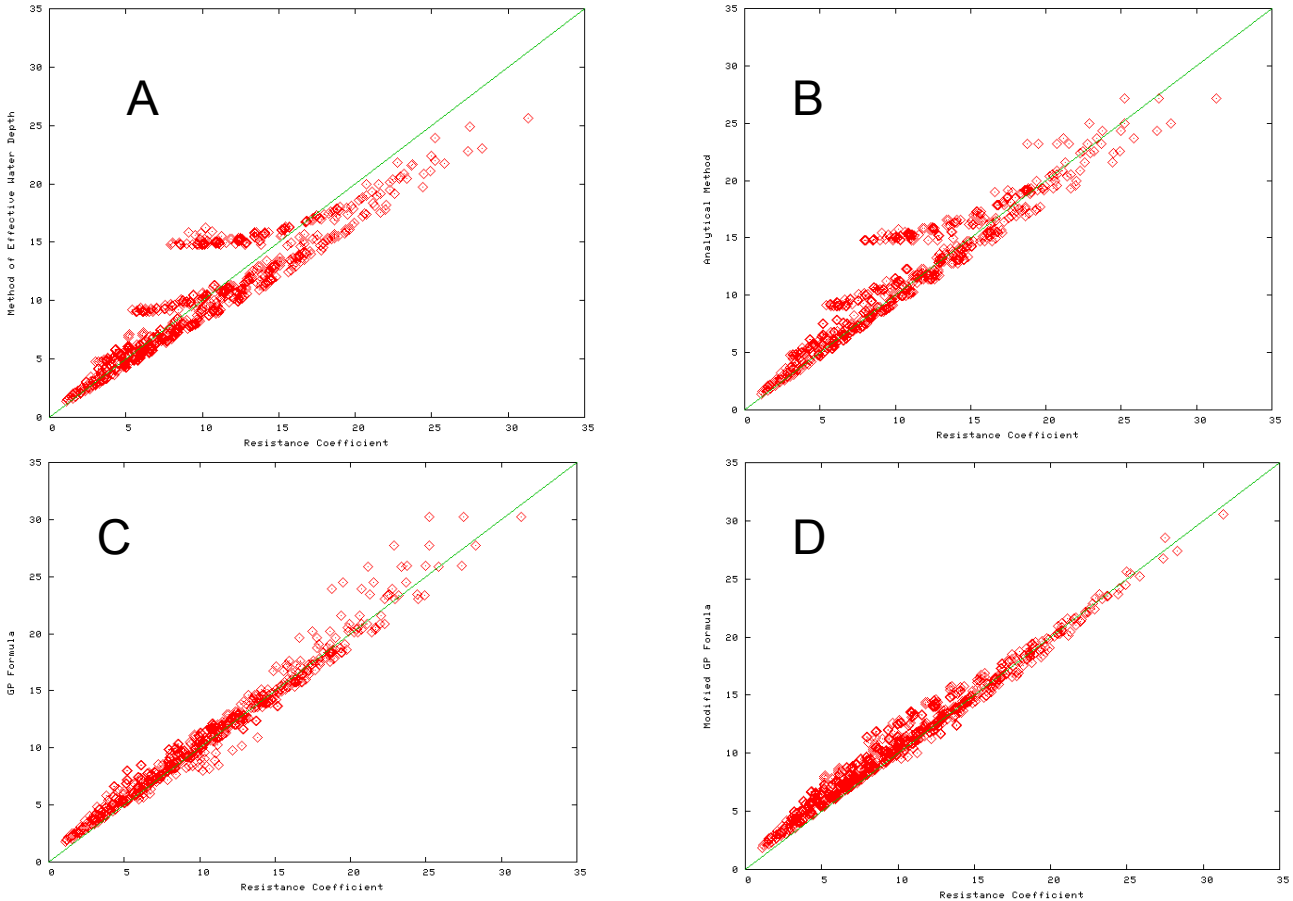
$$\frac{\partial C_r}{\partial h} = \frac{\sqrt{g}}{\kappa h}$$

How Good is All This?

Equation	*RMSE 1-DV Data [$m^{1/2}s^{-1}$]*	*CoD 1-DV Data[-]*
Method of effective water depth (A)	**2.2**	**0.831**
Analytical solution method (B)	**2.06**	**0.892**
Original GP-formula (C)	**0.98**	**0.971**
Modified GP-formula (D)	**1.13**	**0.977**
1-DV numerical model	**0**	**1.00**

1 DV validation data

How Good is All This?



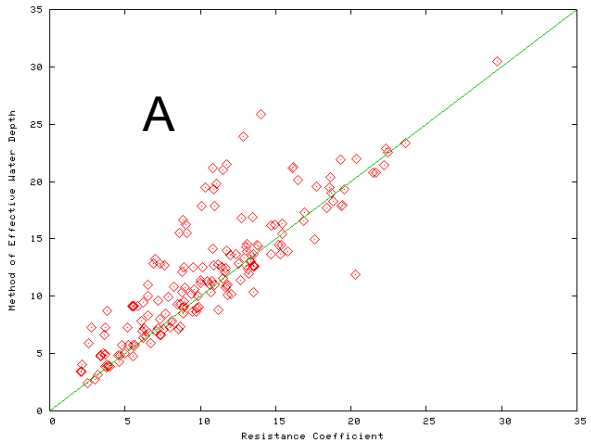
1 DV validation data

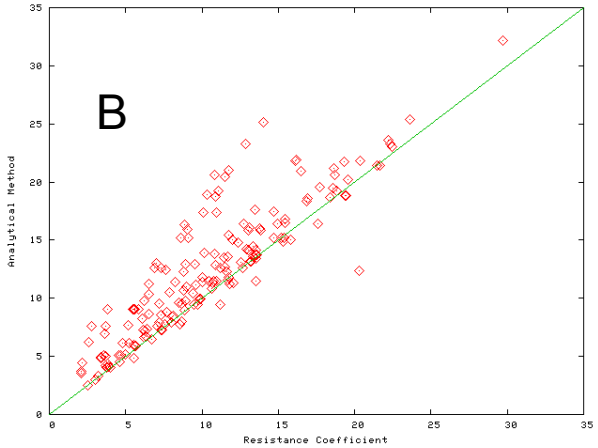
How Good is All This?

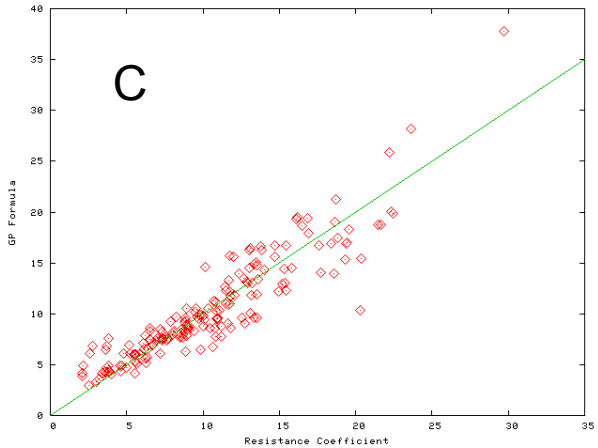
Equation	*RMSE 1-DV Data [$m^{1/2}s^{-1}$]*	*CoD 1-DV Data[-]*
Method of effective water depth (A)	3.17	0.734
Analytical solution method (B)	3.15	0.787
Original GP-formula (C)	2.11	0.838
Modified GP-formula (D)	2.11	0.872
1-DV numerical model	1.86	0.873

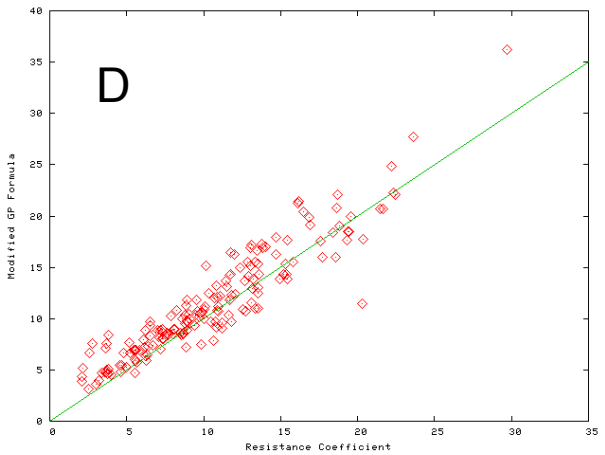
Flume data

How Good is All This?



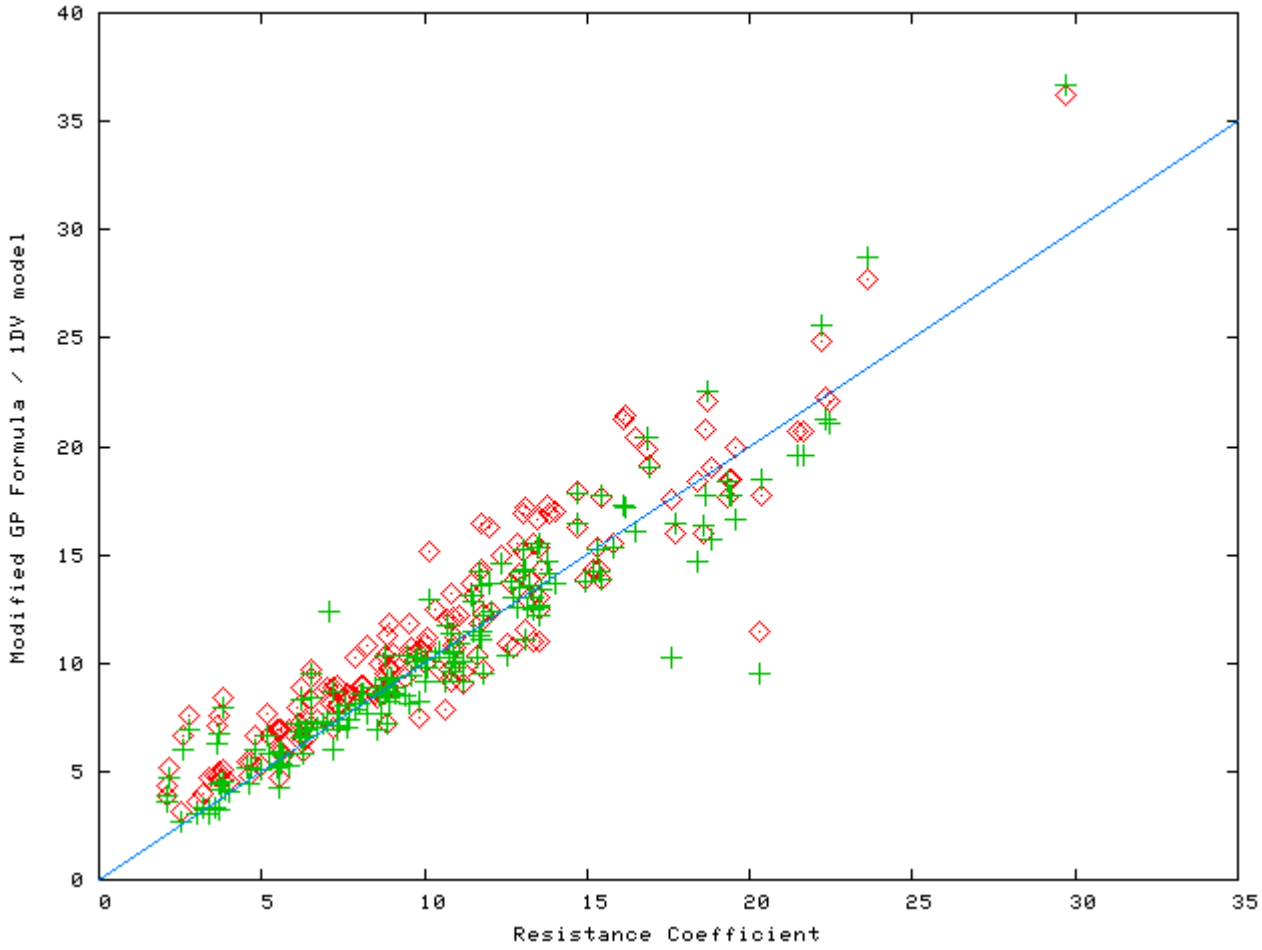




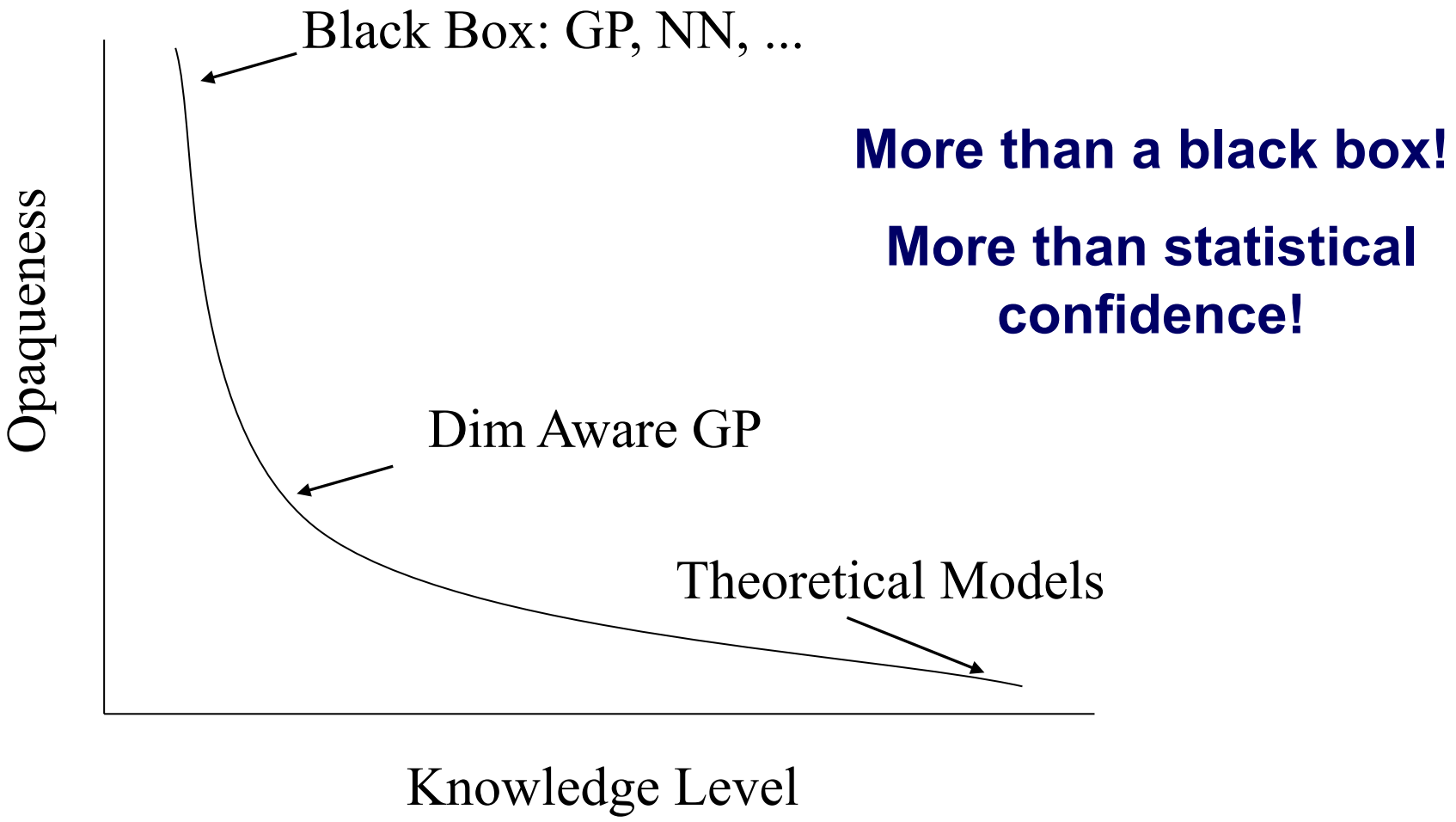


Flume data

Eq. D vs. 1DV


Modified GP Formula / 1DV model
Resistance Coefficient
0
5
10
15
20
25
30
35
40

Model Induction Trade-Off



More than a black box!

More than statistical confidence!

Computational Scientific Discovery?

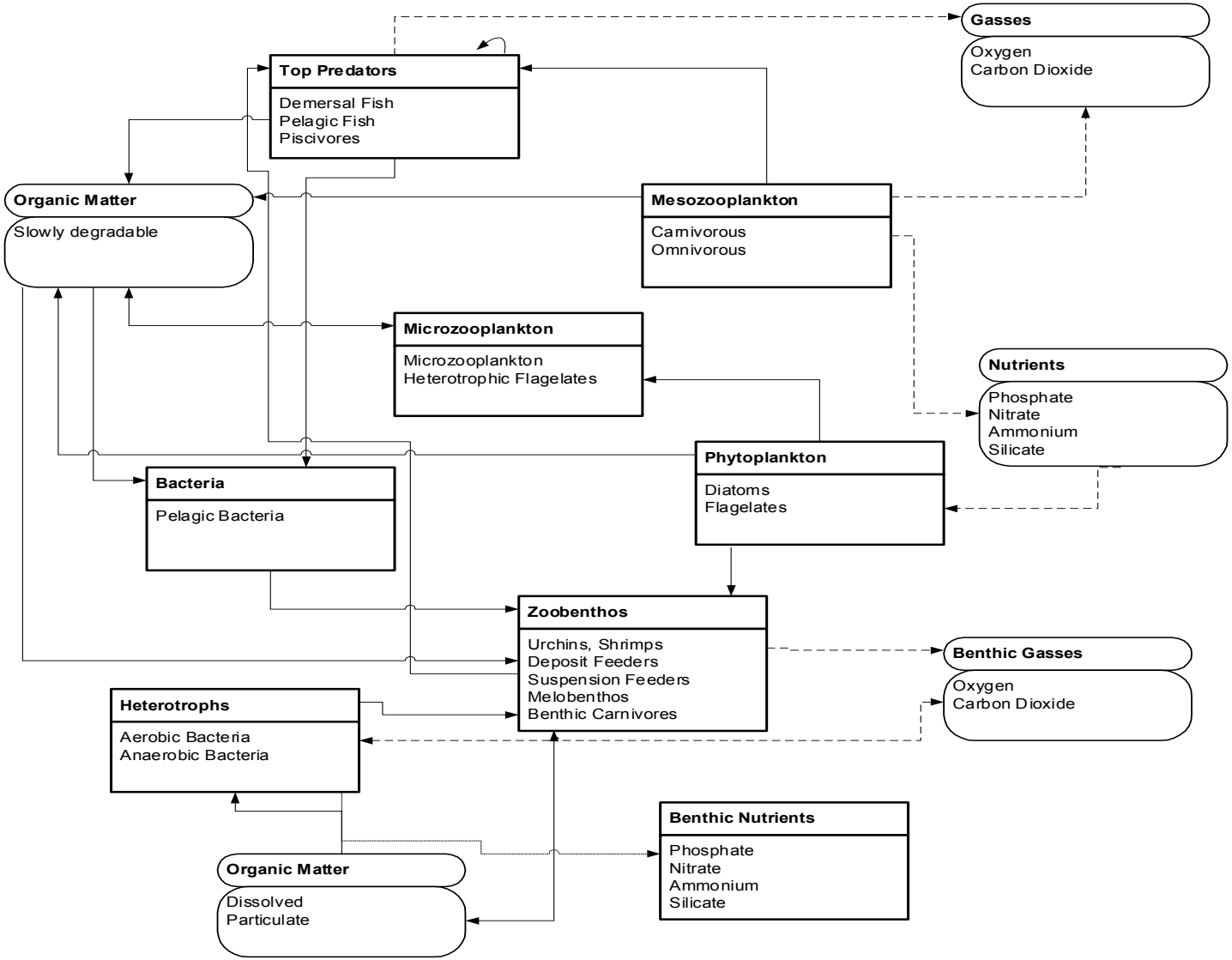
- Scientific Discovery represents the pinnacle of human creative thought
- Is it likely candidate for automation by computer?
- Certainly yes!!!

How far can we push...

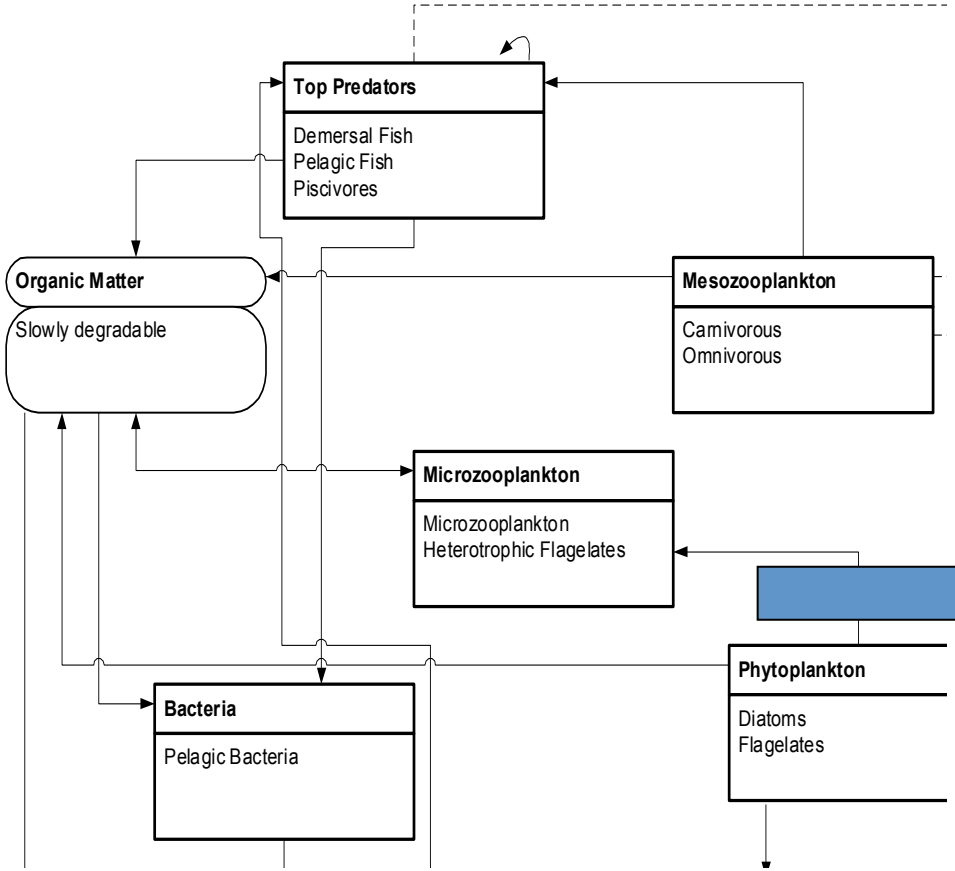
- Evolution of Systems of Coupled Ordinary Differential Equations

$$\frac{dx_1}{dt} = f_1(t, x_1, x_2, \ldots, x_n)$$
$$\frac{dx_2}{dt} = f_1(t, x_1, x_2, \ldots, x_n)$$
$$\vdots$$
$$\frac{dx_n}{dt} = f_n(t, x_1, x_2, \ldots, x_n)$$

Process Flow Diagram



Process Library



F:\pletdata\BOD_DECAY.for

```
      SUBROUTINE            ( PMSA   , FL     , IPOINT , INCREM, NOSEG ,
     +                        NOFLUX , IEXPNT , IKNMRK , NOQ1  , NOQ2  ,
     +                        NOQ3   , NOQ4   )
C
C*******************************************************************************
C
      IMPLICIT NONE
C
C     Type    Name         I/O Description
C
      REAL(4) PMSA(*)     !I/O Process Manager System Array, window of routine to process library
      REAL(4) FL(*)       ! O  Array of fluxes made by this process in mass/volume/time
      INTEGER IPOINT(  2) ! I  Array of pointers in PMSA to get and store the data
      INTEGER INCREM(  2) ! I  Increments in IPOINT for segment loop, 0=constant, 1=spatially varying
      INTEGER NOSEG       ! I  Number of computational elements in the whole model schematisation
      INTEGER NOFLUX      ! I  Number of fluxes, increment in the FL array
      INTEGER IEXPNT      ! I  From, To, From-1 and To+1 segment numbers of the exchange surfaces
      INTEGER IKNMRK      ! I  Active-Inactive, Surface-water-bottom, see manual for use
      INTEGER NOQ1        ! I  Nr of exchanges in 1st direction, only horizontal dir if irregular mesh
      INTEGER NOQ2        ! I  Nr of exchanges in 2nd direction, NOQ1+NOQ2 gives hor. dir. reg. grid
      INTEGER NOQ3        ! I  Nr of exchanges in 3rd direction, vertical direction, pos. downward
      INTEGER NOQ4        ! I  Nr of exchanges in the bottom (bottom layers, specialist use only)
      INTEGER IPNT(  2)   !    Local work array for the pointering
      INTEGER ISEG        !    Local loop counter for computational element loop
C
C*******************************************************************************
C
C     Type    Name         I/O Description                                       Unit
C
      REAL(4) BOD_D_Rate  ! I  Decay rate for Biochemical Oxygen Demand          1 / day
      REAL(4) BOD         ! I  Biochemical Oxygen Demand                        g O2 / m3
      REAL(4) BOD_D_flux  ! F  Decay flux of Biochemical Oxygen Demand          g O2 / m3 / day
      INTEGER IBOD_D_flux !    Pointer to the Decay flux of Biochemical Oxygen Demand
C
C*******************************************************************************
C
      IPNT        = IPOINT
      IBOD_D_flux = 1
C
      DO 9000 ISEG = 1 , NOSEG
C
          BOD_D_Rate = PMSA( IPNT(  1) )
          BOD        = PMSA( IPNT(  2) )
C
C     *****     Insert your code here  *****
C
          BOD_D_flux = ??????
C
C     *****     End of your code       *****
C
          FL   ( IBOD_D_flux ) = BOD_D_flux
C
          IBOD_D_flux = IBOD_D_flux + NOFLUX
          IPNT        = IPNT        + INCREM
C
 9000 CONTINUE
C
      RETURN
      END
* * * End of File * * *
```

The Role of Scientist

- Problem formulation
- Effective representation (representational engineering)
- Data manipulation
- Algorithm manipulation
- Post-processing

Innovation Engine

- Logic is not the key to invention! What human's bring to invention process is not their logic, but their illogic!
- Logic should be used as a constraining factor (*i.e.* dimensional correctness, mass and momentum conservation)
- For the first time in history, we can tell to computer what to do, and not how to do it!

Conclusions

- Knowledge Discovery is coming to a theater near you
- The best results are generated ONLY when a tight interaction between domain and method knowledge is secured
- This is MAINLY Hypothesis generation!