# Understanding the most Massive Stars in the Universe using Genetic Algorithms

**Paco Najarro**

Instituto de Estructura de la Materia

Rohied Mokiem, Amsterdam

Alex de Koter, Amsterdam

Annique Lenorzer, Tenerife

Elena Puga, Madrid

Artemio Herrero, Tenerife

Jo Puls, Munich

CSIC

CESGA, 12-Jun-2008

DAMIR

- *T*
- *L*
- *M*
- *a*

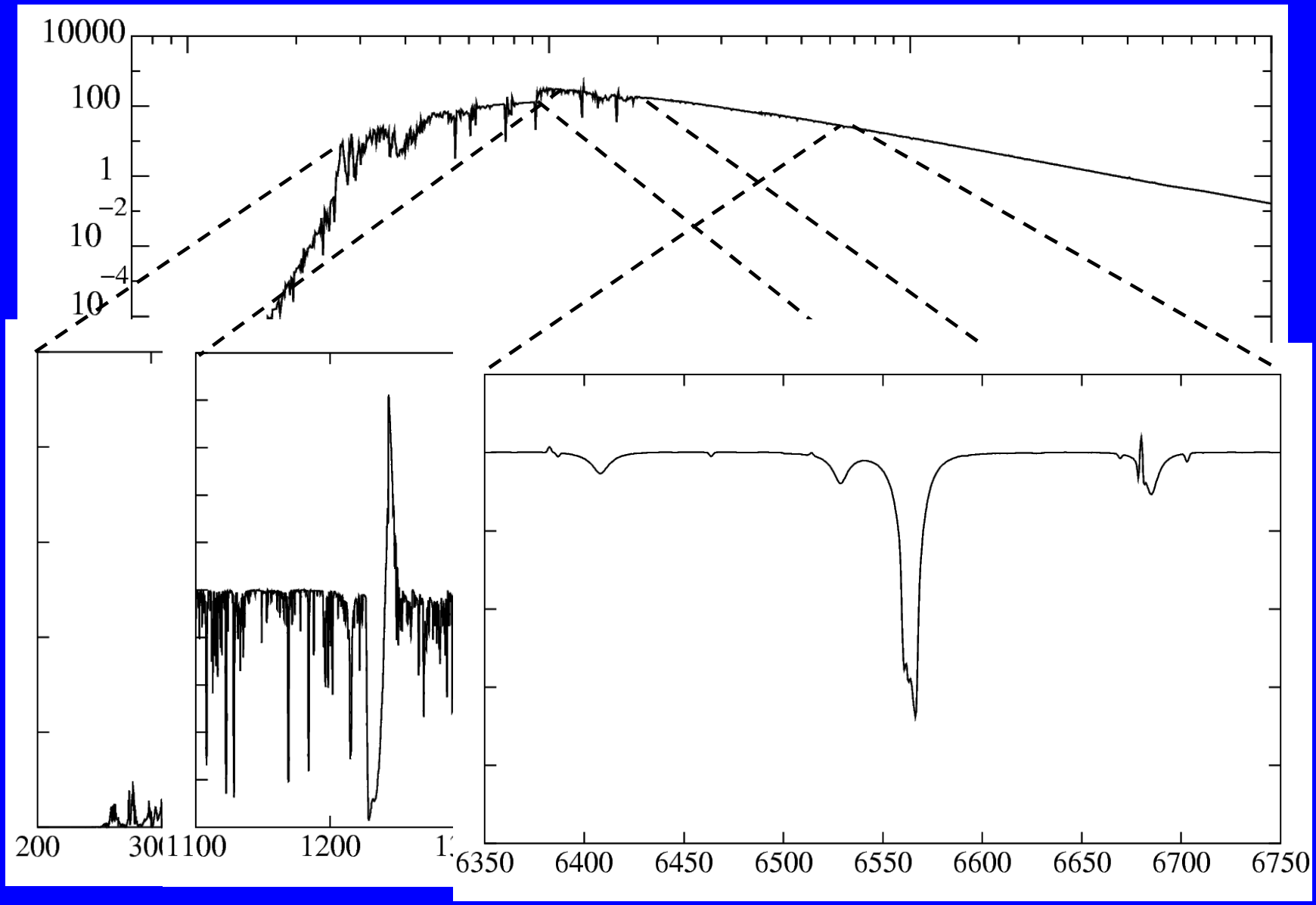
$M_\odot$/yr

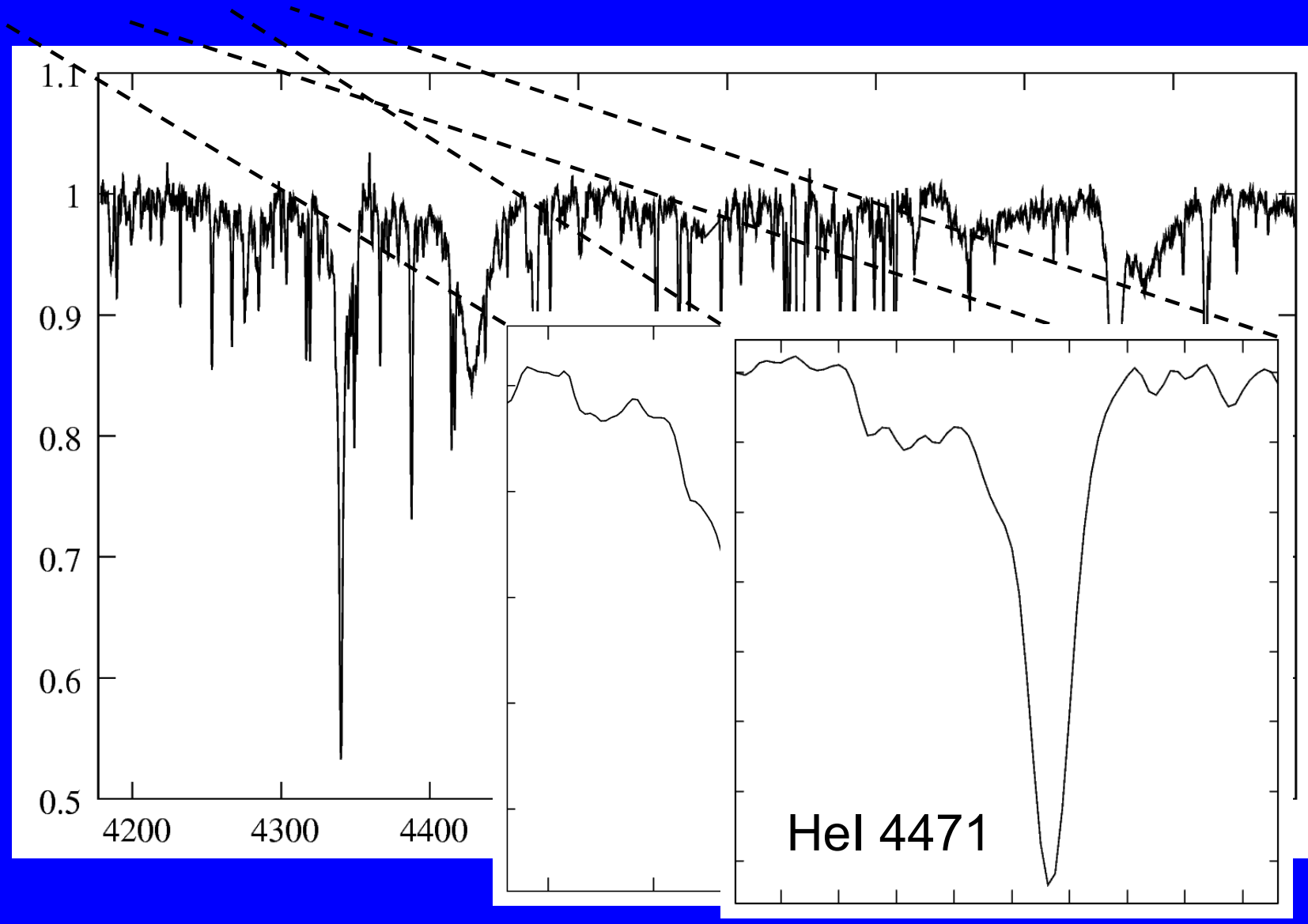
# Importance of massive stars

- Evolution of Galaxies
- Super Novae
- Gamma-ray bursts
- Neutron Stars & Black Holes
- Re-ionization of early universe

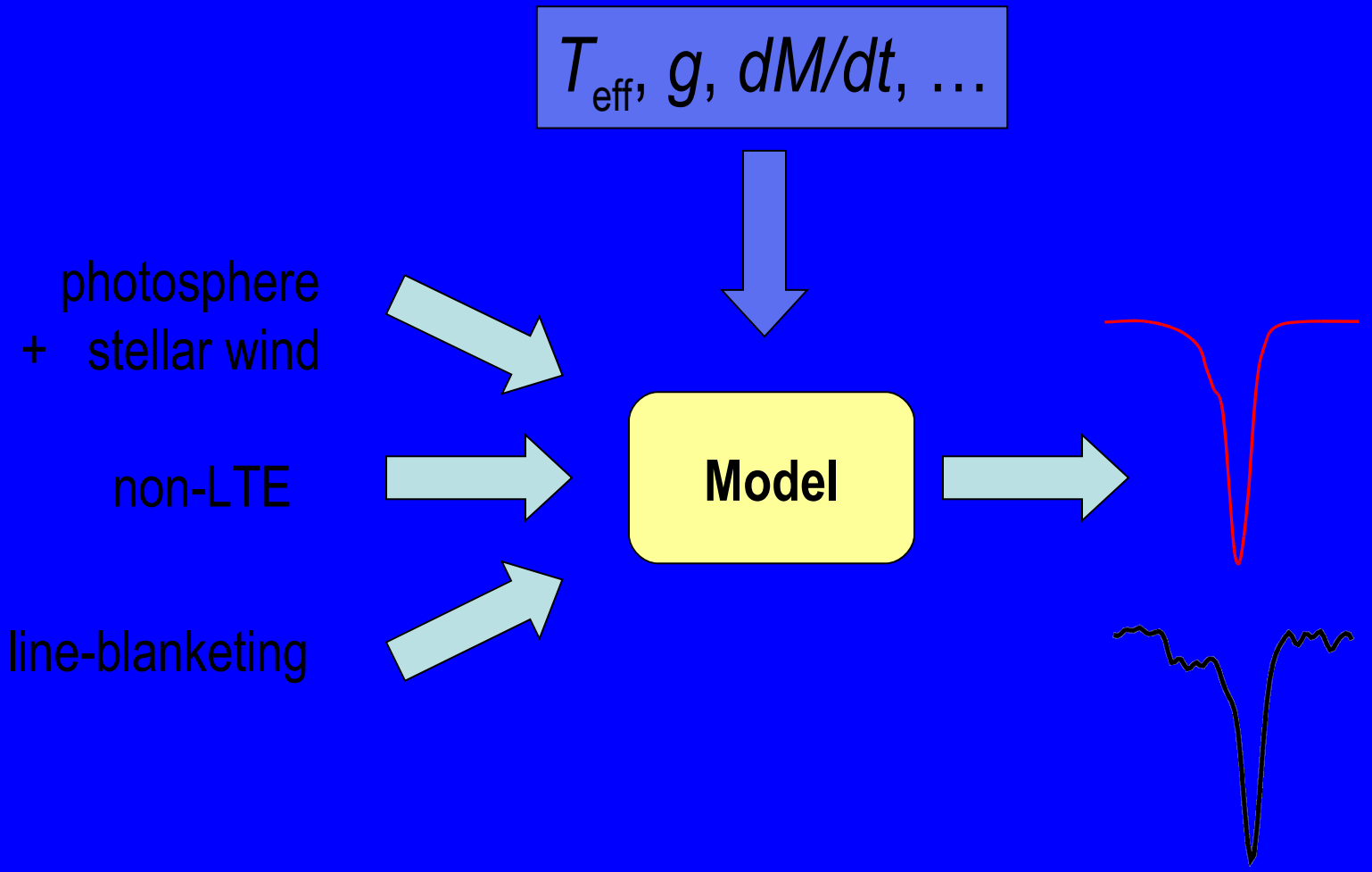
# Large sample analysis

- ~100 analyzed until now
- Flames Survey
  - ~100 hours VLT time
  - Galactic, SMC and LMC fields
  - will double total number of analyzed stars
  - calibration of fundamental parameters
- Quantitative spectroscopy


10000
100
1
10 −2
10 −4
200
300
1100
1200
6350
6400
6450
6500
6550
6600
6650
6700
6750

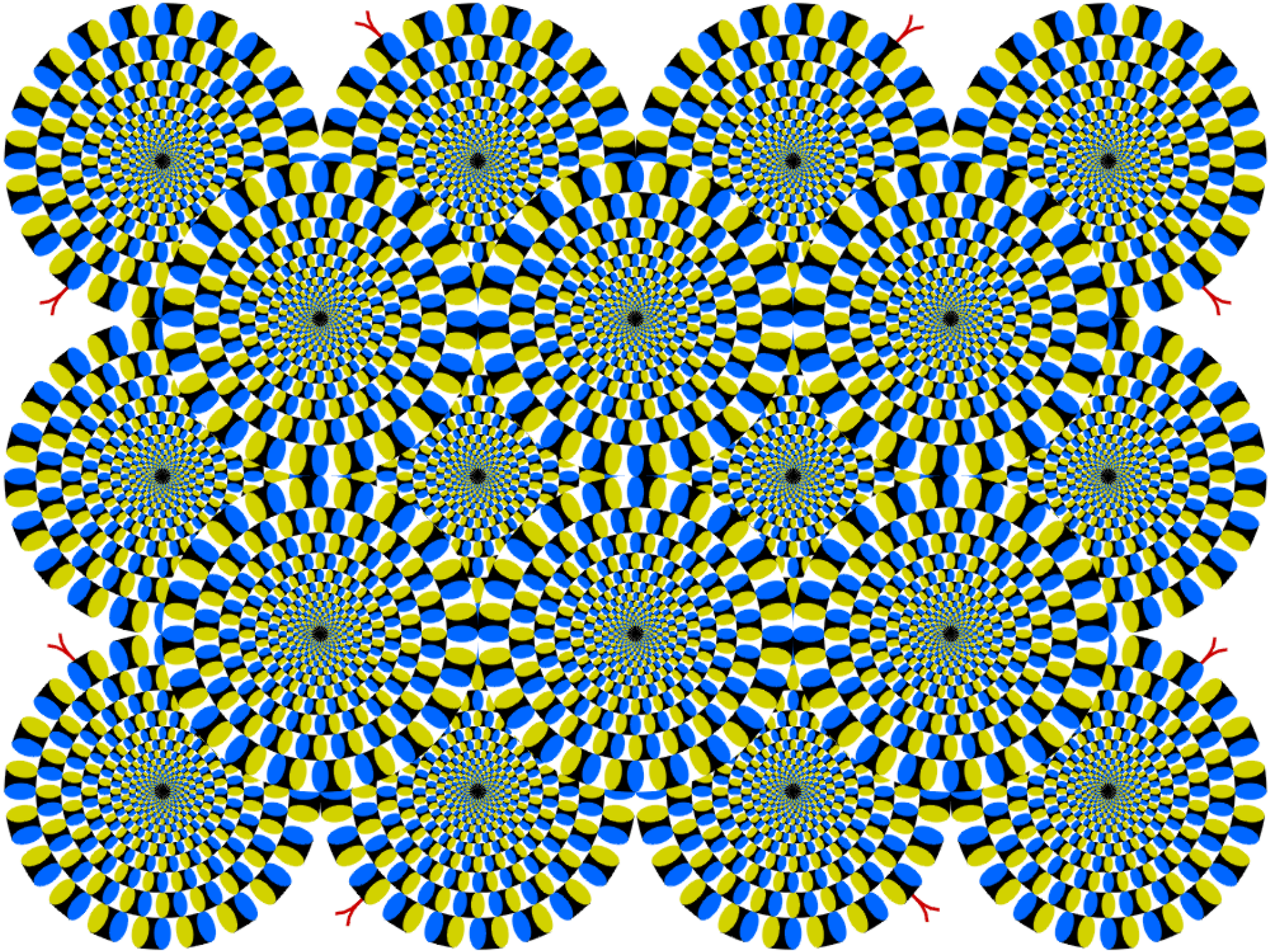

1.1
1
0.9
0.8
0.7
0.6
0.5
4200
4300
4400
HeI 4471

# Quantitative spectroscopy



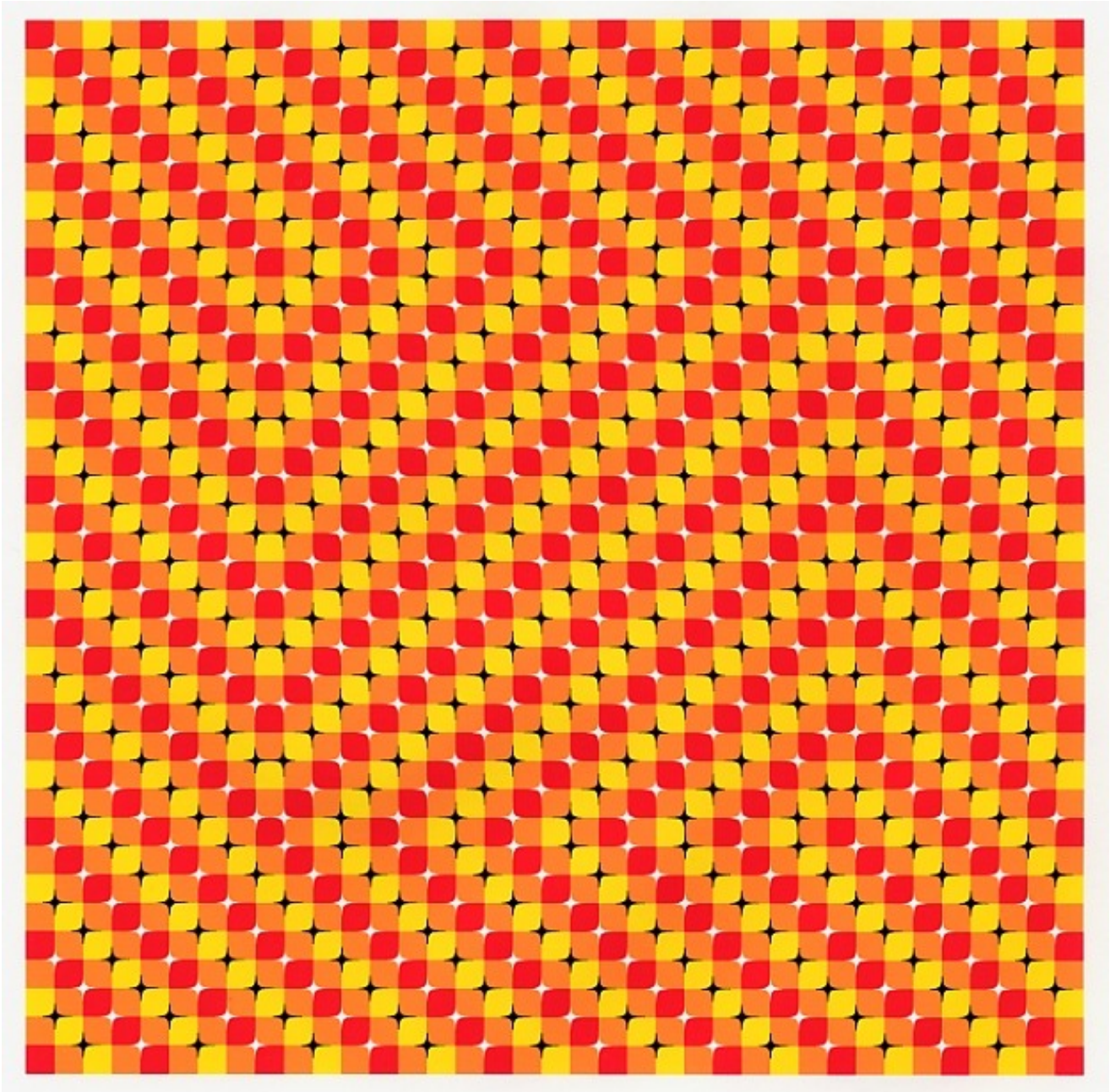
# “by eye”

- The conventional method
- However, …



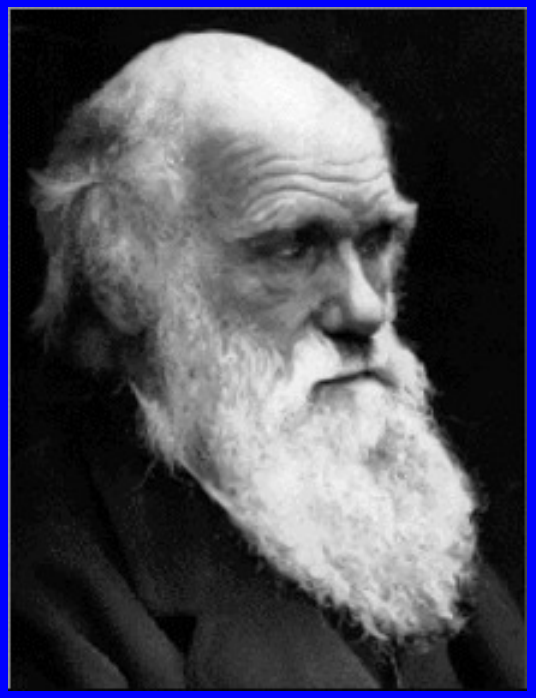
# Drawbacks of fitting “by eye”

- Laborious
- Limited number of physical parameters
- Not reproducible
  - biases
  - best fit?
- No large sample, and …



It gives you a headache

# Natural selection



Charles Darwin
1809 -1882

“I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection.”

# Genetic Algorithms

Adaptative heuristic search method using Darwinian principles

**Applications:**

Scheduling, Data fitting,Trend spotting, Budgeting, ...

Any search for which all solutions cannot be tested.

**Scheme:**

1.Initialization: compute solutions for random parameters

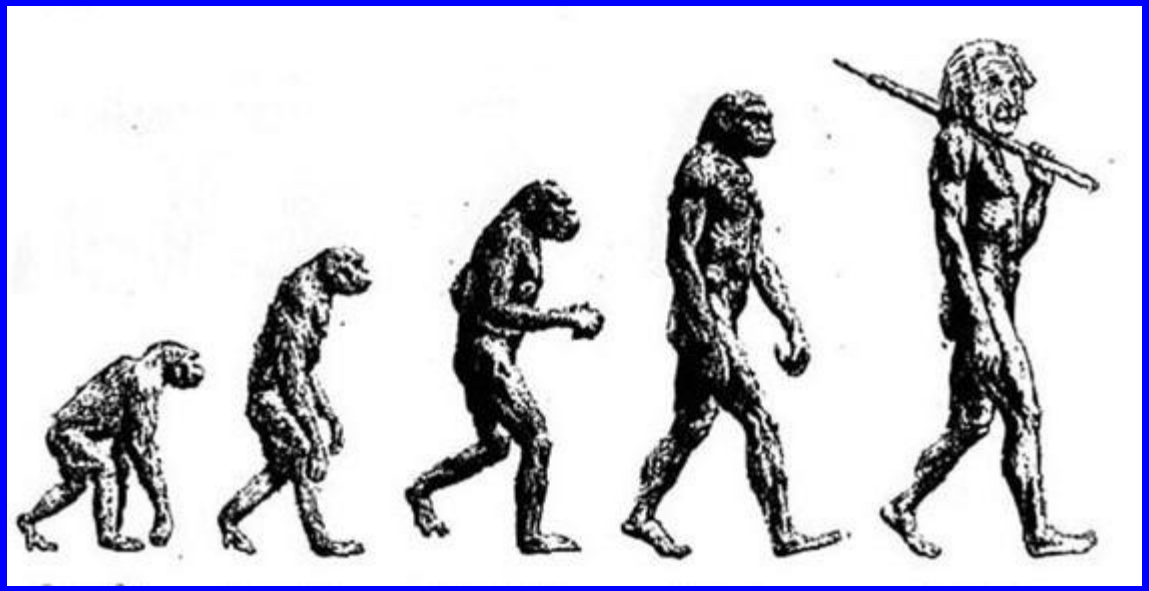
2.Selection: retain successful solutions

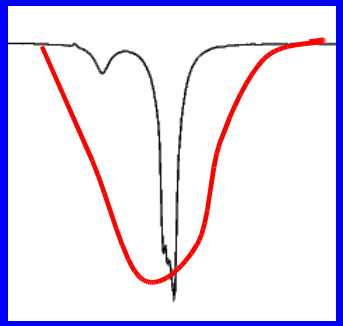
3.Reproduction: deduce potentially better model parameters

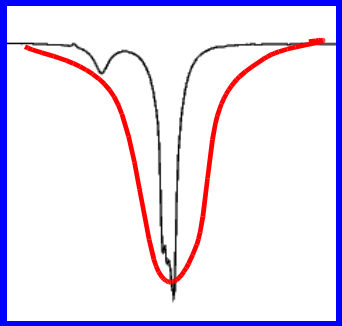
4.Evaluation: compute and evaluate solutions

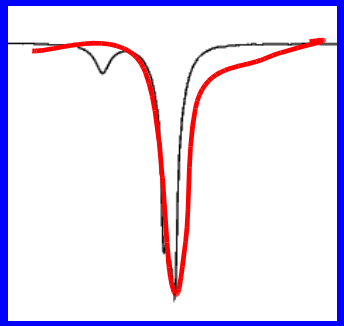
5.Termination: the best solution is satisfactory

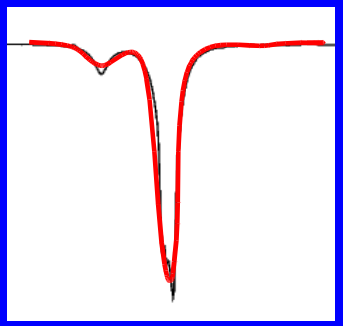
# Optimization technique







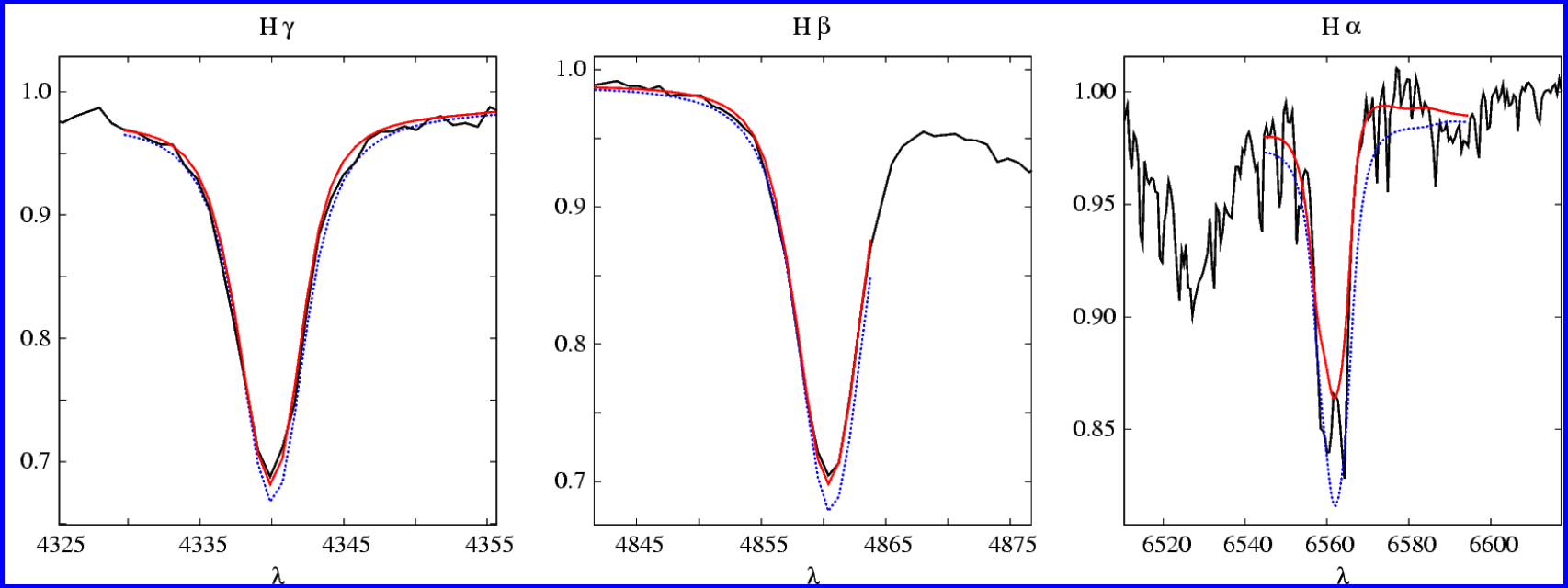




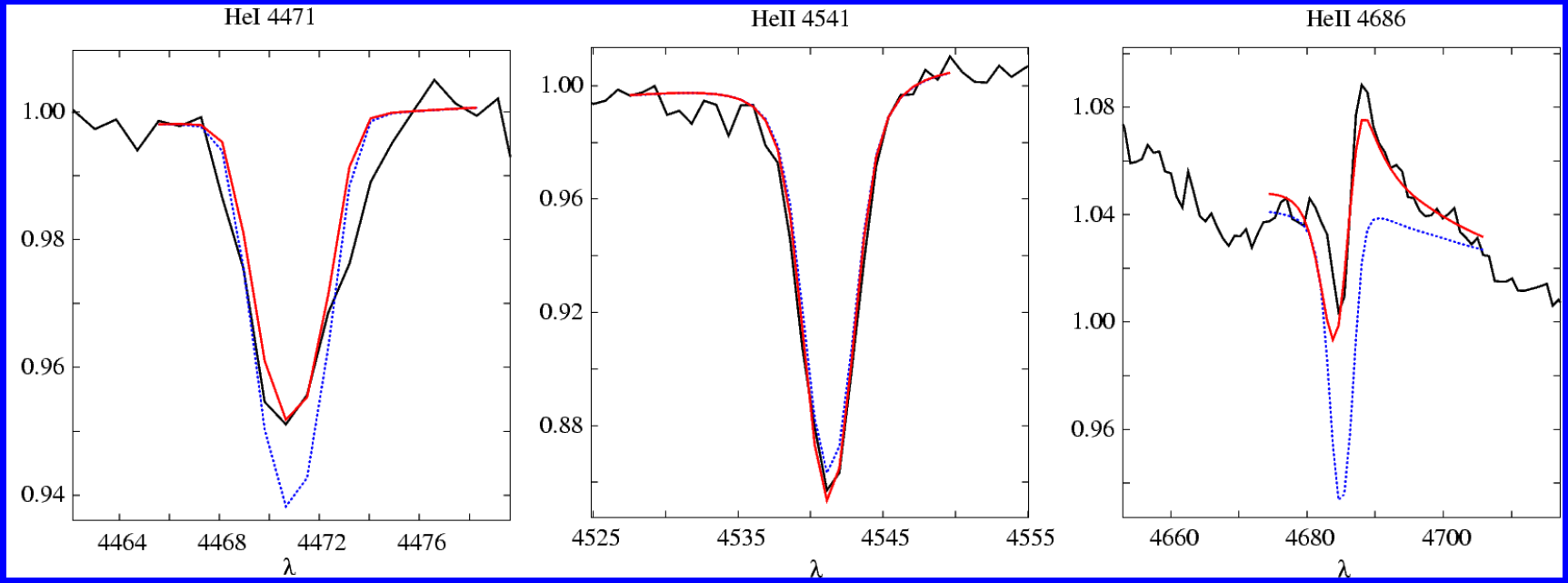
# Basic genetic algorithm

- S

# Cyg OB2 8C O5 If
# Hydrogen lines



# Helium lines



# Encoding: from parameters to gene

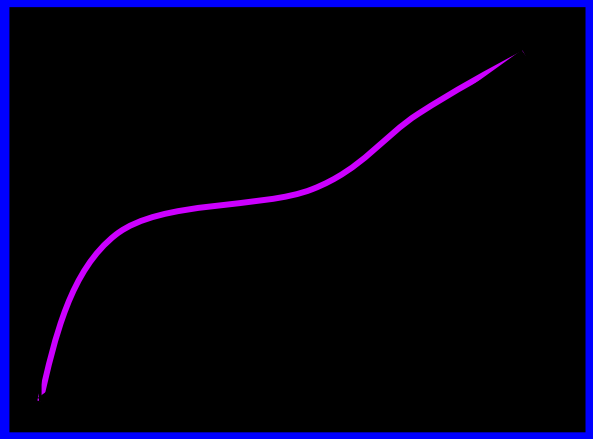
**Bounds of parameters search domain:**

$P_{imin} < P_i < P_{imax}$

**Coding Function:**

$F(P_i)=ph_i$

F(31000K)=0.274396



**Precision required:**

$ph_i$=0.27 (2 digits)

**Choice of gene base:**

2 (computers), 4 (life), 10 (physics)...

$g_i$=27

**PARAMETERS TO GENE**

**(31000K,3.65,10E-8,0.8) ≈ 27421809**

# Reproduction

**Cross parents:**

1, 2 or more cross points

Include complementary kid (base 2)

P1(7395725439)

P2(9243872374)

K1(7393872339)

K2(9245725474)

**Mutate kids:**

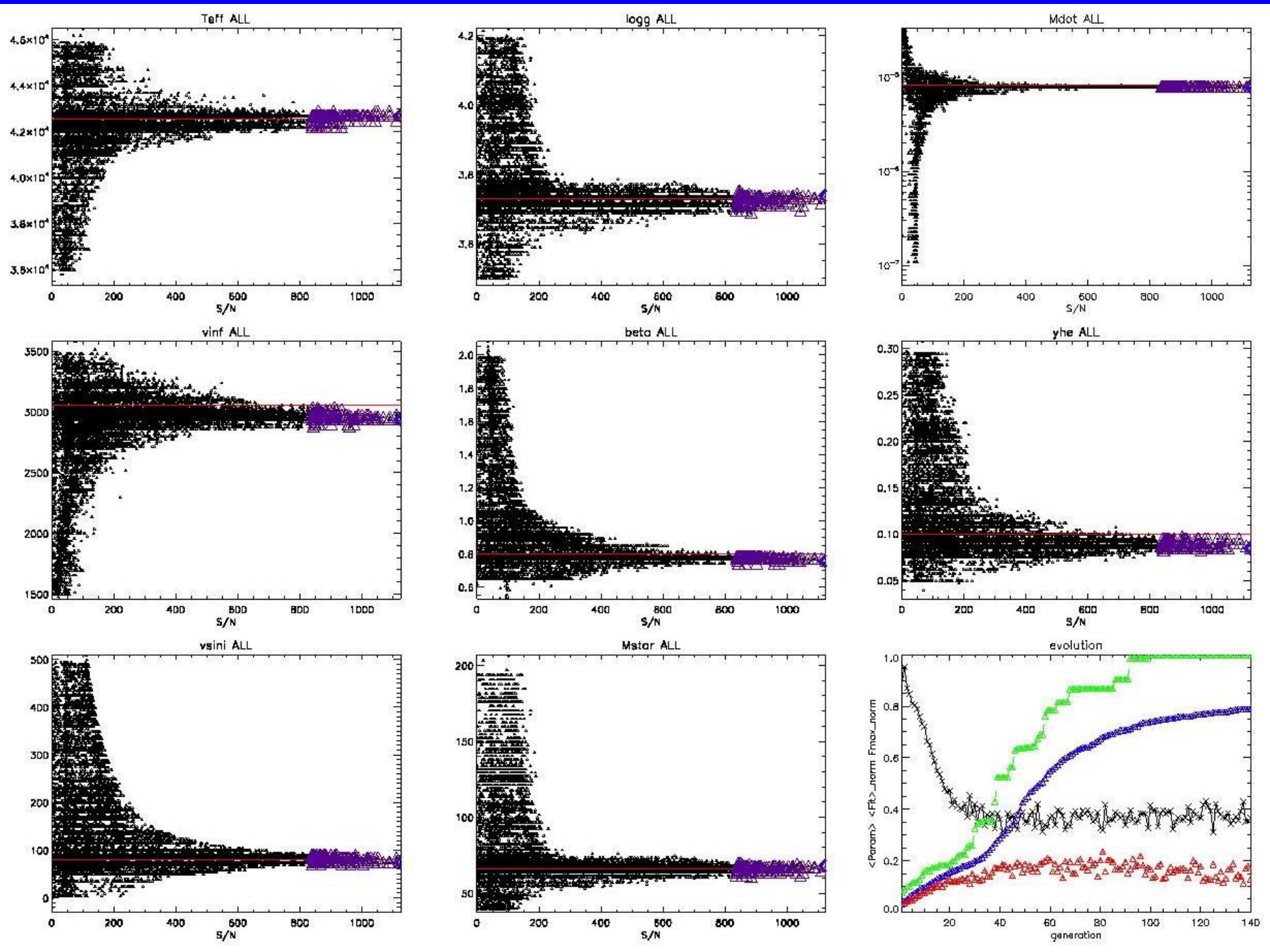
Normal mutation, creep mutation or both

Constant or varying mutation rate

Same or different for each gene locus

K1(7393972340)

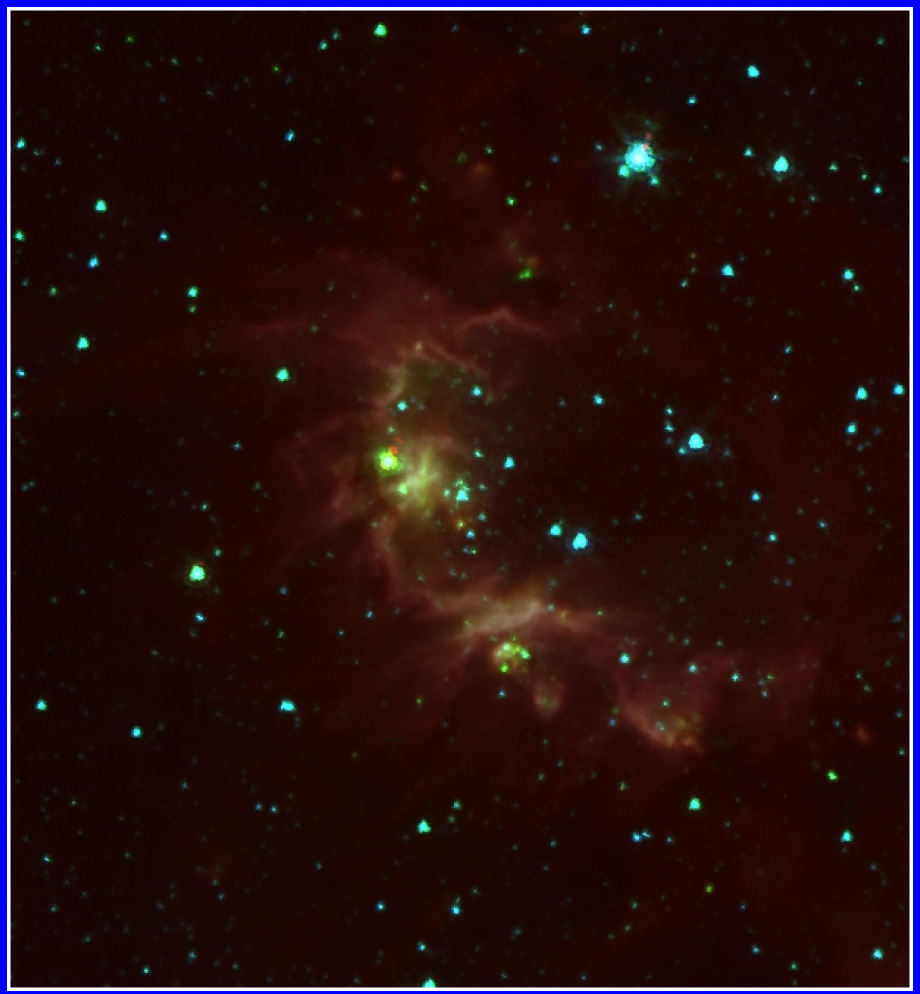
K2(9145725874)

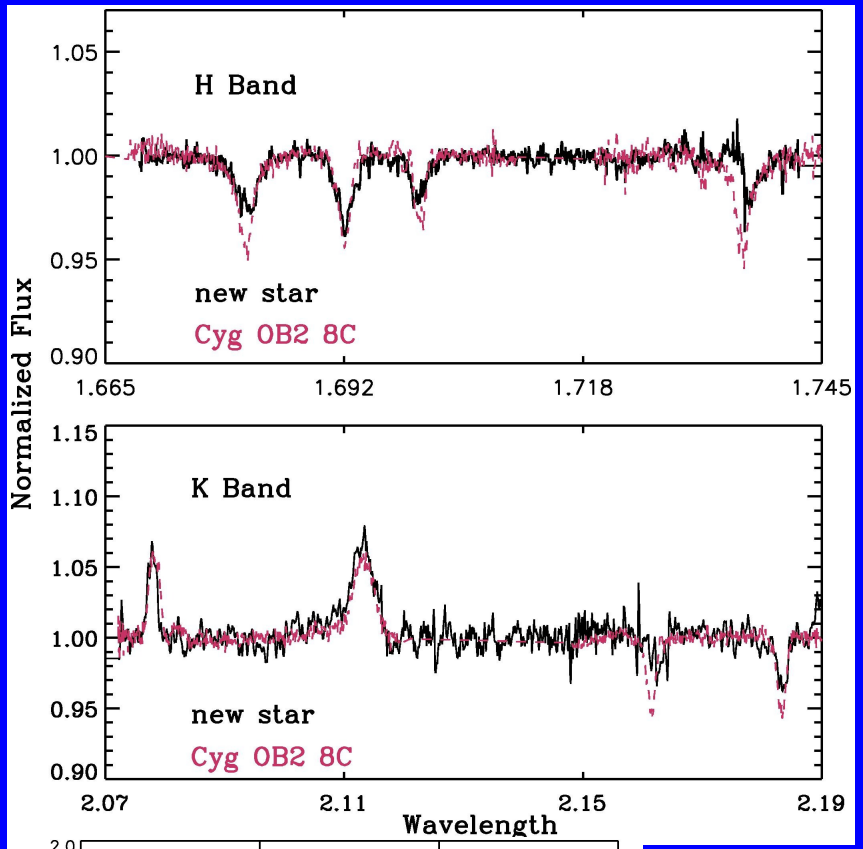

Teff ALL
logg ALL
Mdot ALL
vinf ALL
beta ALL
yhe ALL
vsini ALL
Mstar ALL
evolution
S/N
generation
<Param> <Fit>_norm Fmax_norm

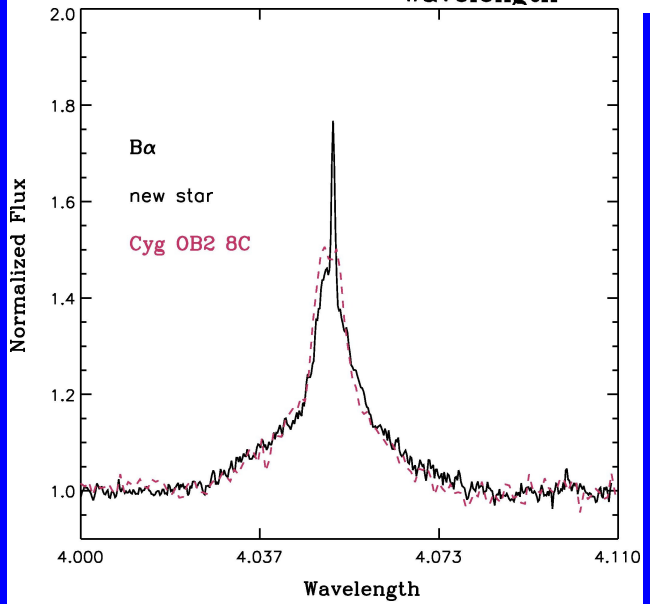
# CESGA PROJECT:

## New Obscured clusters in the Milky Way

Lenorzer et al 2008





# OUR CASE

About 10 parameters coded on 2 digits each: gene length ≈ 20
Model running in 30 min, 50 processors: population size ≈ 100

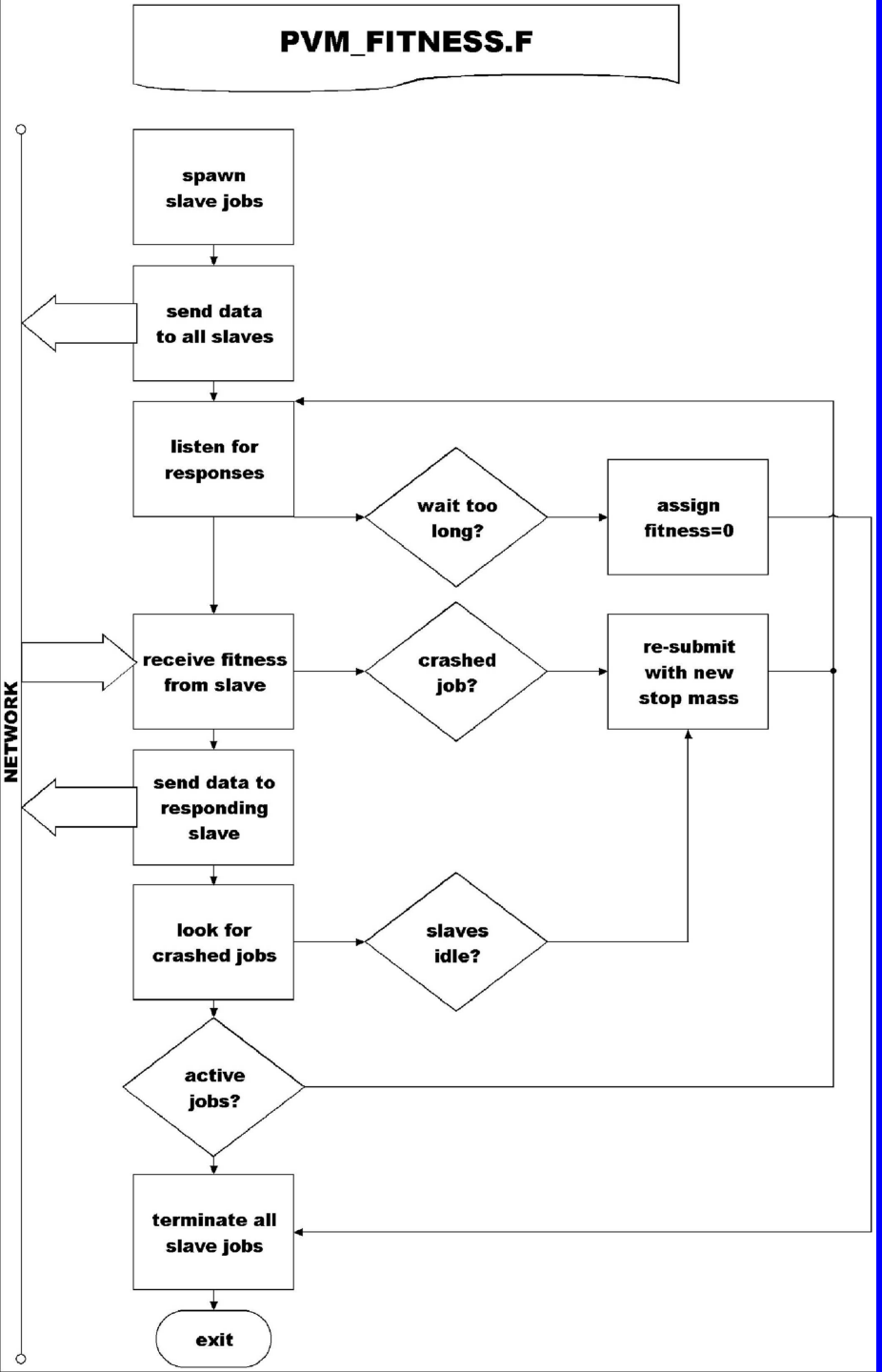
Random search would take <1.E8 generations> for 1 digit accuracy
Need a high selection pressure to be “fast”

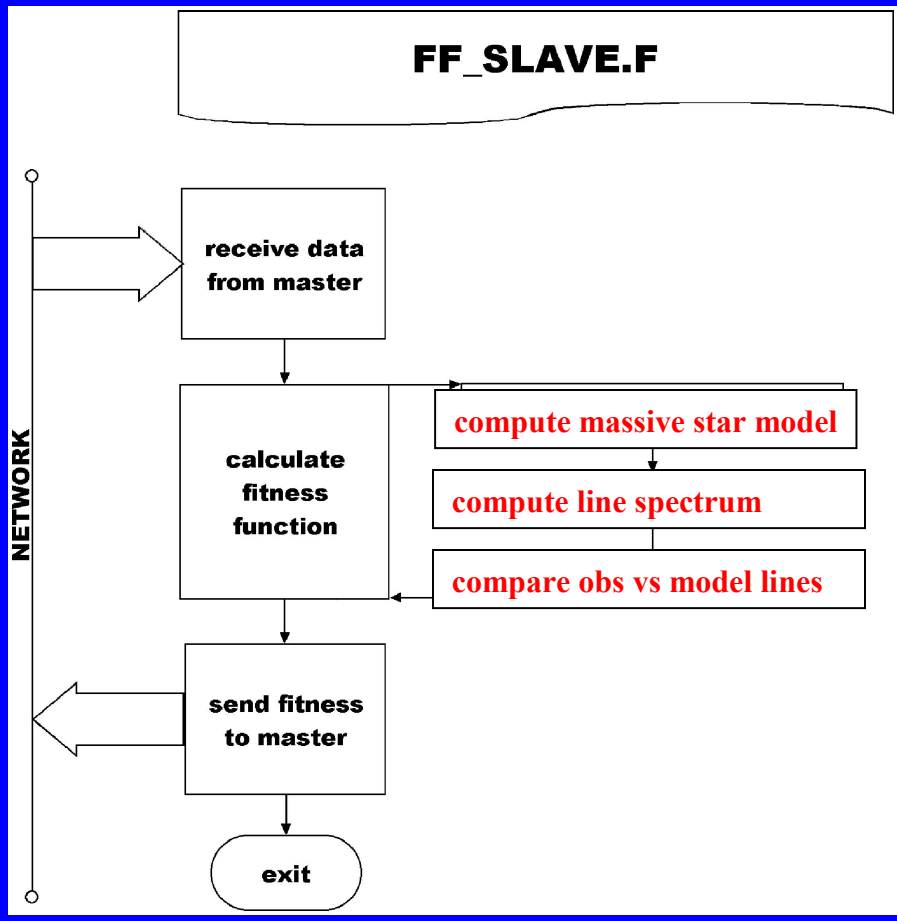
Initial population: each gene value present <pop/base=10 times>
Without mutation all genes are the same after ≈ 50 generations

Mutation needs to be high enough to keep gene diversity
Mutation rate needs to be low enough to allow gene convergence
Uniform mutation ⇨ genetic based adaptative mutation

Mutation rate needs to be low enough to maintain evolution
Full generation replacement ⇨ Steady state replace worst

$$Fgn = Fgn\text{-}1 \times Psel \times (1-Pmut) + Pmut \times pop/base$$


PVM_FITNESS.F
spawn slave jobs
send data to all slaves
listen for responses
wait too long?
assign fitness=0
receive fitness from slave
crashed job?
re-submit with new stop mass
send data to responding slave
look for crashed jobs
slaves idle?
active jobs?
terminate all slave jobs
exit
NETWORK


FF_SLAVE.F
receive data from master
calculate fitness function
compute massive star model
compute line spectrum
compare obs vs model lines
send fitness to master
exit
NETWORK

# Scalability on parallel computers