Complexity of gap-filling CH_4 flux data

Dr Sigrid Dengel

University of Helsinki

Outline

- Ecosystems known for CH_4 emissions;
- Gap-filling methods applied so far;
- Why is gap-filling of CH_4 data not as easy and straight forward as CO_2;
- What can we take over from CO_2?
- Neural networks – a suggested method;

Ecosystems with CH_4 emissions

- Main ecosystems relevant to CH_4 flux studies:

  - Fens;
  - Mires;
  - Bogs;
  - Sedge
  - Water bodies
  - Tundra;
  - Forests;
  - Peat swamp forests;
  - River plains/delta;
  - ???

Gap-filling methods applied so far

- Low resolution daily mean: Hargreaves *et al.* (2001); Rinne *et al.* (2007); Riutta *et al.* (2007); Wille *et al.* (2008); Sachs *et al.* (2008) Long *et al.* (2010); Jackowicz-Korczyński *et al.* (2010); Tagesson *et al.* (2012) – peat temp at various depths;
- Wille et al. (2008) and Sachs et al. (2008) – soil temp and friction velocity;

- Employed models: Wille et al. (2008); Parmentier *et al.* (2011) and Forbrich *et al.* (2012) recover missing data in their daily, 3 hourly and 30 min mean data, respectively.

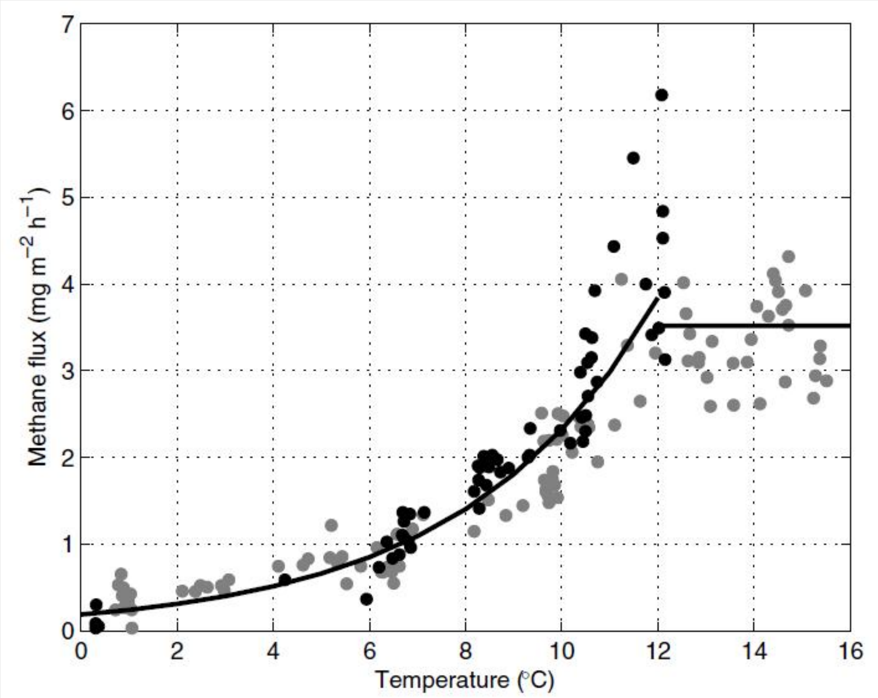
Why is gap-filling of CH_4 data not as straight forward as CO_2?

- Majority gap-filled only daily values;
- Data were split by season, even than still limited;
- Relationships with a variety of drivers;
- Always labour intensive;
- Diurnal vs. non diurnal variation;
- Seasonal complexity;
- Ecosystem location (Nordic, temperate, tropical);
- Grazing patterns (ruminants);

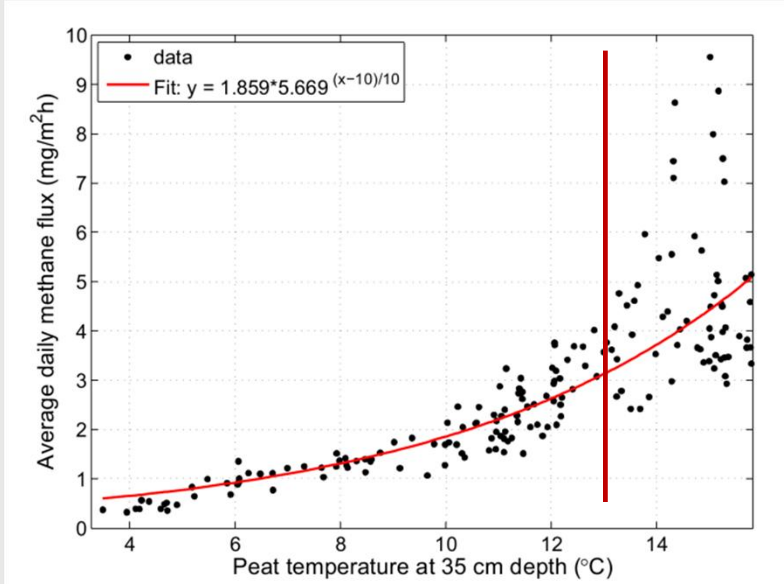
Variables influencing(?) CH_4 fluxes(?)

- Soil temperature;
- Water table level;
- Soil moisture;
- Air pressure
- Air temperature
- Wind speed/velocity
- Ebullition (bubbling effect)
- Soil type
- Vegetation type
- Solar radiation;

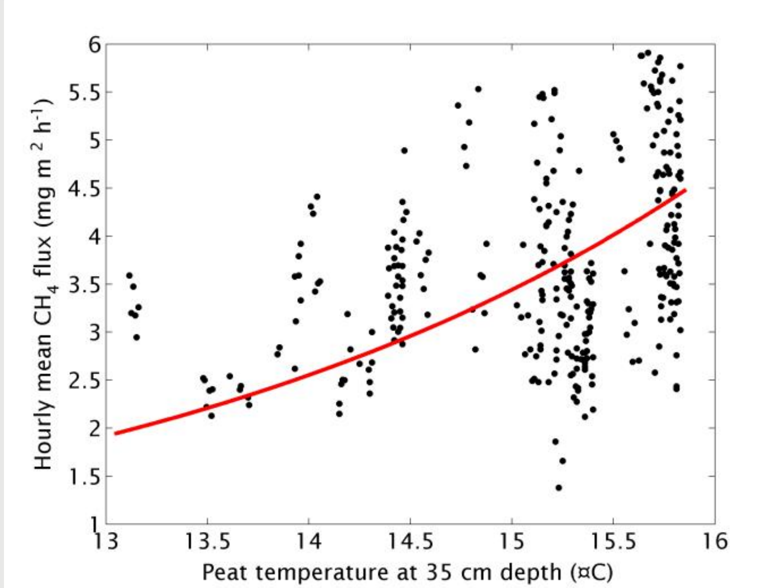
As an example: Rinne et al. (2007)



Rinne et al. (2007). Data representing Siikaneva 2005.

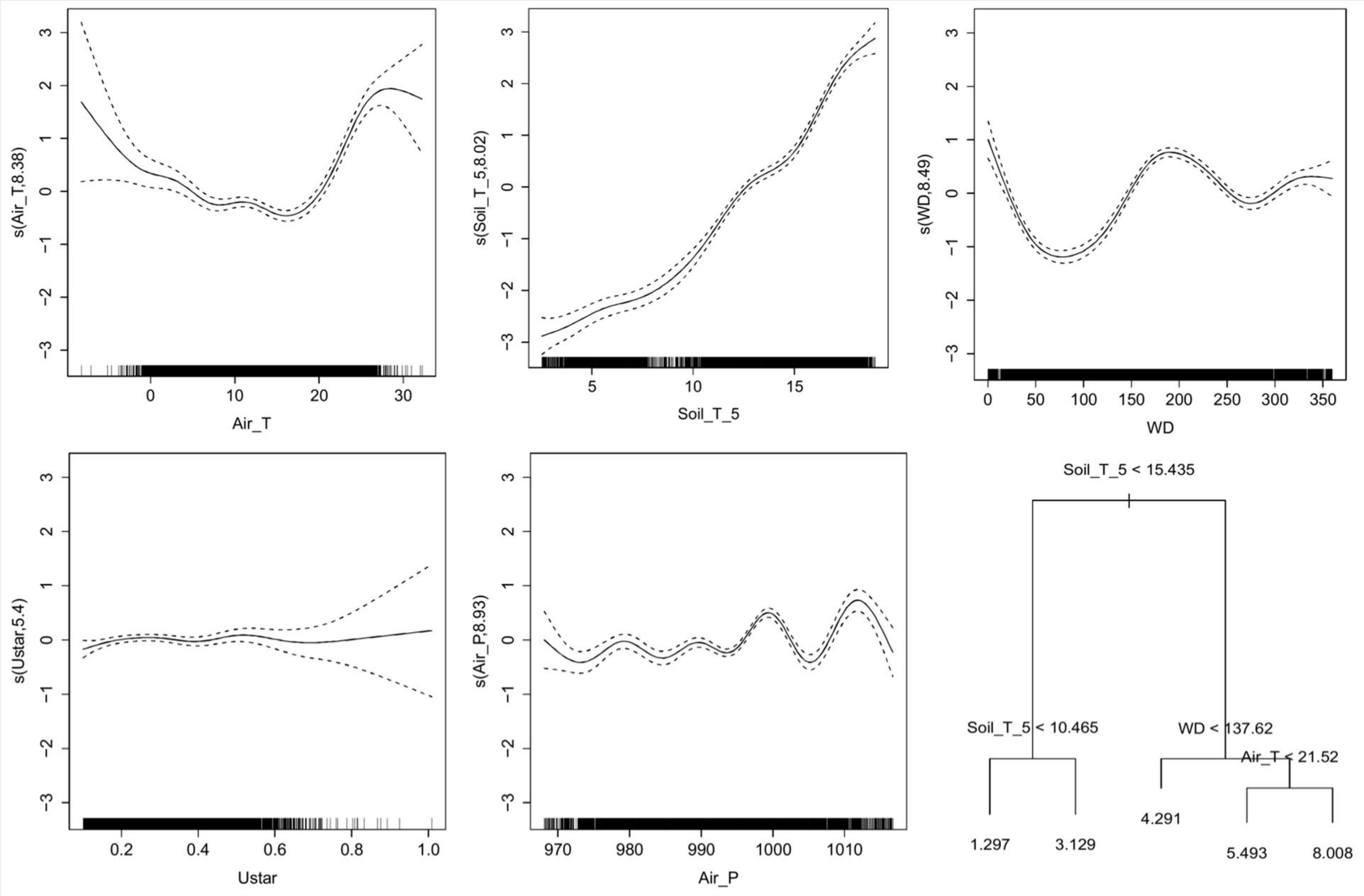


Peltola (2011). Data representing Siikaneva 2010.



Peltola/Haapanala (unpublished data). Data representing Siikaneva 2010.

What now?

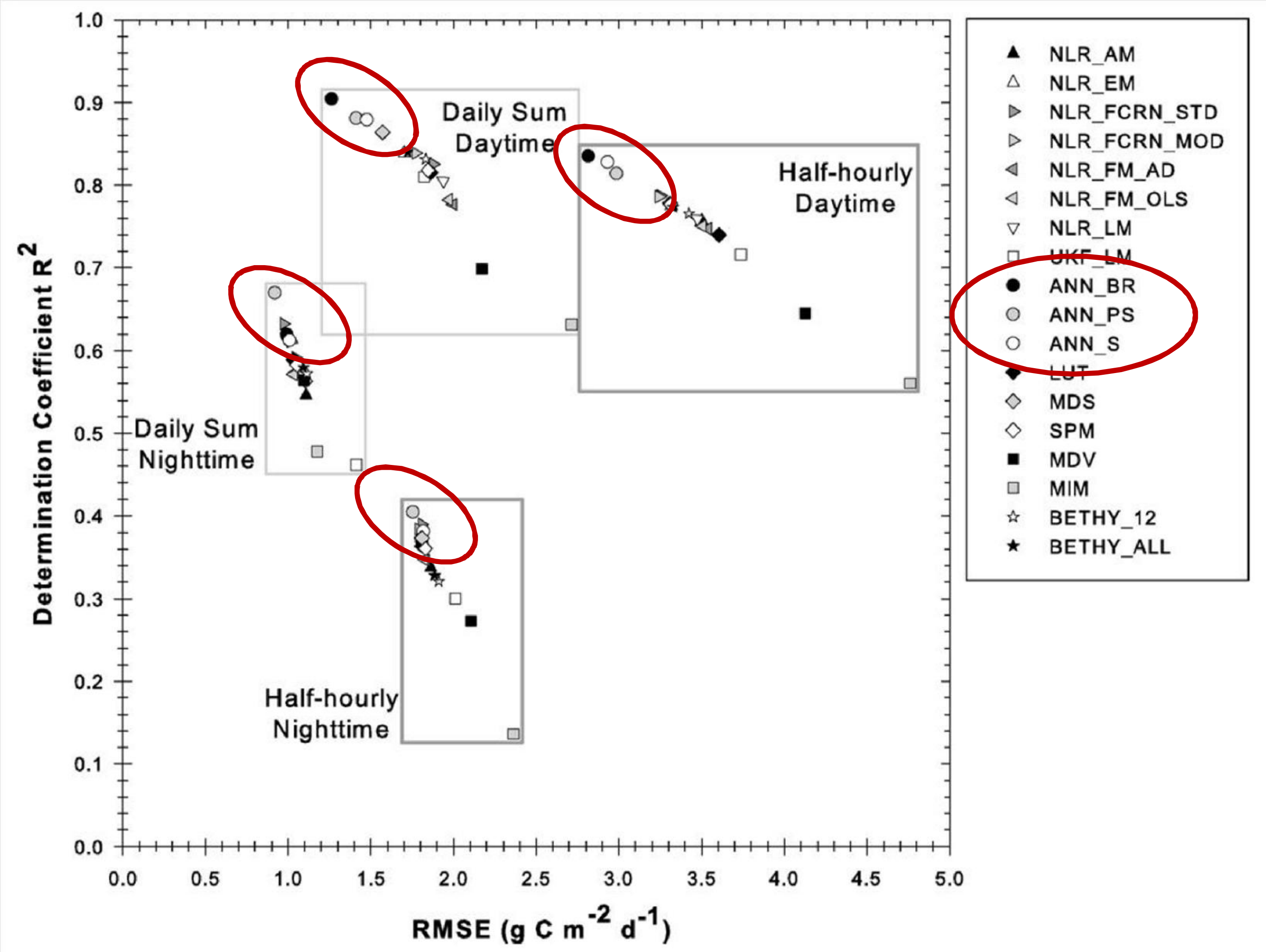


s(Air_T,8.38)
Air_T
s(Soil_T_5,8.02)
Soil_T_5
s(WD,8.49)
WD
s(Ustar,5.4)
Ustar
s(Air_P,8.93)
Air_P
Soil_T_5 < 15.435
Soil_T_5 < 10.465
WD < 137.62
Air_T < 21.52
1.297
3.129
4.291
5.493
8.008

What can we take over from CO_2?

- Falge *et al.* (2001);
- Moffat *et al.* (2007);
- Linear interpolation – short gaps;
- Mean Diurnal Variation (MDV);
- Kalman filtering(?);
- Artificial Neural Networks (ANN’s);

Moffat et al. 2007



Applying neural networks (here as gap-filling)

<u>1st step:</u>

Training

(activation function)

<u>2nd step:</u>

Testing/validating

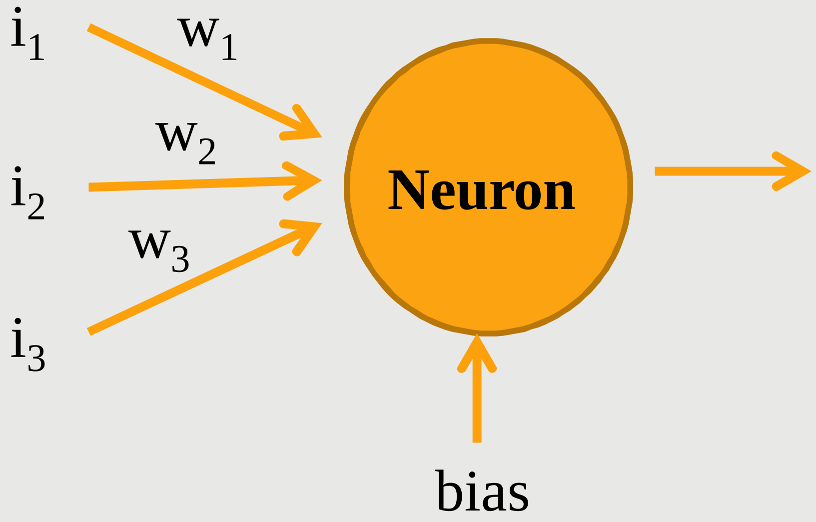
(error function)

<u>3rd step:</u>

Applying

(gap-filling of missing values)

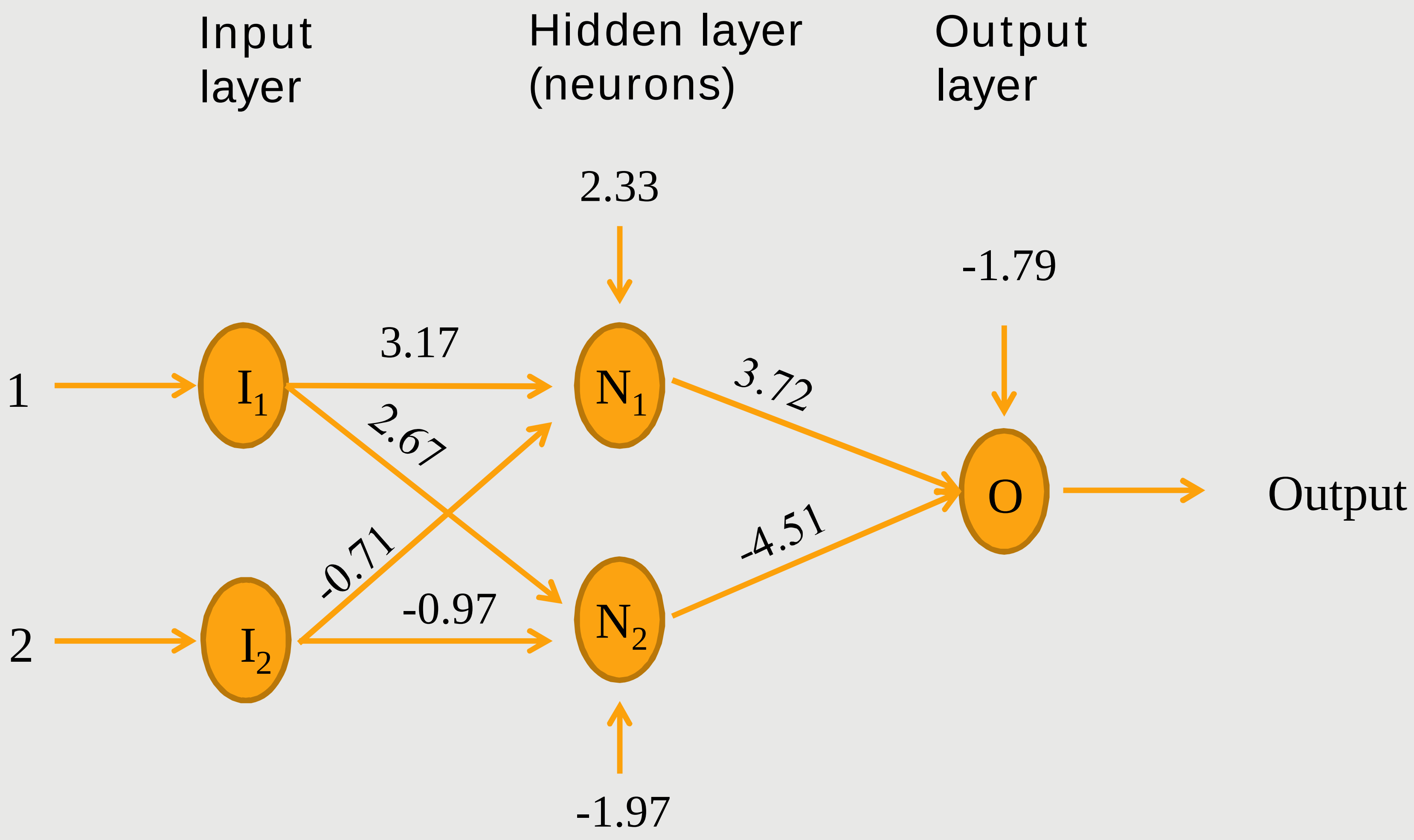
Artificial Neural Networks

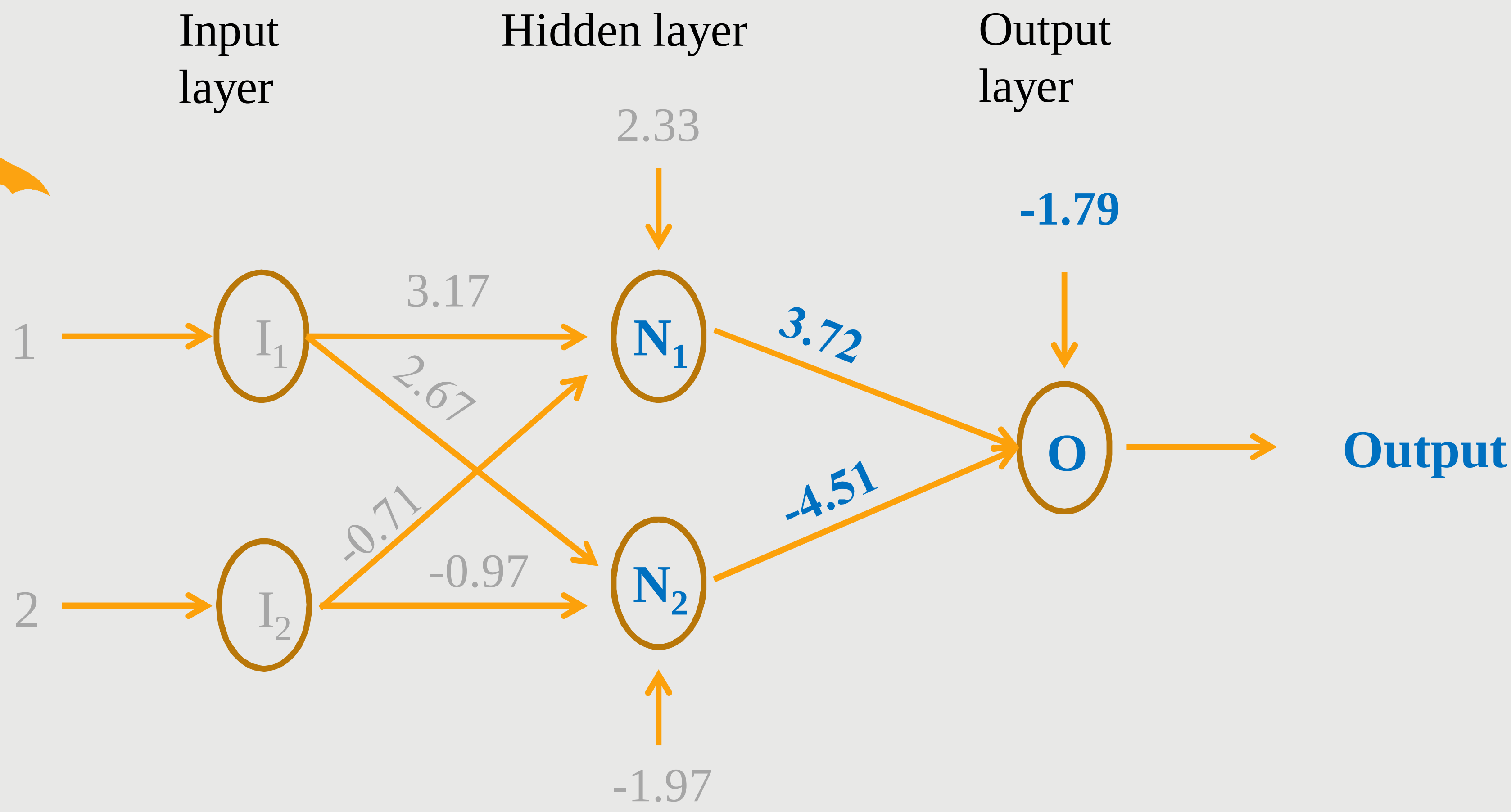


Output = $f(i_1 w_1 + i_2 w_2 + i_3 w_3 + \text{bias})$

Sigmoid function is the most commonly used function: $f_{(x)} = 1/(1+e^{-x})$

Artificial Neural Networks





$$\mathbf{N1} = 1(3.17) + 2(-0.71) + 2.33 = 4.08 = 1/(1 + e^{-4.08}) = \mathbf{0.983374}$$

$$\mathbf{N2} = 1(2.67) + 2(-0.97) - 1.97 = -1.24 = 1/(1 + e^{1.24}) = \mathbf{0.224436}$$

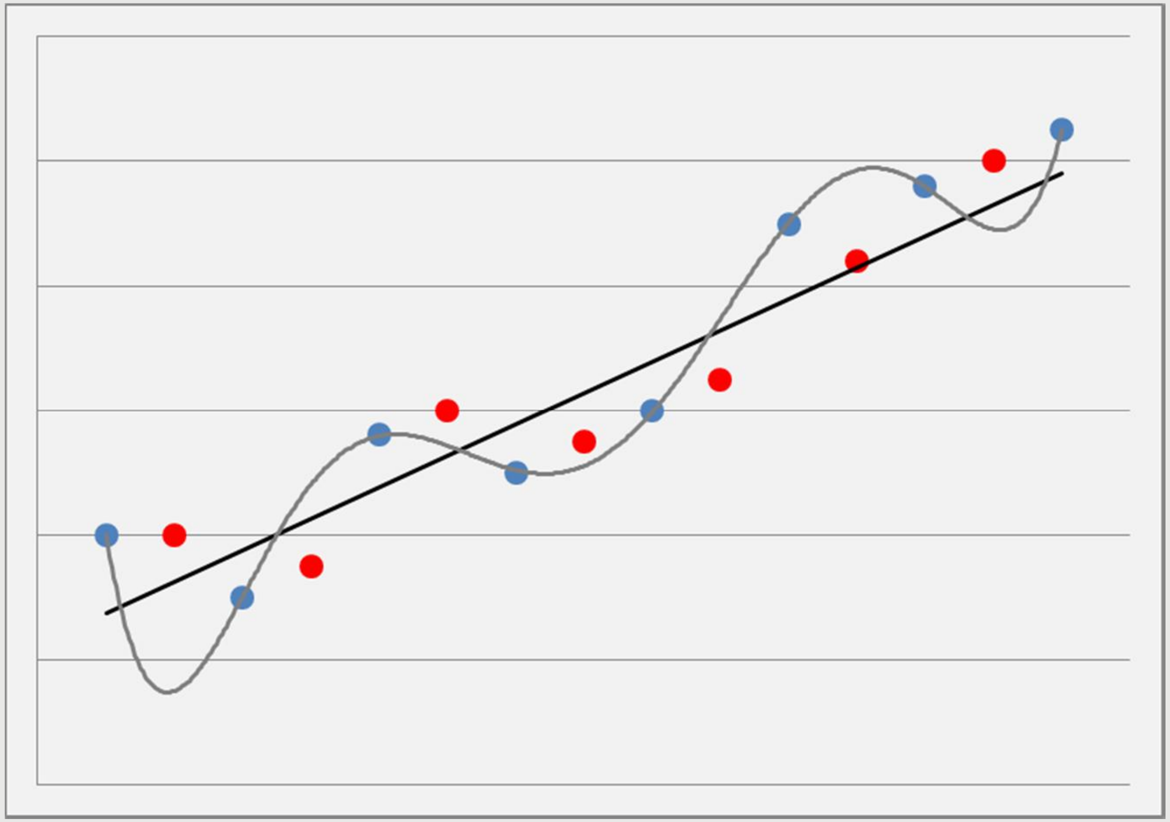
$$\mathbf{O = 0.983374(3.72) + 0.224436(-4.51) - 1.79 = 0.855944}$$

$$\mathbf{Output = 1/(1 + e^{-0.855944}) = 0.701812}$$

How well do our data fit?

- Is the result of **<u>0.70</u>** good enough???
- Common measure of error:
  - Pearson correlation coefficient;
  - (Root) Mean Square Error;
- Pruning/adding neurons and re-adjustment of weights;
- Under vs. over-fitting (quality of the data & network)

How well do our data fit?



- Training data set
- Test data set

— Well fitted
— Overfitted

The number of neuron is very important!

Too few – underfit the data and NN can not learn the details;
Too many – overfit the data and NN learns insignificant details (outliers);