

データ駆動モデルによる生物の空間分布解析

福田信二

東京農工大学大学院農学研究院



Shinji Fukuda

Institute of Agriculture,

Tokyo University of Agriculture and Technology

E-mail: shinji-f@cc.tuat.ac.jp

Web: <http://shinjifukuda-medaka.com>



ニューラルネットワーク in 農業土木学会論文集

著者名	論文名	出版日付	巻	号	ページ
近森,永井	SOLOネットワークを用いた実時間洪水予測	2006	241		25-33
小林,吉武,勝山,岡林	ため池地震危険度評価システムの構築：芸予地震による愛媛県の溜池被害を事例として	2002	70	6	697-703
長谷川,工藤,石井	ニューラルネットワークを用いたコンクリートダム景観の定量的評価	1995	63	5	31-37
平松,四ヶ所,森	多層パーセプトロンモデルによる河川感潮域の塩素量解析	1995	63	4	83-92
猿渡,四方田	水田用水の戦略的管理とニューロネットワークの導入:水管理システム化に関する研究 (V)	1994	172		95-103

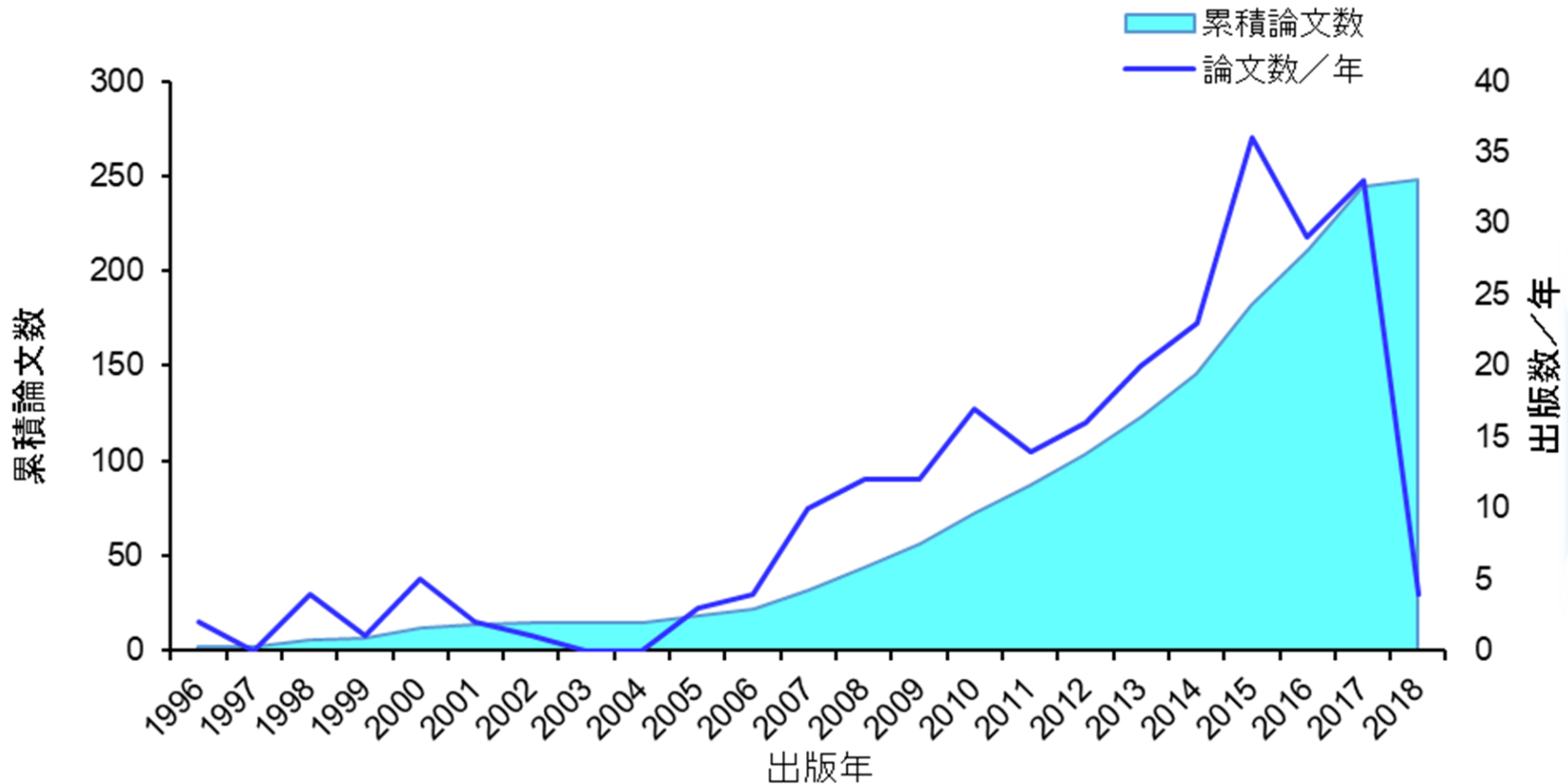
遺伝的アルゴリズム in 農業土木学会論文集

著者名	論文名	出版日付	巻	号	ページ
秋葉,菊池,加納	人工生命による最適値探索手法と斜面安定解析における最小安全率同定への適用	2005	73	6	549-557
福田,平松,森,四ヶ所	農業用水路におけるメダカの生息場選好性のあいまいさに関する数理表現	2005	73	5	505-511
小林,青野,山本,青山	電気探査によるコンクリート構造物の診断手法の検討	2004	72	6	671-677
平松,福田,四ヶ所	ファジィ推論によるメダカの環境応答モデルの開発	2003	71	6	789-796
藤原,田中丸,畑,多田	進化戦略による流出モデル定数の最適同定	2003	71	5	679-689
一恩,平松,河地,北村	拡張した遺伝的アルゴリズムによる地下水汚染面源の同定について	2003	71	3	413-422
森井	ロックファイルを通る非線形流れのパラメータ推定	2002	70	2	215-221
平松,河地,一恩	遺伝的アルゴリズムを用いた管水路網の最適設計	2001	69	5	647-653
一恩,平松,河地,北村	遺伝的アルゴリズムを用いた地下水面源汚染フラックスの長期変動推定法	2001	69	5	593-601
Fazal,平松,河地	有限要素法と遺伝的アルゴリズムを用いた漏水性不圧帯水層の最適揚水量決定	1999	67	6	659-670
板垣,平松,竹内,河地	遺伝的アルゴリズムを用いた貯水池群管理	1999	67	5	587-592
猪迫,中野,黒田,吉田	ラインソース灌漑のための畑地灌漑スケジューリングモデル	1998	66	6	975-983
Fazal,平松,河地	遺伝的アルゴリズムによるタイスマデルに準拠した帯水層の透水量係数と貯留係数の推定	1998	66	3	477-485
白谷,戸原,四ヶ所,井上	麦作圃場からの窒素流出簡易評価モデルの開発	1996	64	1	107-113
白谷,戸原,四ヶ所,井上	麦作圃場における窒素流出のモデル化	1996	64	1	97-105
平松,四ヶ所,森	多層パーセプトロンモデルによる河川感潮域の塩素量解析	1995	63	4	83-92
田中丸	タンクモデル定数の大域的探索	1995	178		503-512

ファジィ推論 in 農業土木学会論文集

著者名	論文名	出版日付	巻	号	ページ
張,前田,河地	水田への灌漑用水配分のためのファジィ最適化モデル	2007	2007	249	285-292
福田,平松,森,四ヶ所	農業用水路におけるメダカの生息場選好性のあいまいさに関する数理表現	2005	73	5	505-511
平松,福田,四ヶ所	ファジィ推論によるメダカの環境応答モデルの開発	2003	71	6	789-796
Hasan Mahbub,松井,水谷,後藤	ファジィ推論を用いた農業用水の取水量操作モデル	2001	69	6	731-738
河地,土原,平松,一恩	不圧海岸地下水流解析のためのファジィ則内蔵スキーム	2001	69	5	673-678
齊,笹尾,三原,安富,青山	耕作放棄に係る集落特性分析におけるファジィ・クラスタリング手法の適用とその有効性	1997	65	5	587-594
HASAN Mahbub,後藤,水谷,松井	プロダクションルール最適化のためのパラメータ化手法の開発	1997	1997	191	595-602
筑紫,V. F. Eduardo,豊田	ファジィ推論によるドリップ灌漑の灌水量の制御	1996	64	6	909-916
稲垣,國武	ファジィ推論によるパイプラインのデジタル制御	1996	1996	183	365-377
猿渡,四方田	水田用水の戦略的管理とニューロネットワークの導入:水管理システム化に関する研究 (V)	1994	1994	172	95-103
大野	不確定性のある浸透特性情報を考慮した飽和-不飽和浸透流解析	1993	1993	165	1-8
猿渡,四方田	水管理システム化に関する研究-3-ファジィ推論による灌漑用水量予測システムの検討	1992		158	p29-36
猿渡,四方田	水田用水予測管理システムの実証的検討:水管理システム化に関する研究 (IV)	1992	1992	161	75-83
猿渡,四方田	ファジィ推論による灌漑用水量予測システムの検討:水管理システム化に関する研究 (III)	1992	1992	158	29-36
大野,近藤	ファジィ理論による多目的非線形計画問題の最適化を利用した土構造物の飽和-不飽和浸透特性の同定	1991	1991	156	33-40
青山,庄野	フィルダムの耐震性診断に関する二つの試み:年確率地震加速度スペクトル推定と耐震性の簡易診断システム	1991	1991	152	91-99
猿渡,四方田	ファジィ推論による水田の水需要予測の検討:水管理システム化に関する研究 (II)	1991	1991	152	29-37
大野,近藤	土構造物内の浸透流解析についてのファジィ理論の応用	1988		135	p67-73

ANN & irrigation in WOS (by Feb 2018)



大量絶滅が起こりそうな現状をどうするのか？

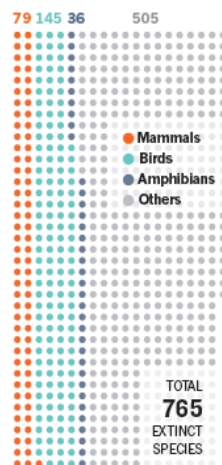
FEATURE NEWS

Life under threat

Thousands of species are currently deemed to be threatened, but the true number of species at risk of extinction may be much higher. Estimates suggest that between 500 and 36,000 species might be disappearing each year. The best data are for well-studied groups — mammals, birds and amphibians. Much less is known about threats to other groups, such as insects and fish.

ALREADY EXTINCT

TOTAL DOCUMENTED SINCE 1500



Mammals
1,199
THREATENED SPECIES
26% of described species

CURRENTLY THREATENED



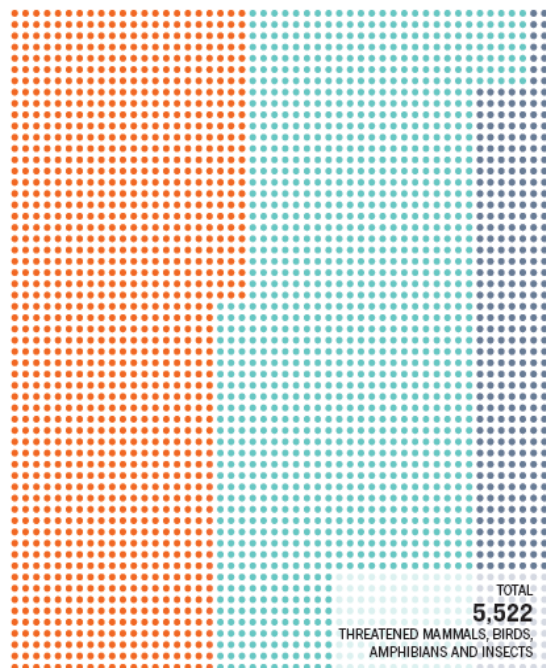
Birds
1,373
THREATENED SPECIES
13% of described species



Amphibians
1,957
THREATENED SPECIES
41% of described species

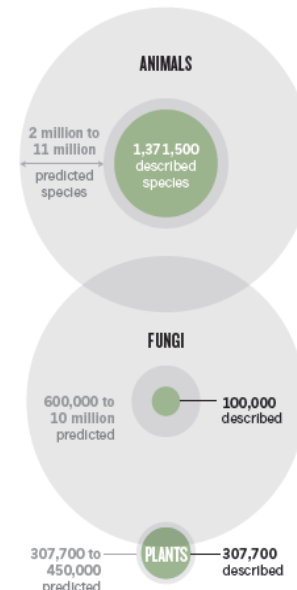


Insects
993
THREATENED SPECIES
(Only 0.5% of roughly 1 million described have been evaluated. Number of living species may exceed 5 million)



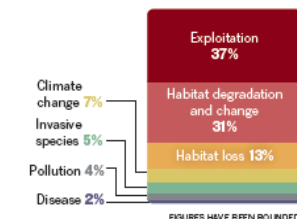
How many species are there?

Estimates of the number of species of animals, fungi and plants vary significantly. That uncertainty clouds understanding of how many species are threatened and how many are going extinct.



Main threats

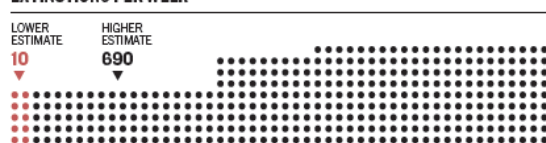
Hunting, fishing and other forms of exploitation are a major factor in declines in animal populations, according to the Living Planet Index. Habitat degradation and loss are also dominant threats. Climate change is expected to become a bigger factor over time.



March towards mass extinction

Mass extinctions — loss of 75% of existing species — have happened 5 times in the planet's history. If there are 5 million animal species and they are disappearing at rate of 0.72% per year (the upper end of estimates), a sixth mass extinction could happen by the year 2200. At the low end of the estimated range, a mass extinction would not happen for thousands of years.

EXTINCTIONS PER WEEK



BY RICHARD MONASTERSKY | GRAPHIC BY SW INFOGRAPHIC

 PHOTO CREDITS: *B. parvus* and *N. americanus*: Joel Sartore/National Geographic Creative; *S. demersus*: Life on white/Alamy; *R. sumneri*: Joel Sartore/National Geographic Creative/Getty.

 SOURCES: Already Extinct, Currently Threatened: IUCN Red List. How many species are there?: S. L. Pimm et al. *Science* 344, 12467-52 (2014); B. R. Schaffers et al. *Trends Ecol. Evol.* 27, 501-510 (2012); IUCN Red List. March towards mass extinction: Pimm et al.; C. Mora et al. *Science* 341, 237 (2013). Main threats: WWF *Living Planet Report* 2014.

世界的な潮流：分布データの公開



Global Biodiversity Information Facility

Free and Open Access to Biodiversity Data

639,472,777

OCCURRENCES

1,611,321

SPECIES

15,249

DATASETS

774

DATA PUBLISHERS

Sharing biodiversity data for re-use

[Learn about GBIF](#)

[Publish your data through GBIF](#)

[Technical infrastructure](#)

Providing evidence for research and decisions

[Using data through GBIF](#)

[Enabling biodiversity science](#)

[Supporting global targets](#)

Collaborating as a global community

[Current Participants](#)

[How GBIF is funded](#)

[Enhancing capacity](#)

Search news items and information pages...

Search

- 研究や政策決定などの目的に使用する生物多様性情報基盤の整備
- 生物多様性情報の集積と提供
- 情報集積・解析ツールの開発
- 生物多様性情報に関わる活動の支援と能力開発

よりよい 分布データの 管理のために

理解



マルチスケール
の空間モデル化



トレンド解析と
予測†



生物学的
システムの
モデル化†



可視化と
普及†



得るべき
新規データの
優先順位付け

証拠



用途適合性と
アノテーション



分類学的体系



統合された
分布データ



集約された種
の形質データ



包括的な
知見への
アクセス

データ



出版物



コレクションと
標本



野外調査と
観察



塩基配列と
ゲノム



自動センシング
およびリモート
センサによる観察†

文化



オープンアクセス
と
再利用の文化



標準データ規格



永続的保存と
アーカイブ†



政策による
インセンティブ†



生物多様性に
関する知見の
ネットワーク

進歩のキー：

——— 顕著

——— 一定の成果

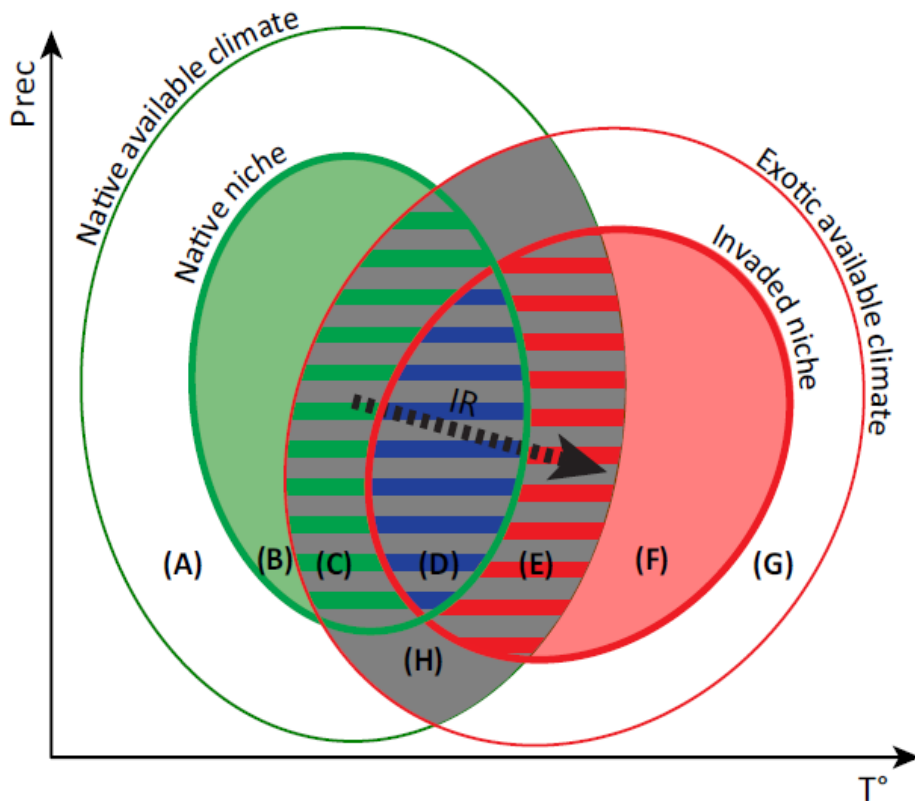
——— 限定的

——— 最小

† 緊急性の高い問題と考えられるが、これまでの進展は限定的なものにとどまっている。

生物の空間分布から分かること

- ❖ どこに生息しているか？
- ❖ どのような環境条件が分布に影響を与えているのか？

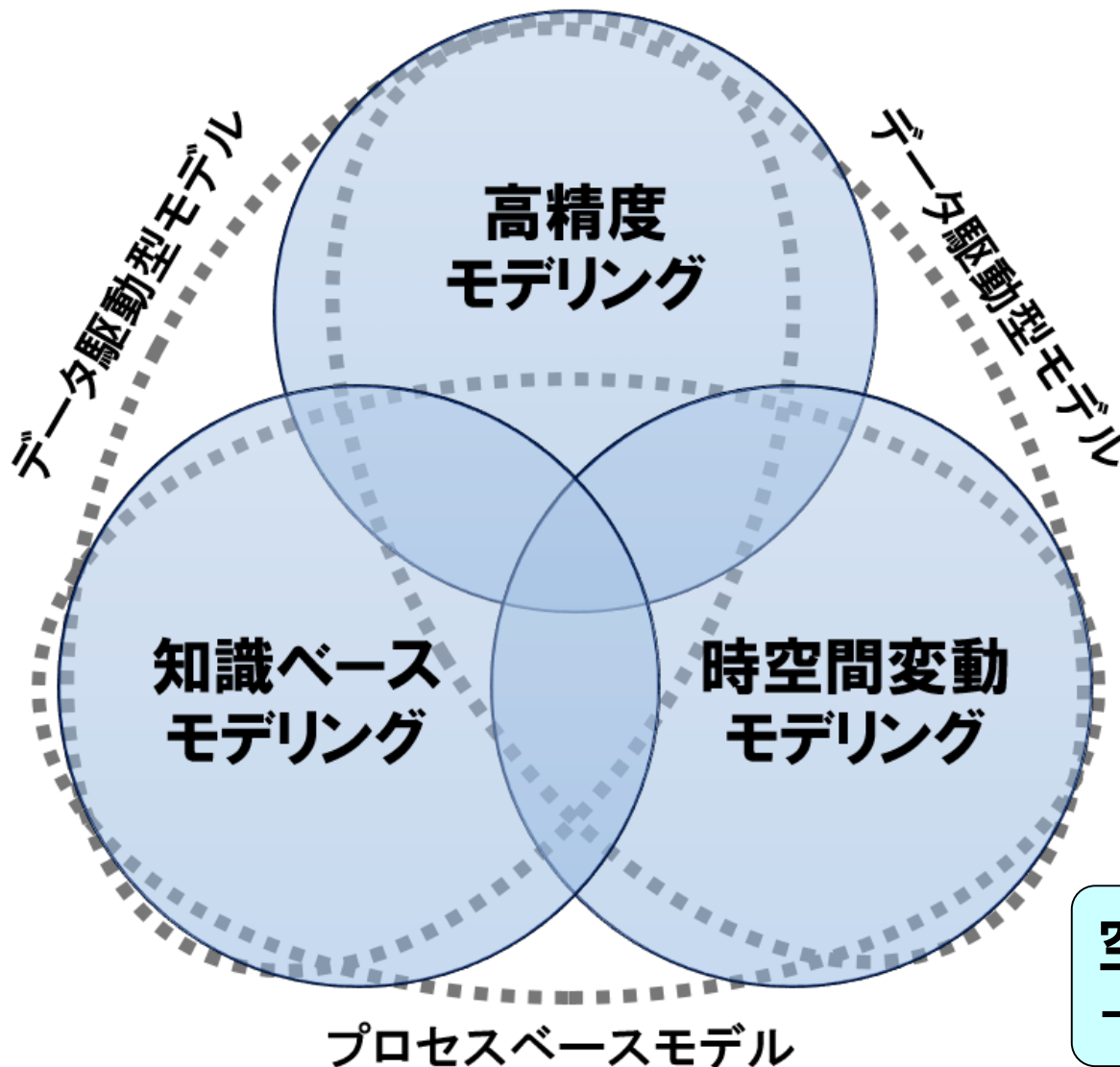


TRENDS in Ecology & Evolution

例えば、ニッチシフトの解析も可能
(データが揃っていれば...)

Guisan et al. (2014)

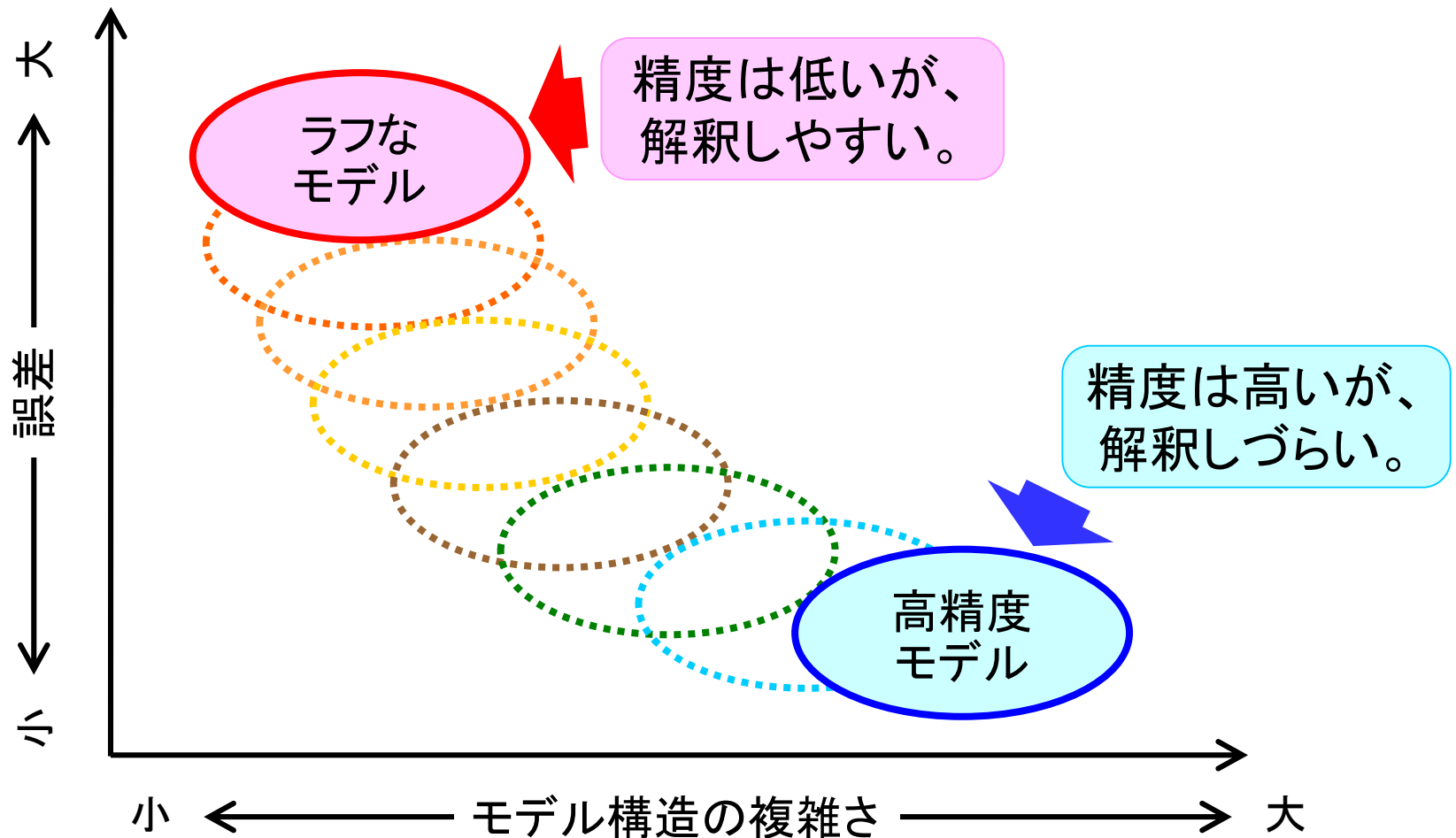
数理解析手法の概要



空間分布モデル
→データ駆動型モデル

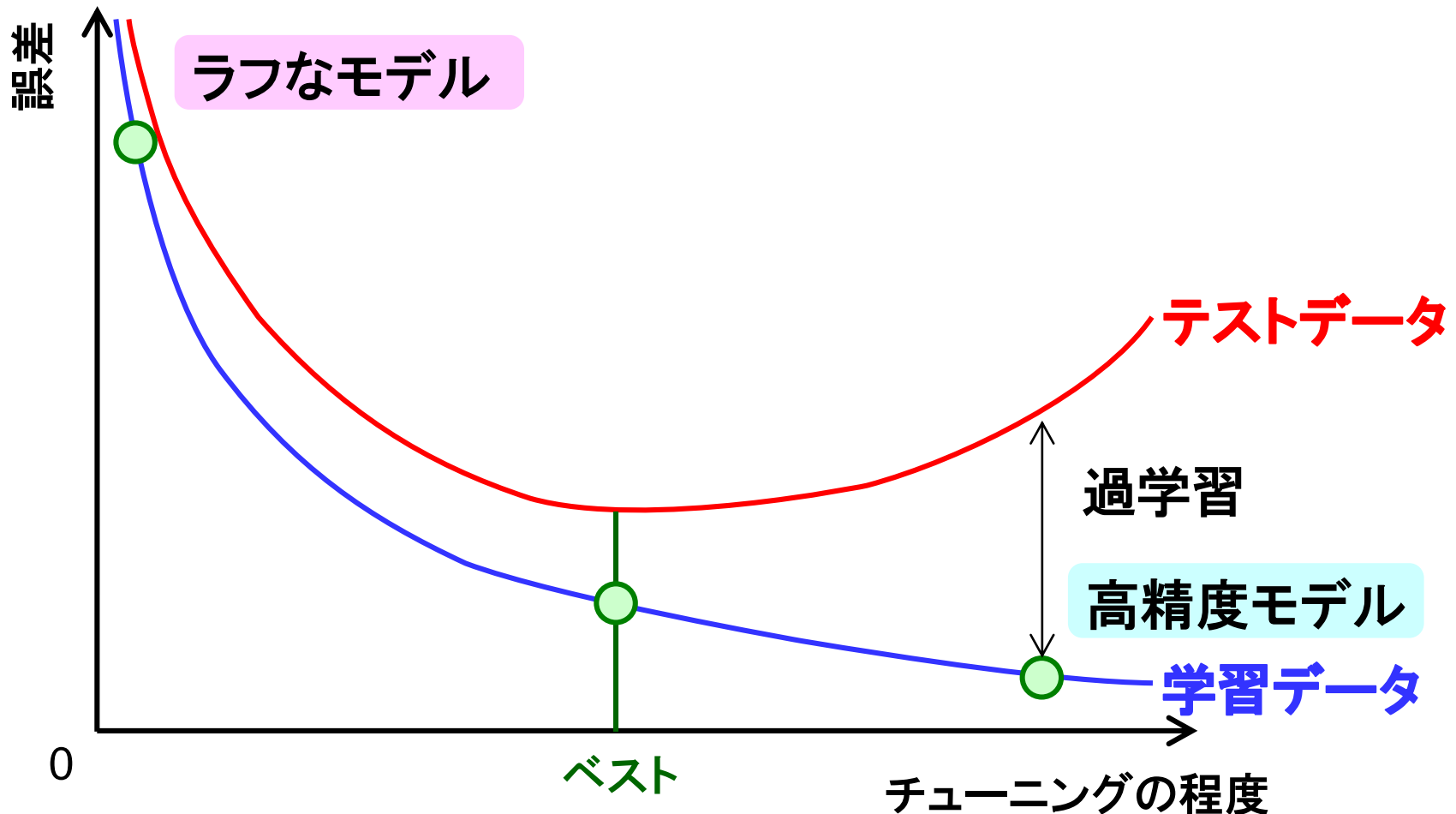
空間分布モデルの使用上、注意を要すること

- ❖ 様々なモデルがあり、特徴がバラバラ
→どのモデルを使用するのが良いのか?

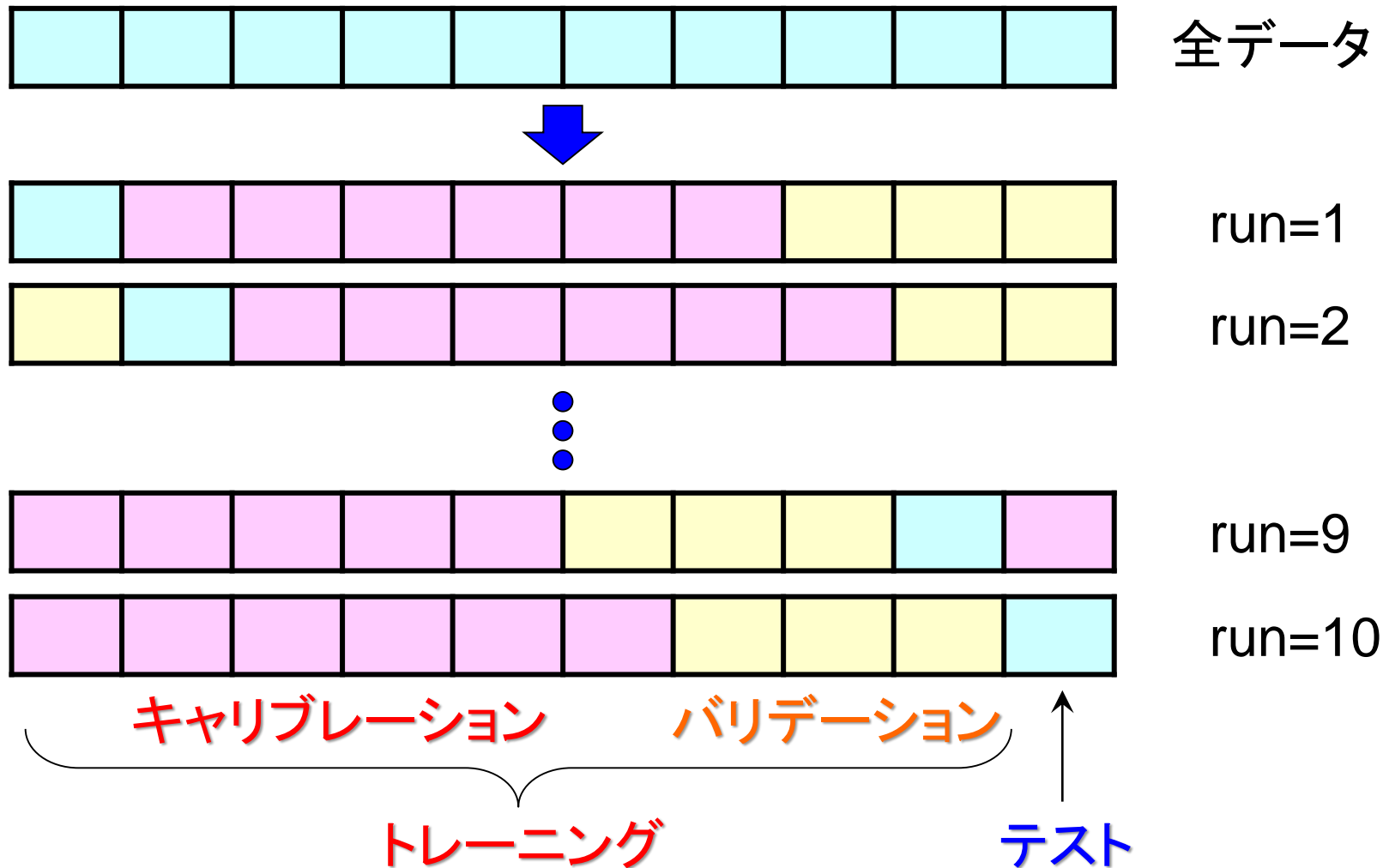


空間分布モデルの使用上、注意を要すること

❖ 適切な方法でモデルを構築する必要あり！

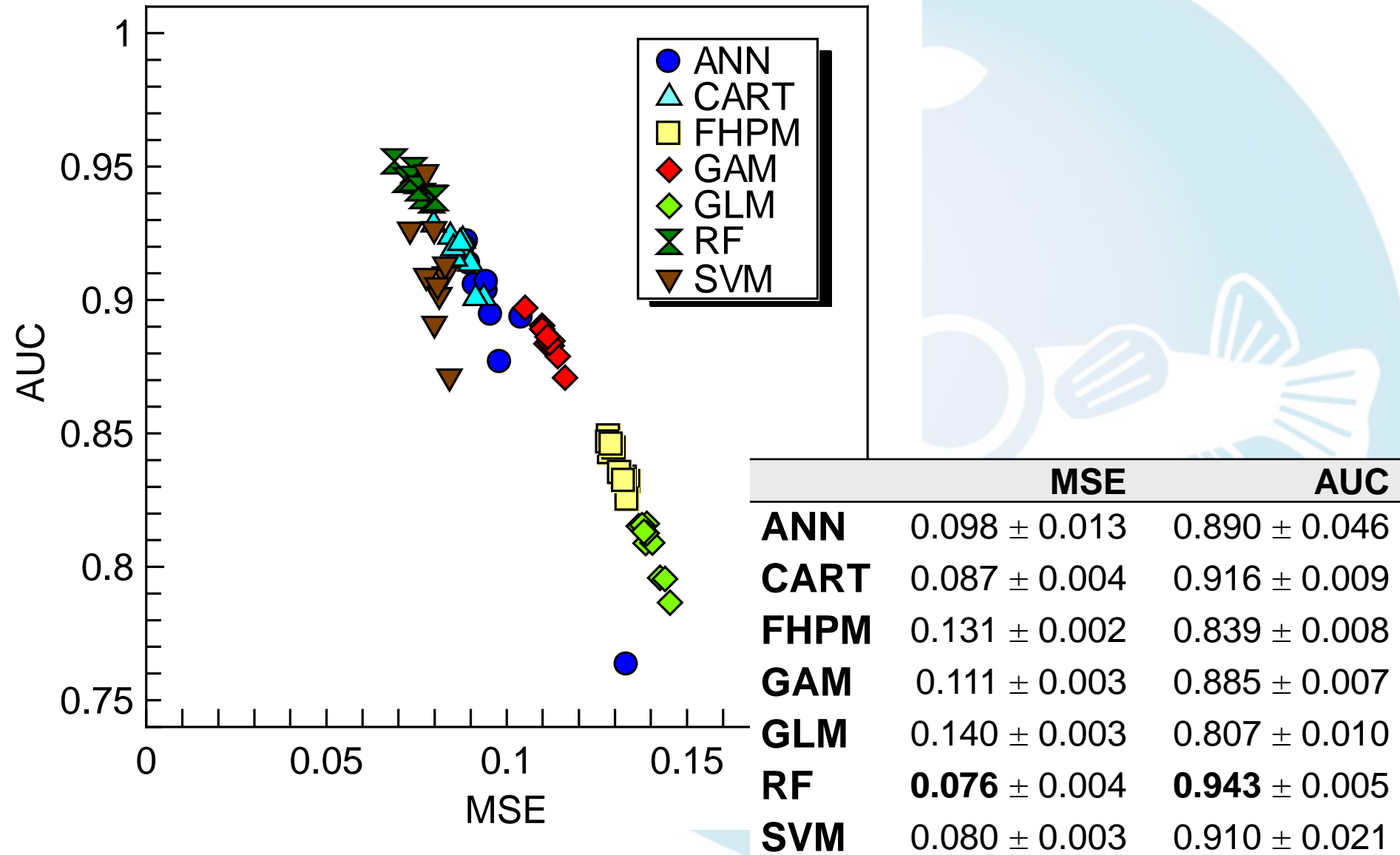


10-fold Nested Cross-Validation



結果の平均と標準偏差に基づいて評価

空間分布モデリング手法の比較事例



適切な評価指標を選択する必要あり

混同行列		データ	
		在	不在
モデル	在	TP	FP
	不在	FN	TN

$$TP=a; FP=b; FN=c; TN=d$$

評価指標
の例

$$CCI = (a + b) / n$$

$$Sn = a / (a + c)$$

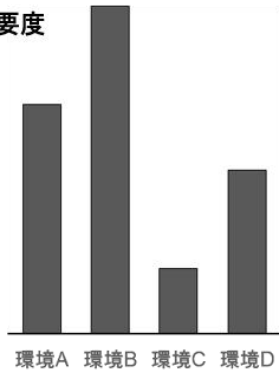
$$Sp = d / (b + d)$$

$$\kappa = \frac{((a + b) / n) - ((a + b)(a + c) + (c + d)(d + b)) / n^2}{1 - ((a + b)(a + c) + (c + d)(d + b)) / n^2}$$

$$TSS = Sn + Sp - 1$$

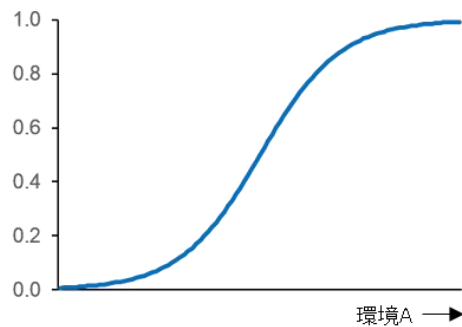
知識獲得

(a)変数の重要度

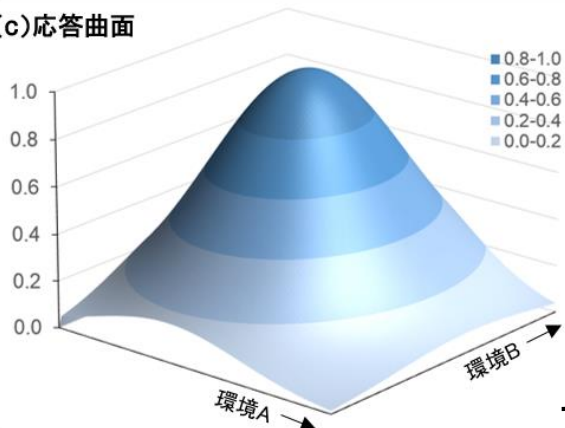


どの環境変数が重要なのか？

(b)応答曲線



(c)応答曲面



どのような環境条件が重要なのか？

Q:空間分布モデリングは難しいのか？

A:アルゴリズムは複雑だが、今では小学生でも使える！

- ❖ 適切なデータ(生物分布と環境条件)の収集
- ❖ 適切な数理解析(モデリングと統計解析)
- ❖ 見やすい説得力のある図表
- ❖ 素敵なストーリー

水資源計画学研究室

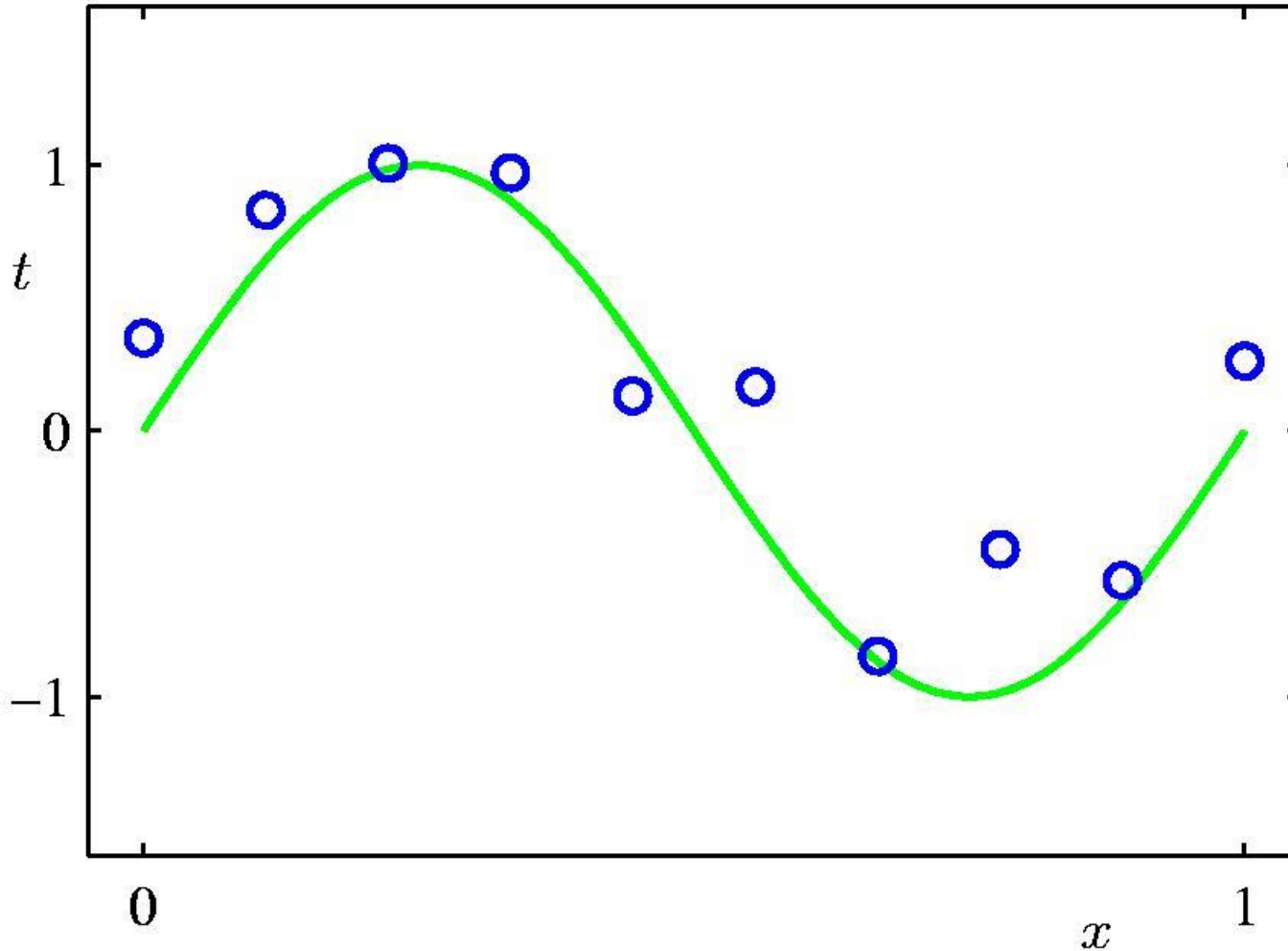
<http://shinjifukuda-medaka.com/>



『農業』-『水』-『生物』
持続可能な水資源管理と
生態系と調和した
農業生産について考える。

Sample data

Bishop (2006)
Pattern Recognition
and Machine Learning

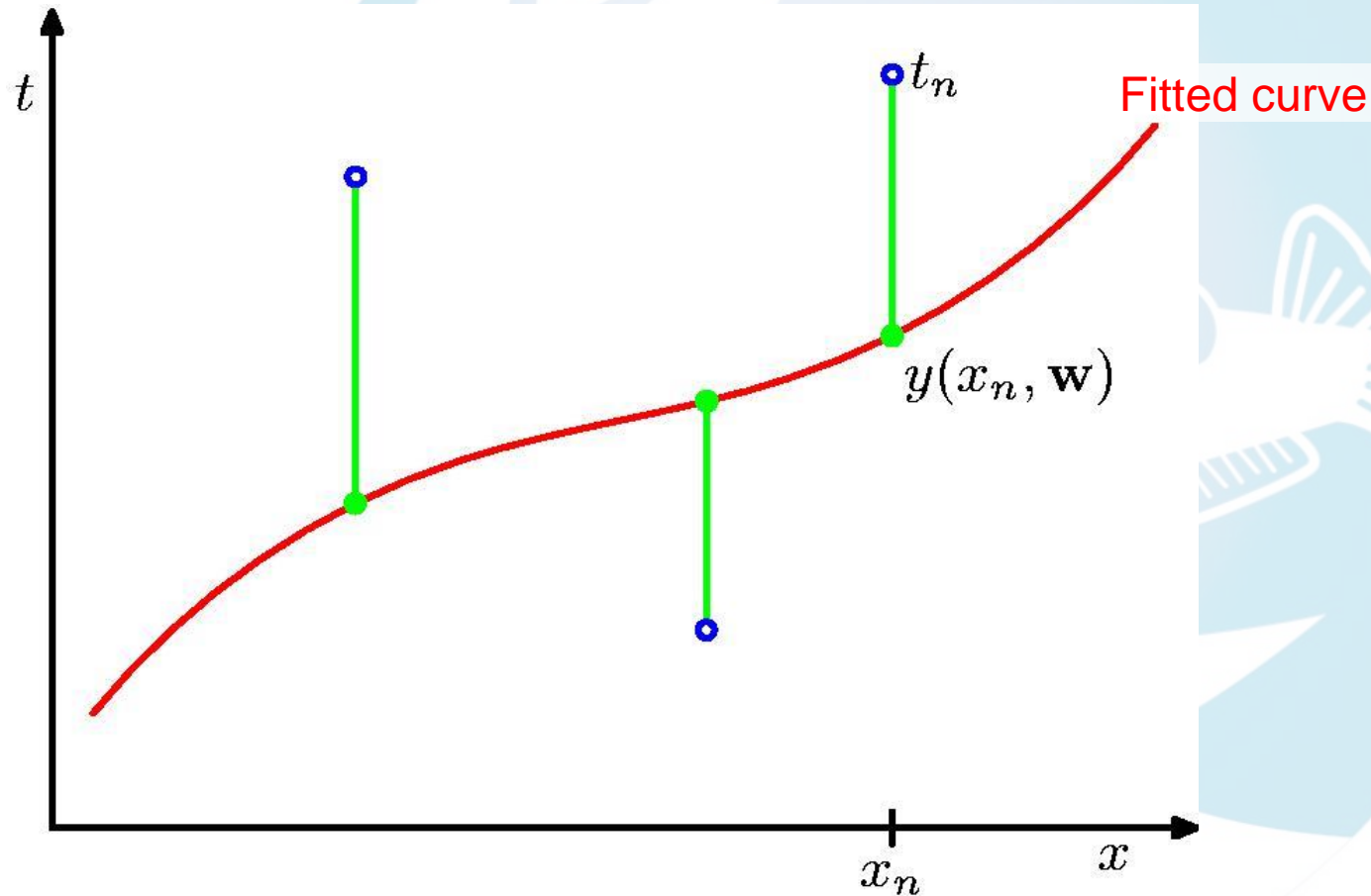


Polynomial curve: $y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$

Least squares fitting

Bishop (2006)
Pattern Recognition
and Machine Learning

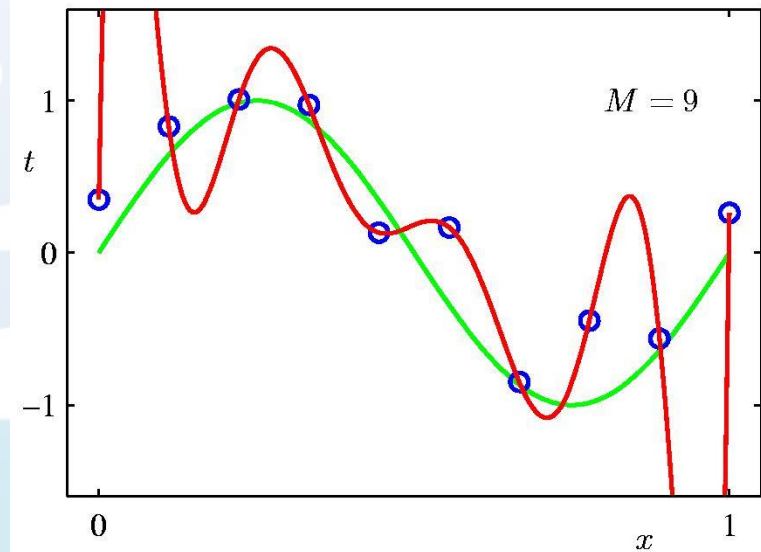
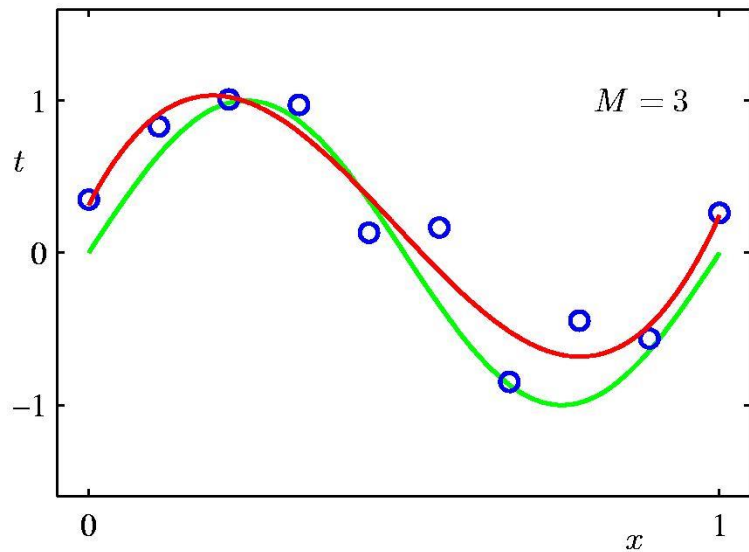
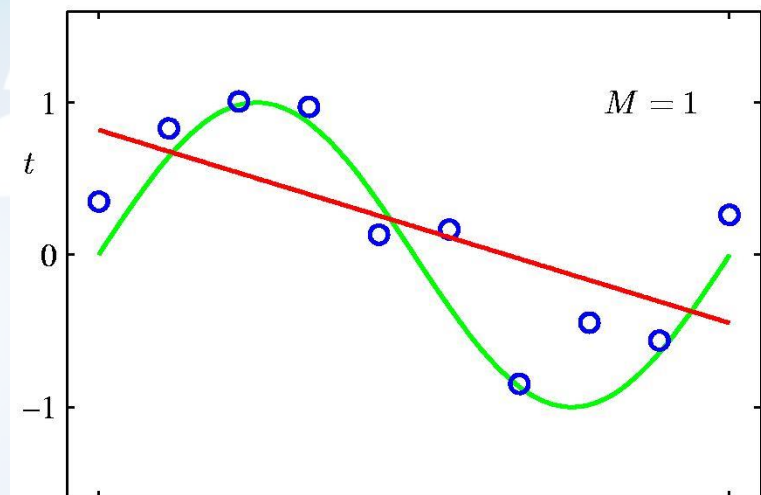
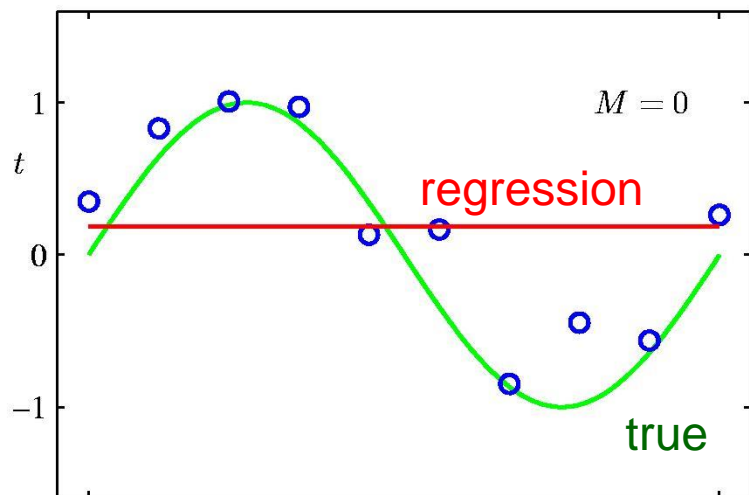
In least squares method,
sum-of-square error function is used when fitting a model



$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

Fitting results (different complexity)

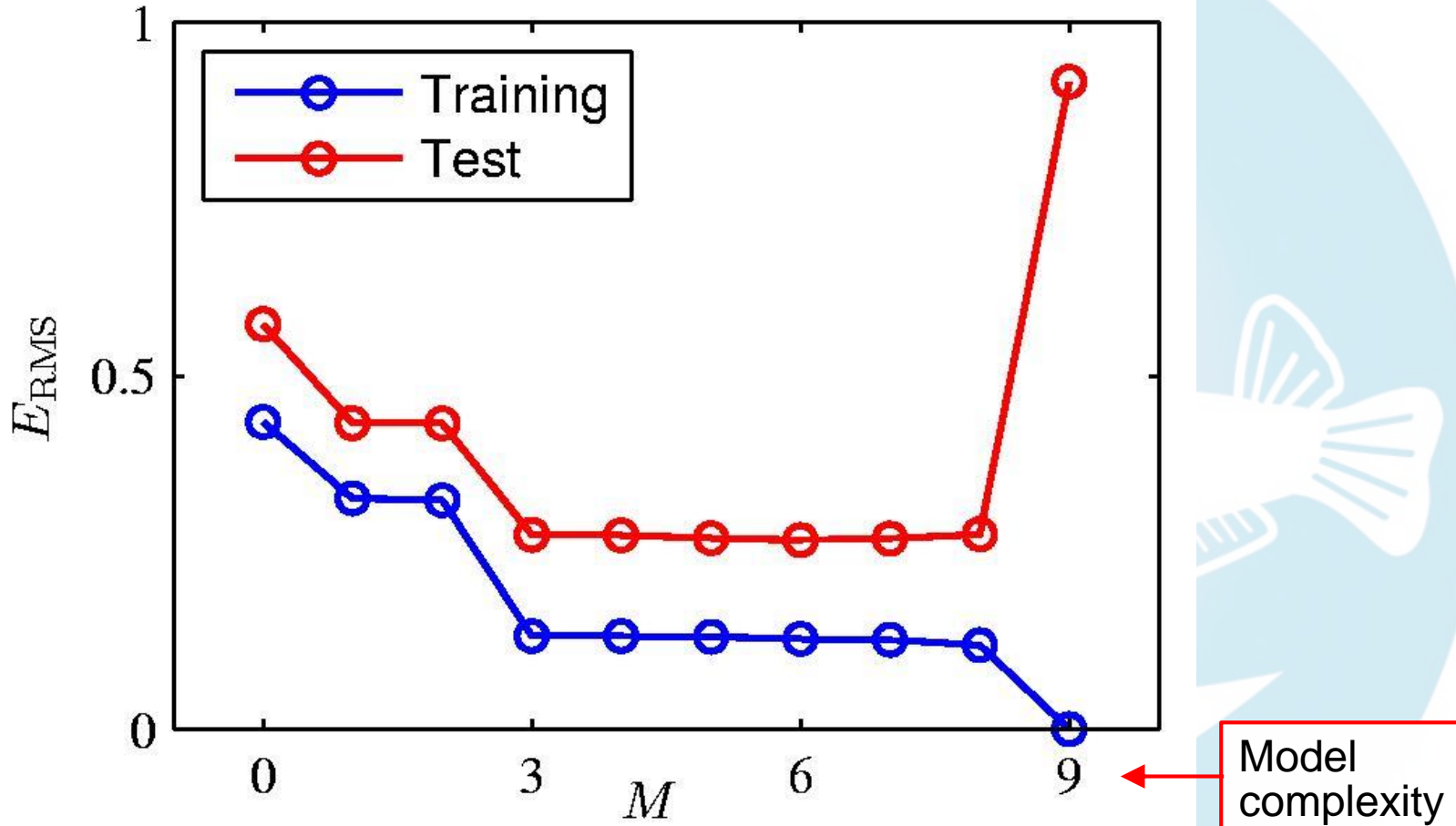
Bishop (2006)
 Pattern Recognition
 and Machine Learning



Polynomial curve: $y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$

Overfitting

Bishop (2006)
Pattern Recognition
and Machine Learning



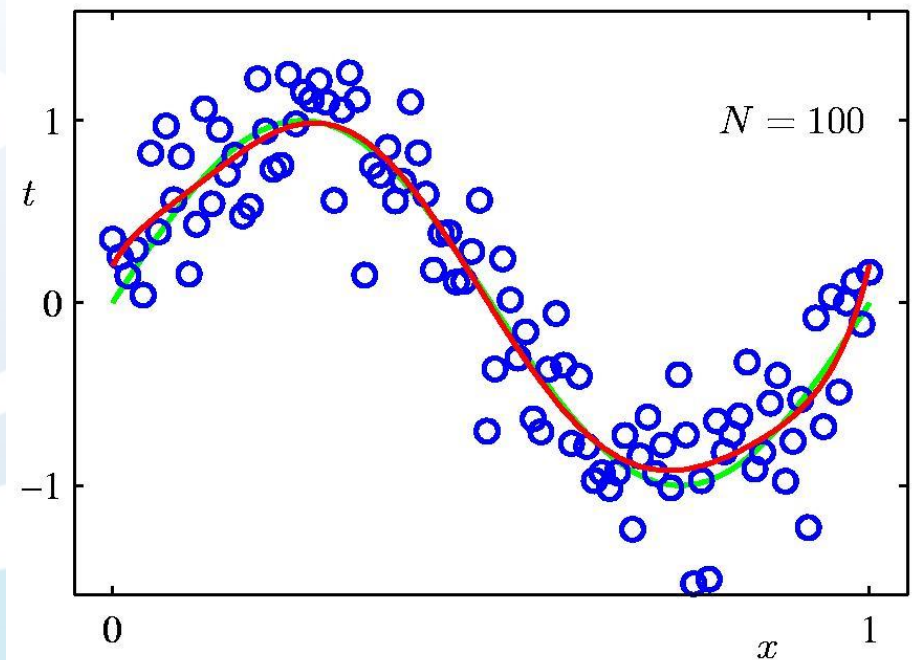
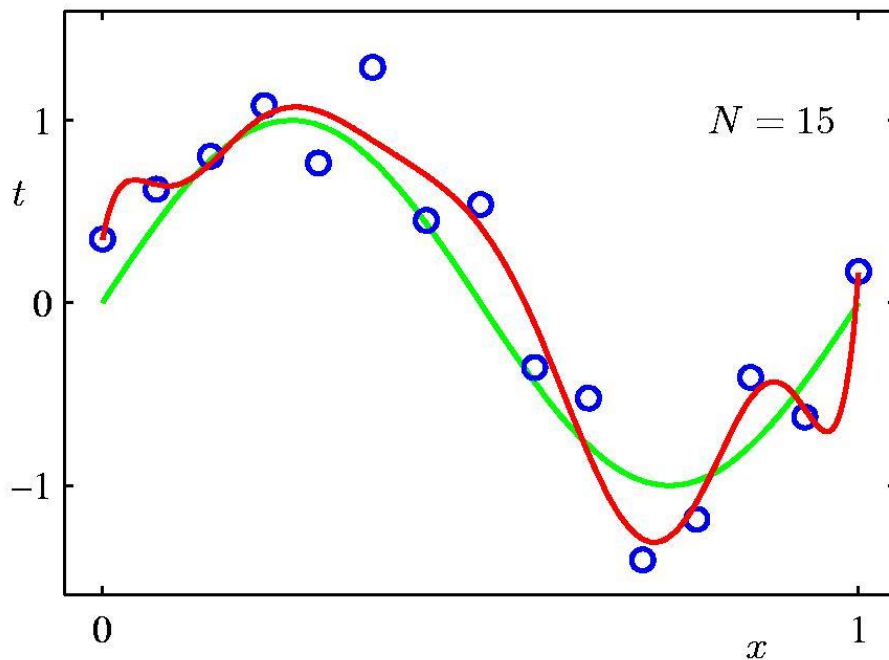
Root-Mean-Square (RMS) Error:
$$E_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{d,i} - Y_{m,i})^2}$$

Solutions for overfitting: larger size of a data set

Bishop (2006)
Pattern Recognition
and Machine Learning

9th Order Polynomial

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$



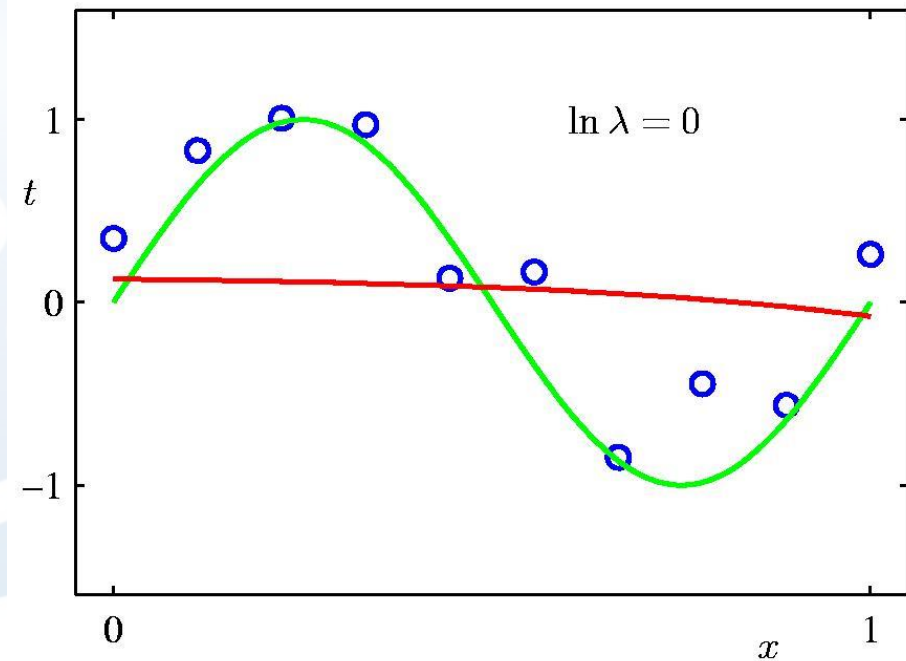
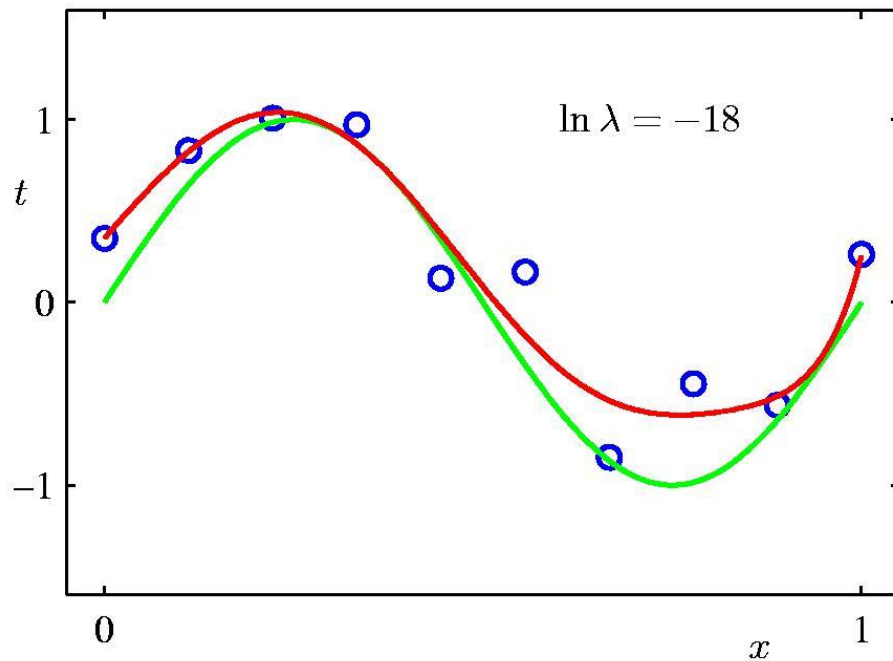
Small

Size of the data set

Large

Regularization: a conventional method in statistics

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$



Bishop (2006)
Pattern Recognition
and Machine Learning

Example of regularization: fuzzy neural networks

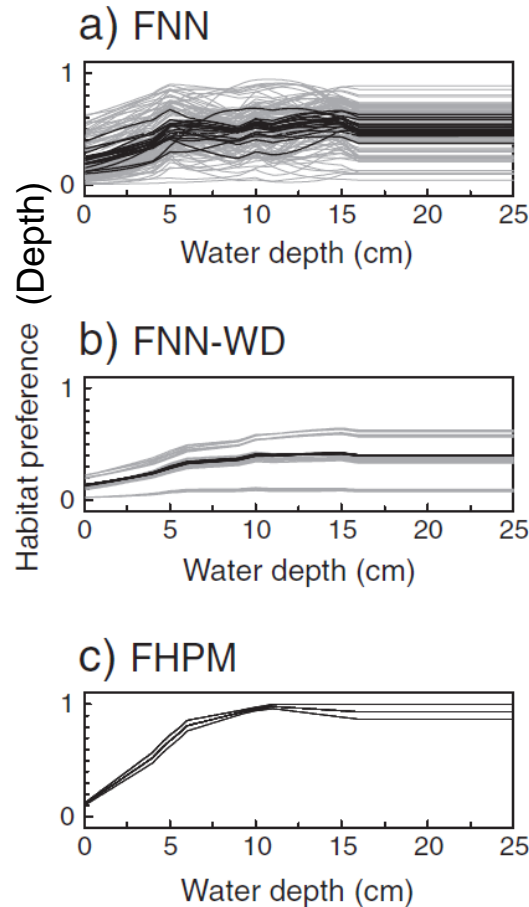
Fukuda, S., Assessing the applicability of fuzzy neural networks for habitat preference evaluation of Japanese medaka (*Oryzias latipes*), *Ecological Informatics*, 6, 286–295, 2011.

Results of habitat preference evaluation using fuzzy neural networks (FNN):

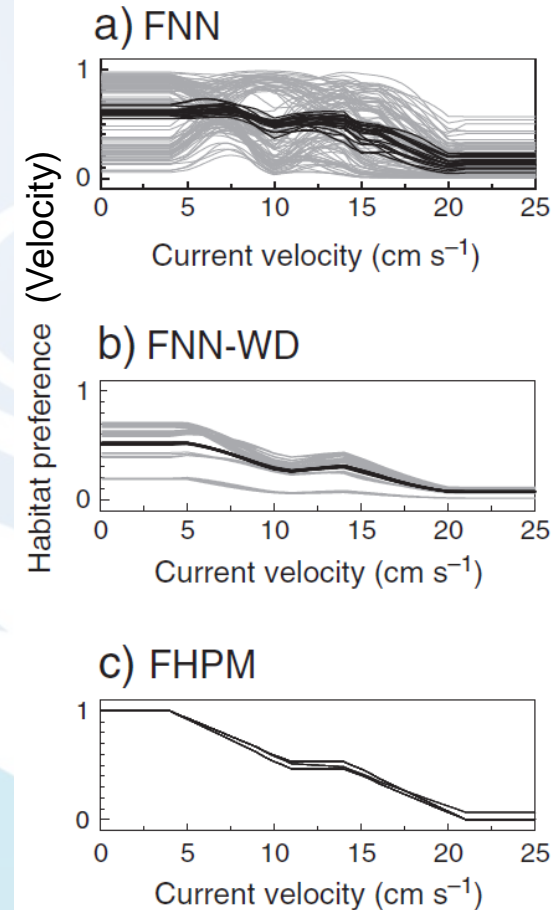
a) FNN without **weight decay**

b) FNN with **weight decay**

c) Fuzzy Habitat Preference Model



(i) 14 October, 2004



(i) 14 October, 2004