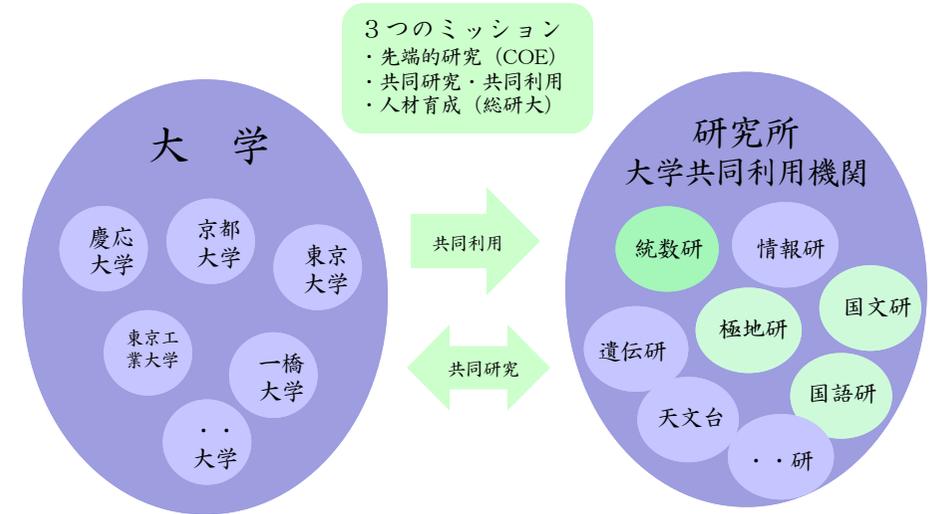


サイバー世界のモデリング

大学共同利用機関法人
情報・システム研究機構
北川 源四郎

数学月間懇話会
東大数理科学研究科 7/22/2011

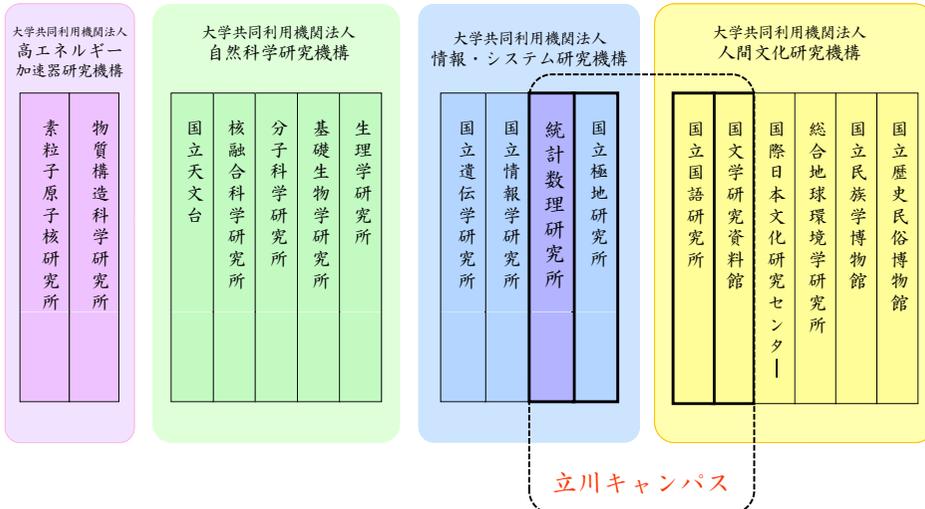
大学共同利用機関とは



大学共同利用機関法人
情報・システム研究機構

2

4機構・17大学共同利用機関



大学共同利用機関法人
情報・システム研究機構

3

概要

1. 情報化にともなう社会と科学研究の変化
2. 知識社会における科学的方法論
3. データ中心科学の課題
4. 統計的モデリング

大学共同利用機関法人
情報・システム研究機構

4

歴史の転換点

歴史にも境界がある。

・・・数百年に一度、際立った転換が起こる。社会は数十年をかけて、次の新しい時代のために準備する。世界観を変え、価値観を変える。社会構造を変え、政治構造を変える。技術と芸術を変え、機関を変える。やがて50年後には、新しい世界が生まれる。

・・・この転換は2010年ないし20年まで続く。

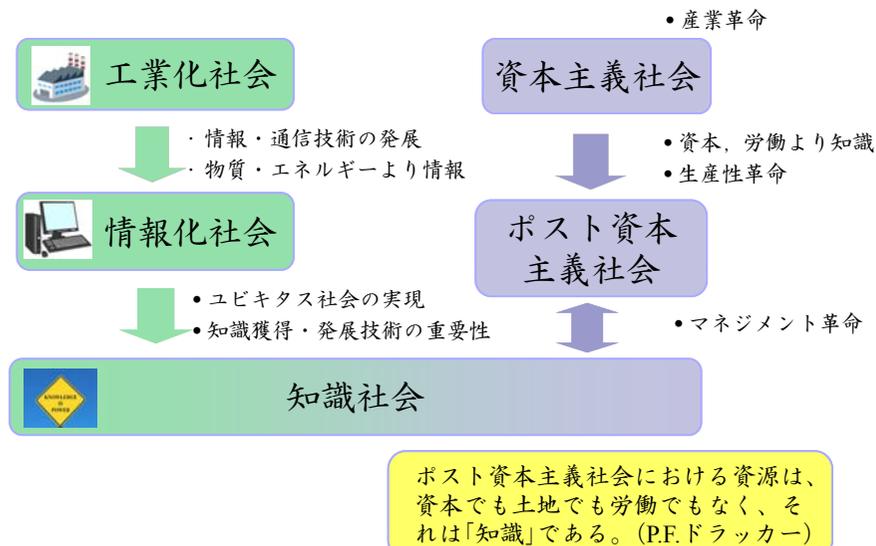
P.E. Drucker (1993) 『ポスト資本主義社会』

社会の変化と学術の変化

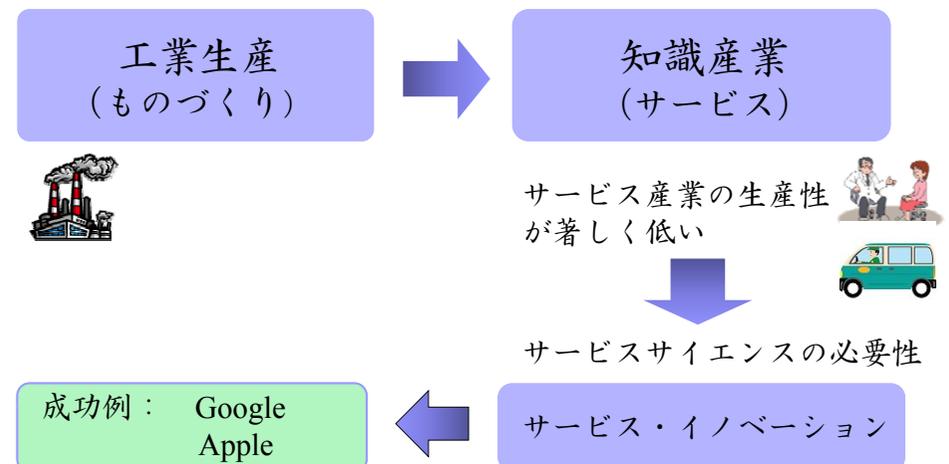
情報化・知識化によって、震災以前から社会も学術・科学技術も大きく変化している。

数理科学も変わらなければならない。

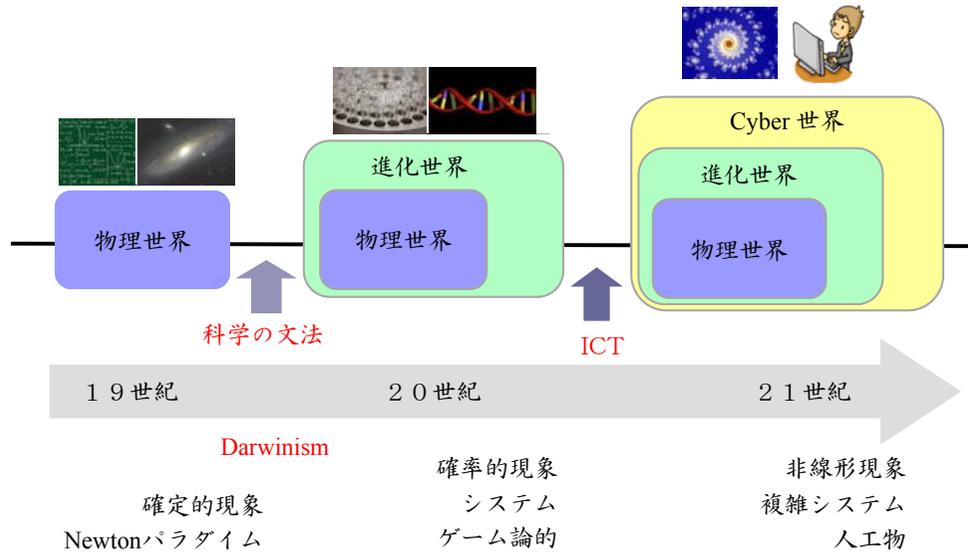
社会の変化



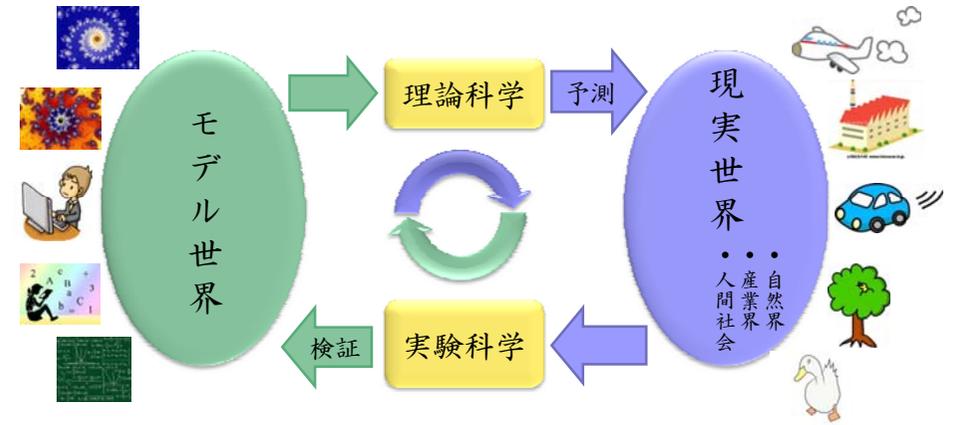
産業の変化



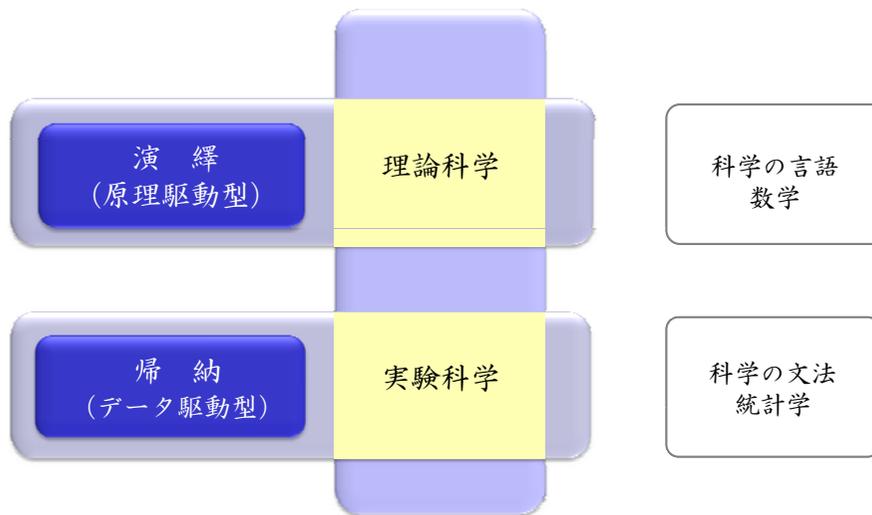
科学の対象の拡大・変化



モデル世界と現実世界のかけ橋

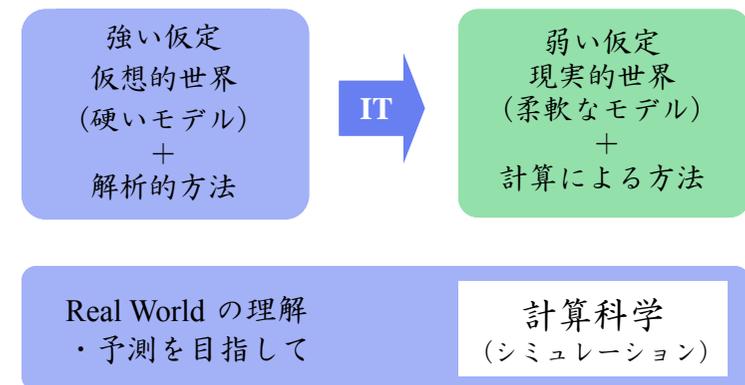


科学的方法論 (実験と理論)

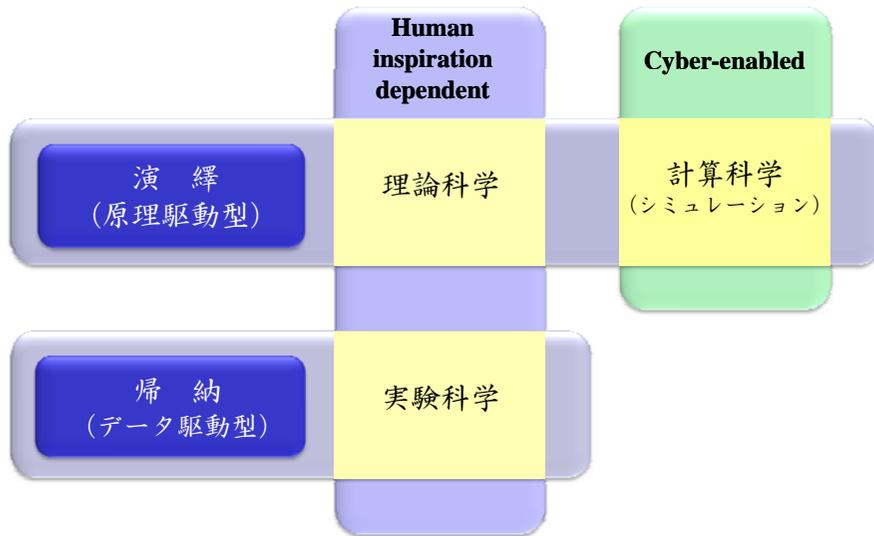


計算科学 - Cyber-Enabled Science -

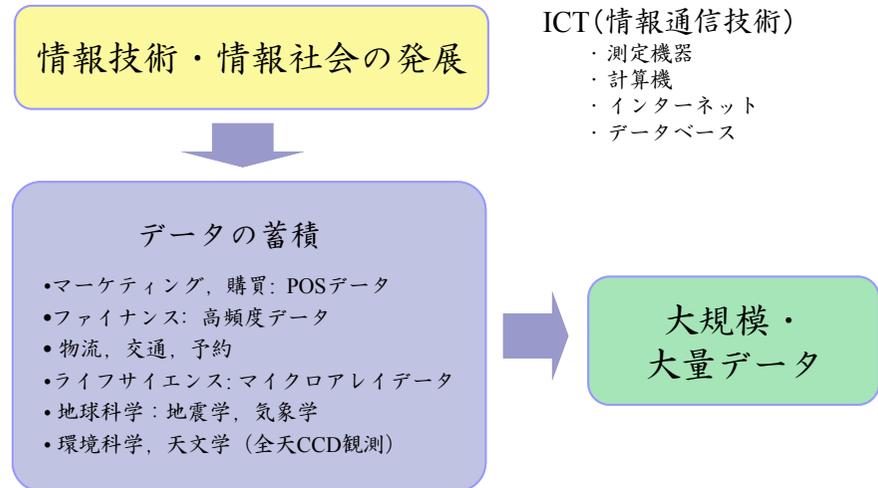
- ・非線形現象
- ・複雑・大規模システム (生命, 人工物)
- ・人間社会 (経済, 人間行動)



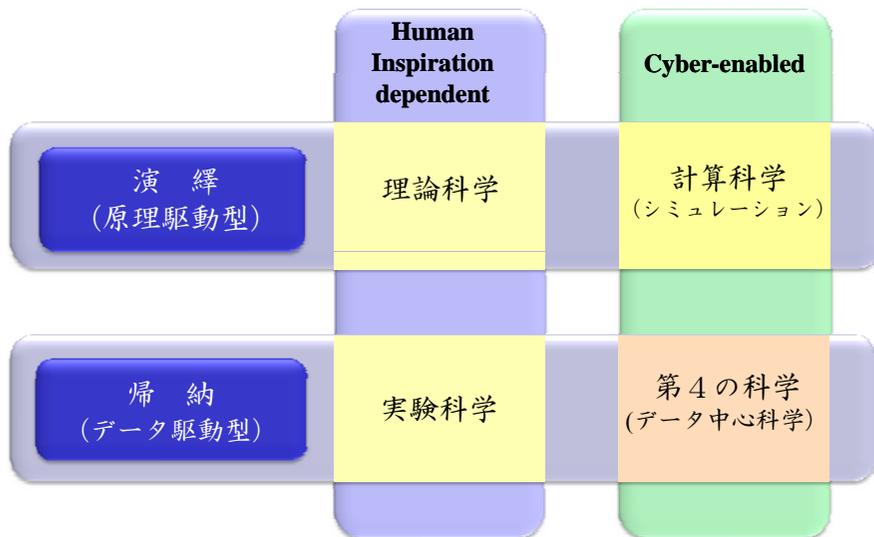
第3の科学 (計算科学)



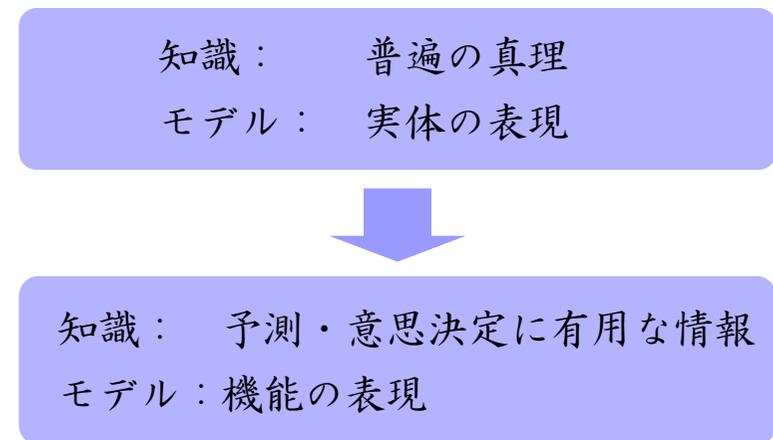
情報化社会と大規模・大量データ



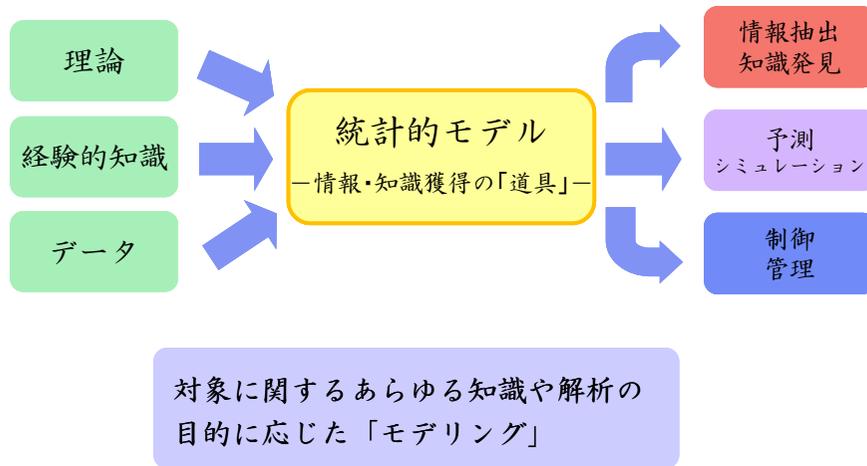
第4の科学 (データ中心科学)



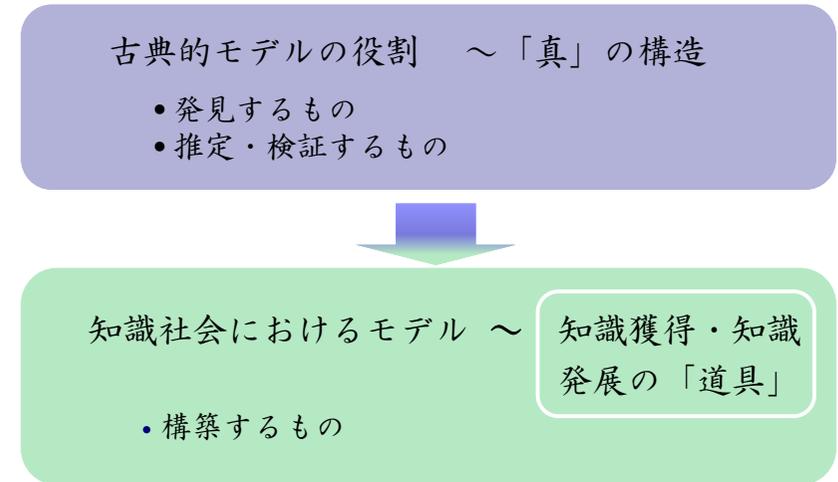
「知識」の変化とモデルの変化



知識社会におけるモデルの役割



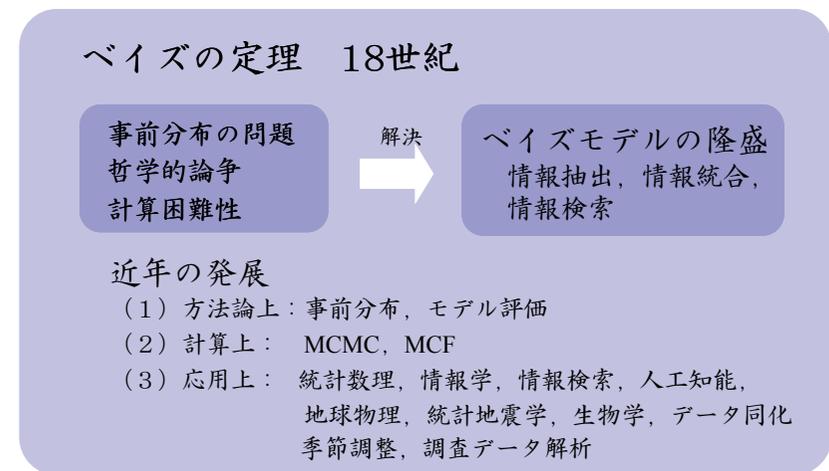
モデルからモデリングへ



モデリングによる異種情報の統合

1. 事前情報とデータ情報の統合
ベイズ推論
2. 時間発展とデータ更新
フィルタリング (Sequential filtering)
3. 普遍的知識と個別的情報の統合
個別化 (Personalization)
4. シミュレーションモデルとデータの統合
データ同化 (Data assimilation)
5. 知識発展のサイクル
能動的モデリング (Active modeling)

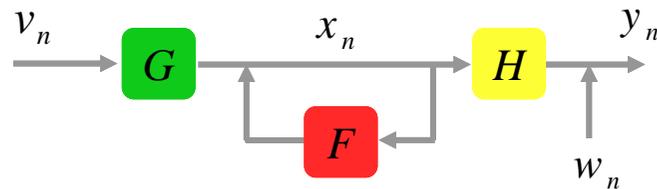
(1) ベイズモデリングの新展開



(2) 状態空間モデル：時間発展とデータ更新

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n && \text{状態モデル} \\ y_n &= Hx_n + w_n && \text{観測モデル} \end{aligned}$$

y_n 時系列 v_n システムノイズ
 x_n 状態ベクトル w_n 観測ノイズ



状態空間モデル

線形・ガウス型

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \\ y_n &= Hx_n + w_n \end{aligned}$$

非線形・非ガウス型

$$\begin{aligned} x_n &= f(x_{n-1}, v_n) \\ y_n &= h(x_n, w_n) \end{aligned}$$

一般型

$$\begin{aligned} x_n &\sim F(\cdot | x_{n-1}) \\ y_n &\sim H(\cdot | x_n) \end{aligned}$$

関数：非線形
分布：非ガウス型

条件付分布
離散状態・離散観測値

非ガウス型フィルタ・平滑化

一期先予測

$$p(x_n | Y_{n-1}) = \int_{-\infty}^{\infty} p(x_n | x_{n-1}) p(x_{n-1} | Y_{n-1}) dx_{n-1}$$

フィルタ

$$p(x_n | Y_n) = \frac{p(y_n | x_n) p(x_n | Y_{n-1})}{p(y_n | Y_{n-1})}$$

平滑化

$$p(x_n | Y_N) = p(x_n | Y_n) \int_{-\infty}^{\infty} \frac{p(x_{n+1} | x_n) p(x_{n+1} | Y_N)}{p(x_{n+1} | Y_n)} dx_{n+1}$$

(3) 個別化：統計的ものの見方の変化

帰納：個々の具体的事実から一般的知識へ

● 集団として捉える → 本質を捉える

大量生産・大量消費を目指した20世紀
個人に焦点をあてた科学技術へ

背景

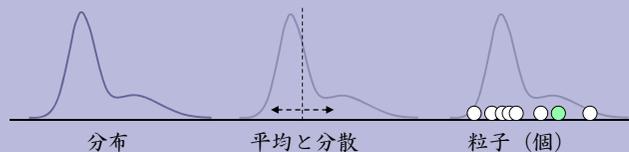
資源の有効利用のための選択と集中
価値観の多様化（個人、個性、個別、固有の尊重）

テラーメイド創薬
オーダーメイド医療
マイクロマーケティング
サービスイノベーション
教育、物流

平均を見る → 個性を捉える

(3) 個別化(Personalization, 平均から個へ)

●平均から個性へ



●本質は究極の条件付け



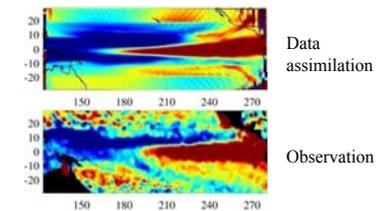
新NP問題 ($n \ll p$)
 ・マイクロレイ
 ・マーケティング

サービスの個別化
 ・テラーメード医療
 ・オーダーメード創薬
 ・マイクロ・マーケティング
 ・テラーメード教育
 ・ロジスティックス
 ・情報提供



(4) データ同化 (Data Assimilation)

地球環境シミュレーション:
 理論モデルと衛星観測データの統合

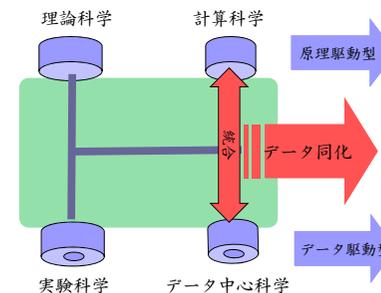


原理 (モデル) 駆動型アプローチと
 データ駆動型アプローチの統合

非線形フィルタリング・平滑化

適用領域:

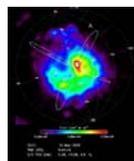
- ・エルニーニョ予測
- ・津波予測
- ・遺伝子ネットワーク推定



統計学におけるデータ同化研究

・リングカレントの3次元構造の解明

海老原祐輔 (極地研), M.-C Fok (NASA)
 大谷晋一, P.C.Brandt (ジョンホプキンス大)



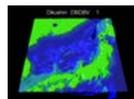
・気象・海洋結合

鍵本崇 (地球環境フロンティア研究センター)
 広瀬直毅 (九州大・応用力学研究所)



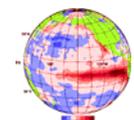
・津波データ同化

広瀬直毅 (九州大・応用力学研究所)



・ゲノム情報, 化学プラント

宮野, 井元 (東大医科研)

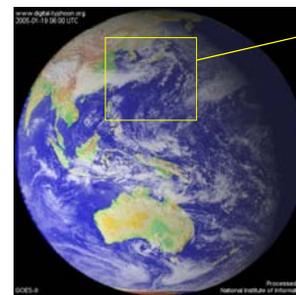


データ同化による高精度予測 (台風進路の予測)

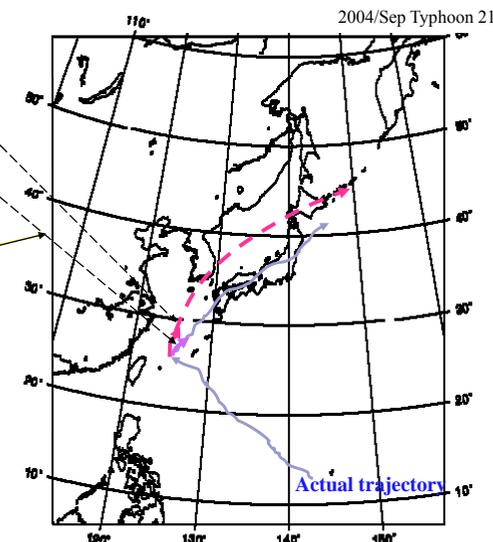
→: シミュレーションのみ

→: データ同化による予測

気象研のデータ同化研究者を
 客員教授に招聘して共同研究



(Taken from NII, Japan)



(5) 能動的モデリング

対象に関するあらゆる取得可能な情報(対象に関する理論, 経験的知識, 観測データ)およびモデリングの目的

事前情報とデータの持つ情報の統合

➡ ベイズモデリング

モデルを通して知識が得られる.
知識はモデルを改良する

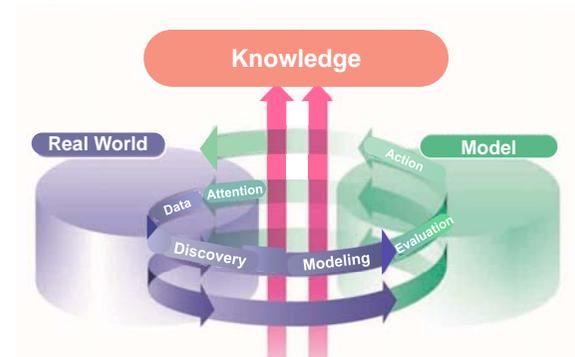
➡ 知識発展のスパイラル

知識発展のスパイラル

Knowledge is being applied to knowledge itself.

(P.E. Drucker: Post Capitalist Society)

↓
知識発展のスパイラル



モデリングの基本的考え方

- モデルはものの見方を表現したもの
- モデルは近似であり, 構築するもの
- 「真のモデル」は未知 (存在しない)

- 絶対的評価基準は利用できない
- BestではなくBetterなモデル

- データの増加, 知識の増加

➡ モデルの持続的改良・発展が必要

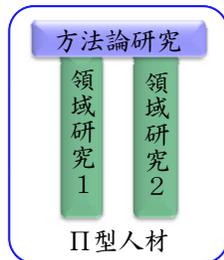
情報統合の切り札

- 逐次フィルタリング (Sequential filtering)
- 能動的モデリング (Active modeling)
- 個別化 (Personalization)
- 情報統合・データ同化 (Data assimilation)

鍵はBayes modeling

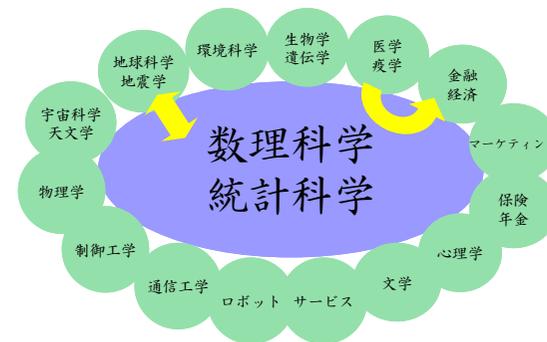
現在求められる研究者像

- T型人材, II型人材
 - 方法論 (横型の知識)
 - モデリング, 数理解析,
 - データ解析, 情報処理
 - 領域科学 (縦型の知識)
- 異分野交流のハブ
- 研究コーディネータ

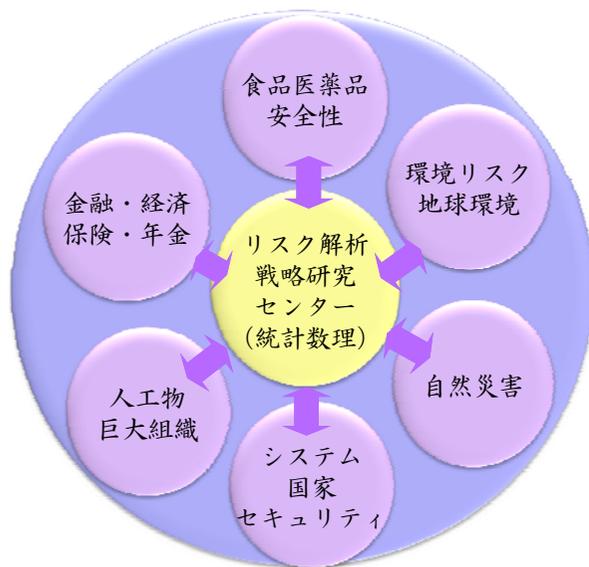


数理学と他分野との連携・融合

1. 科学的方法論を要とした連携
問題解決・問題発見のスパイラル
2. 科学研究のハブ → 知識移転



リスク研究のNOE構築



5つのNOE
(Network of Excellence)

- リスク科学
- 次世代シミュレーション
- 調査科学
- サービス科学
- 統計的機械学習

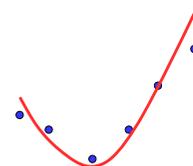
トレンド: パラメトリックモデル

歴史的例題:

$$y_n = f(n) + \varepsilon_n, \quad n = 1, \dots, N$$

y_n Observation
 $f(n)$ Parametric model
 ε_n Noise

例:
 多項式モデル
 三角関数
 ...



真の構造 + 観測ノイズ