IAHS 90th ANNIVERSARY – PUB SYMPOSIUM 2012
Theme 2: Conceptualization of process heterogeneity
Delft, The Netherlands, 23-25 October 2012

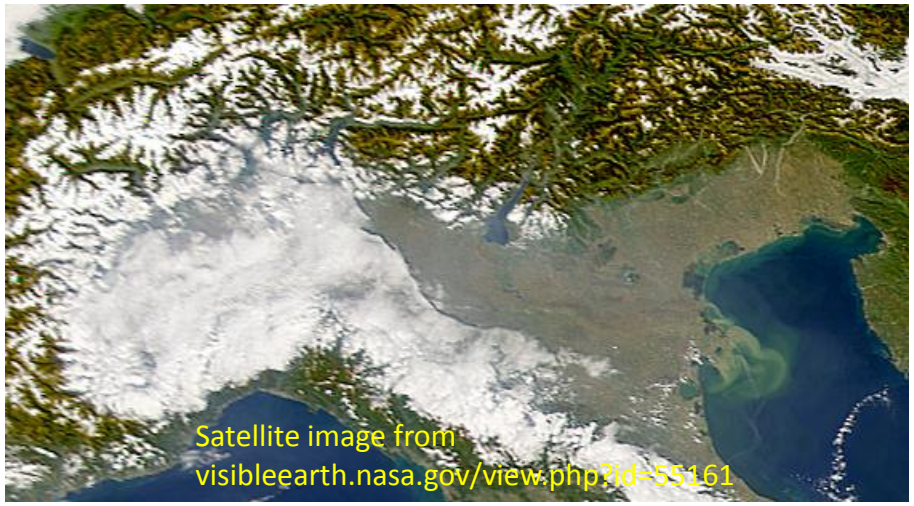
From deterministic heterogeneity to stochastic homogeneity

Demetris Koutsoyiannis
Department of Water Resources and Environmental Engineering School of Civil Engineering
National Technical University of Athens, Greece
(dk@itia.ntua.gr; itia.ntua.gr/dk/)

Presentation available online: itia.ntua.gr/1289/

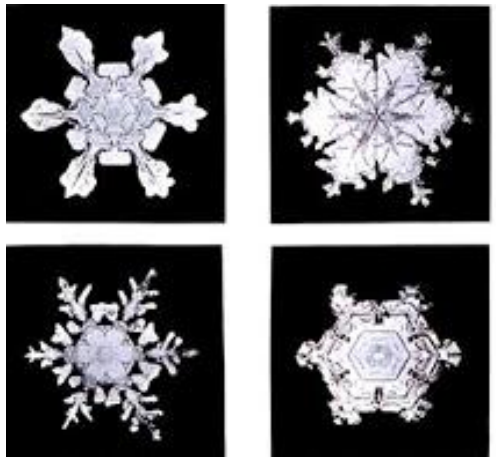
Visualization of diversity and heterogeneity

Megascopic: Po river basin



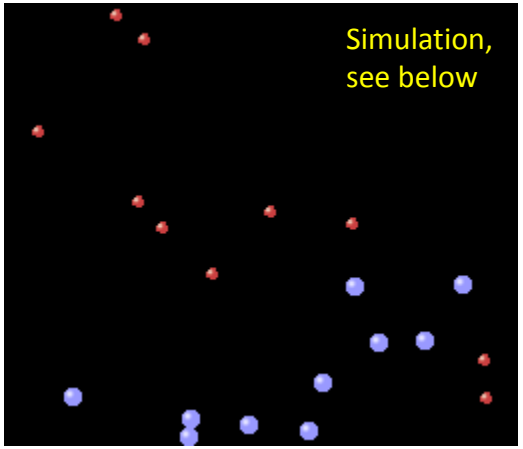
Macroscopic: Gully erosion

Mesoscopic: Snowflakes and soil structure



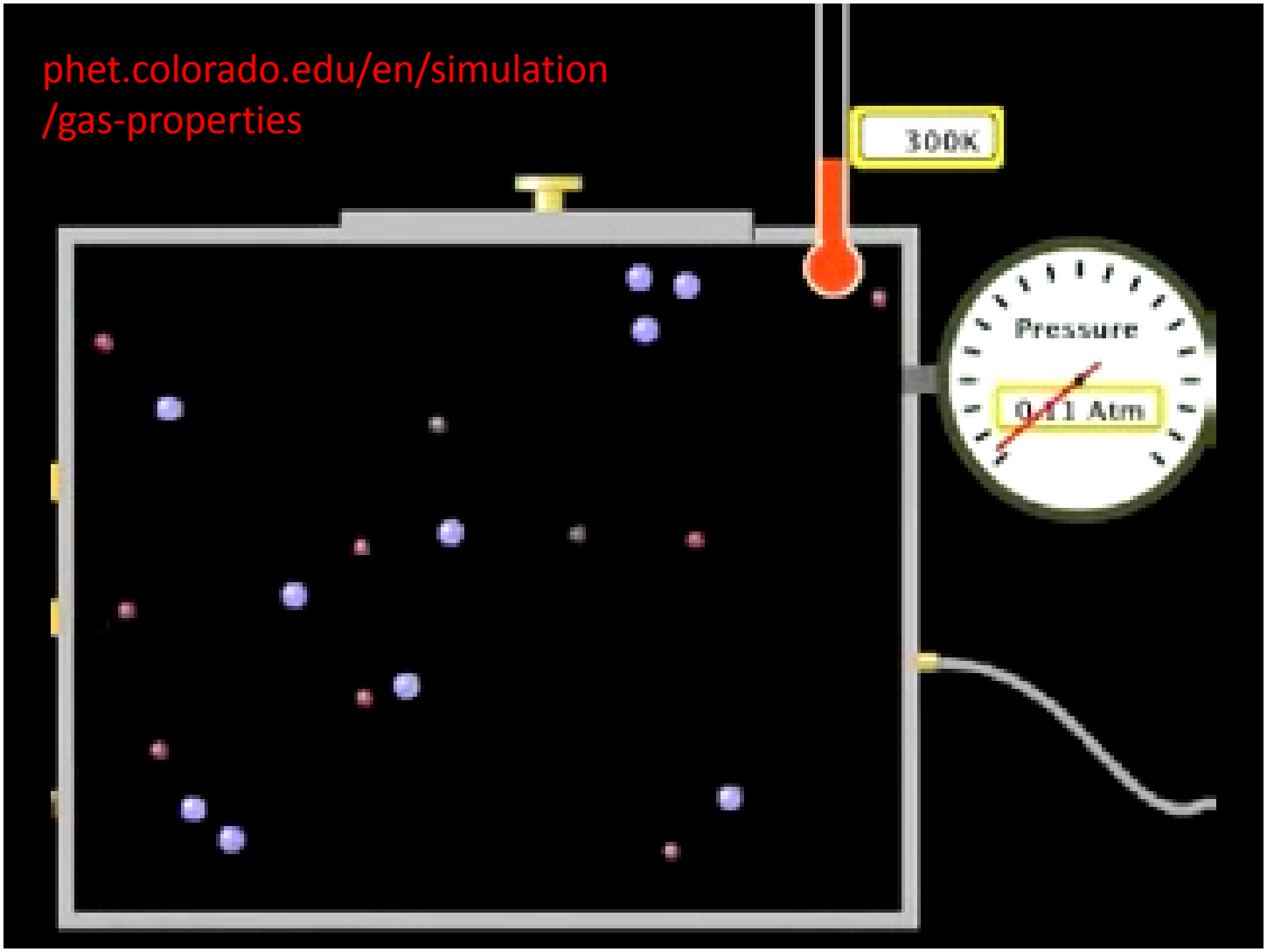
cher-homespun.blogspot.gr/2010/12/no-snowflake-in-avalanche-ever-feels.html

Microscopic: Molecules of gases in motion

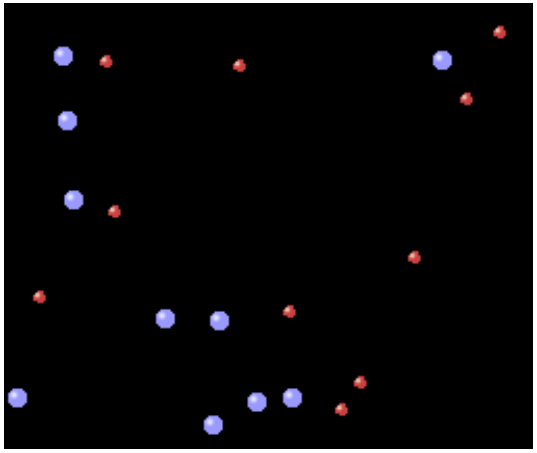


Microscopic world: Gas molecules in motion

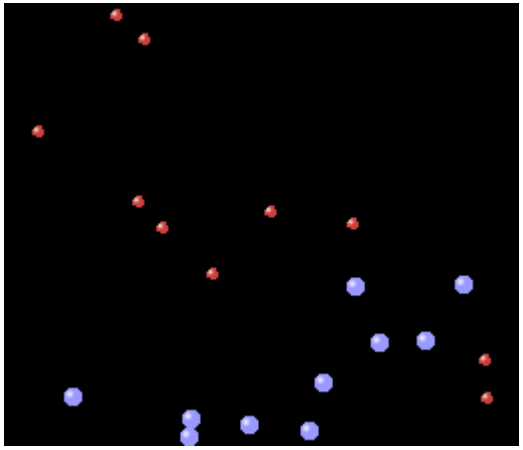
Computer simulation of 20 molecules in motion (10 heavy, 10 light).



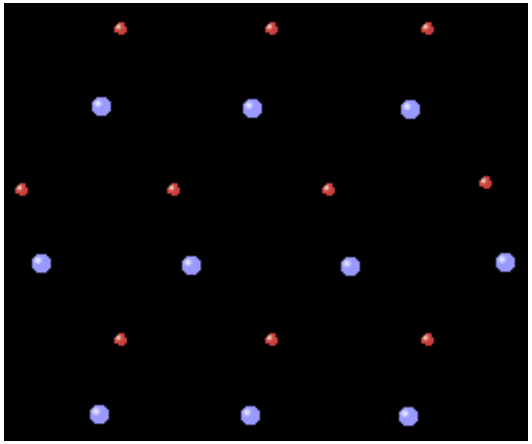
Contemplating on snapshots of gas motion simulation



Time: t_1
Both heavy and light molecules are irregularly distributed and tend to cover all available space

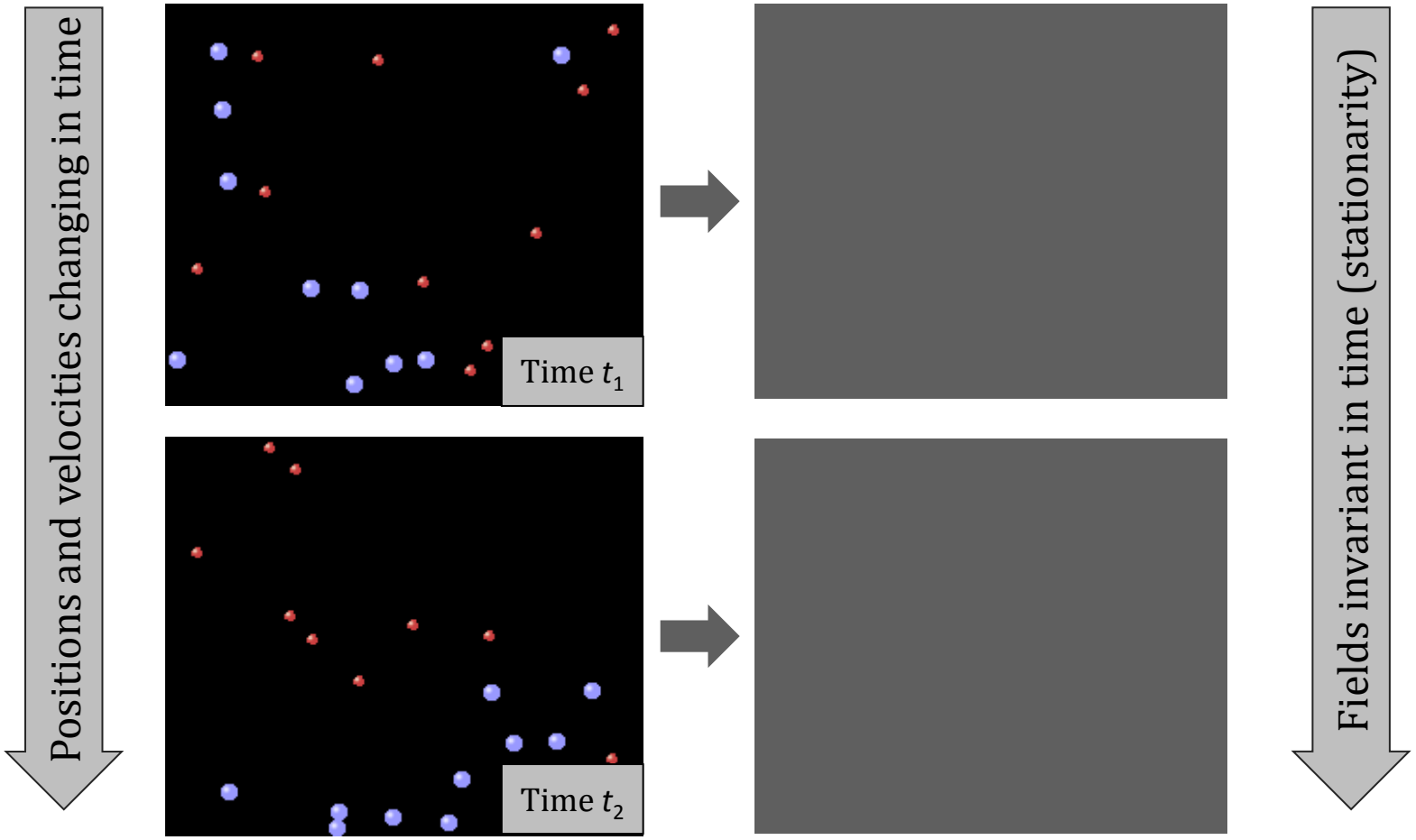


Time: t_2
All heavy molecules lie in the lower part while the upper-right part is free even of light molecules
(but such patterns would hardly appear for larger N)



Time: ?
Regular/uniform arrangement of molecules (this one is fake—not constructed by simulation; perhaps the age of universe is not enough for this to appear)

Deterministic inhomogeneity and nonstationarity vs. statistical homogeneity and stationarity



Positions and velocities irregularly and non-uniformly distributed

Abstract random fields uniformly distributed in space (homogeneity)

Probability theory and the notion of random variable

- By its definition, probability constitutes a mapping of sets into numbers: for a set (event) A, the probability $P(A)$ is a number in the interval [0, 1].
- By its definition, a random variable is not a single number but some function of elementary events; despite its name, a random variable does not necessarily describe anything random.
- Analogy: the equation $x^2(x-1)^3 = 0$ has a double root $x = 0$ and a triple root $x = 1$; the notion "root of the equation" has, at the same time, the values $x = 0$ and $x = 1$, the first twice and the second thrice.
- In the above analogy, the "root of the equation" can be represented as a random variable $\underline{x}$ taking on both values 0 and 1 with $P\{\underline{x} = 0\} = 2/5$ and $P\{\underline{x} = 1\} = 3/5$.
- Note: A random variable needs a special notation to distinguish it from a regular variable x; the best notation devised is the so-called Dutch convention (Hemelrijk, 1966), according to which random variables are underlined, i.e. $\underline{x}$.

Basic quantities associated to a random variable

- Continuing the same logic we can proceed from single-valued variables x to infinite-times-valued variables, or random variables $\underline{x}$, discrete or continuous, which are always associated with a **probability distribution function**, a function of a regular (real or vector) variable $x \in X$ defined as:

  $$F(x) := P\{\underline{x} \le x\}$$

- **Probability mass function** of a discrete random variable (DRV) $\underline{x}$:

  $$p(x) := P\{\underline{x} = x\}$$

- **Probability density function** of a continuous random variable (CRV) $\underline{x}$:

  $$f(x) := P\{x \le \underline{x} \le x + \mathrm{d}x\} \,/\, \mathrm{d}x$$

- **Expectation** of any deterministic function $g(\underline{x})$ of the random variable $\underline{x}$:

  $$\text{(DRV) } \mathrm{E}[g(\underline{x})] := \sum_X g(x)\, p(x); \quad \text{(CRV) } \mathrm{E}[g(\underline{x})] := \int_X g(x)\, f(x)\, \mathrm{d}x$$

- **Entropy**:

  $$\text{(DRV) } \varphi[\underline{x}] = \mathrm{E}[-\ln p(\underline{x})]; \quad \text{(CRV) } \varphi[\underline{x}] = \mathrm{E}[-\ln(f(\underline{x})/h(\underline{x}))]$$

  where $h(x)$ is a reference density (usually the Lebesgue density, i.e. a constant $h = 1$ with dimensions identical with those of $f(\underline{x})$)

Homogeneous description of gas in motion: quantities invariant in space and time

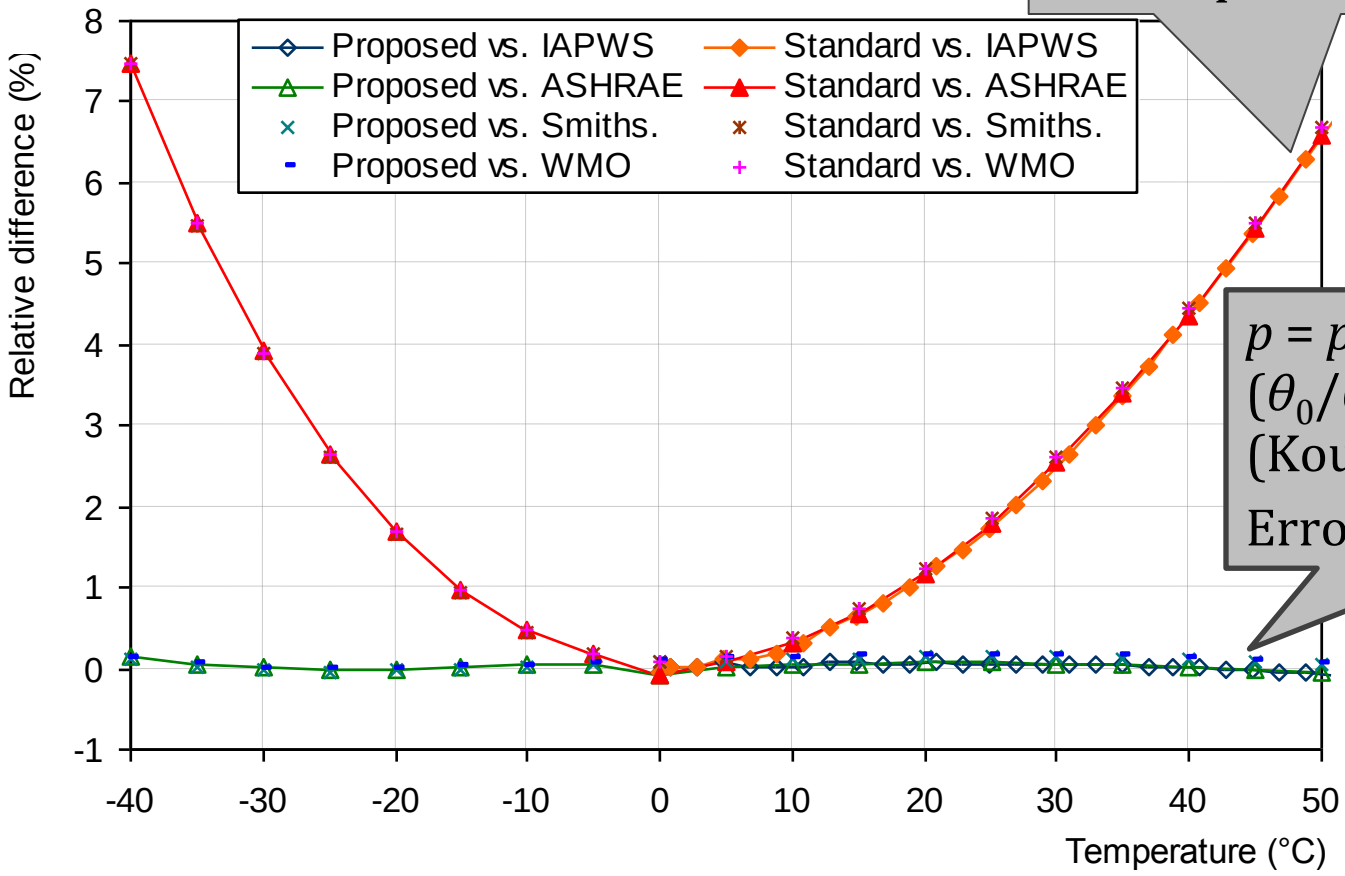
- **Volume per particle** $v := \mathrm{d}V/\mathrm{E}[\mathrm{d}\underline{N}]$ **and density** $\rho := \mathrm{E}[\mathrm{d}\underline{m}]/\mathrm{d}V = m_0\,\mathrm{E}[\mathrm{d}\underline{N}]/\mathrm{d}V = m_0/v$ where $\mathrm{d}\underline{m}$ is the mass contained in in volume $\mathrm{d}V$, m_0 the mass of a particle (here assumed to be of one kind only), $\mathrm{d}\underline{N}$ the number of particles in $\mathrm{d}V$.
- **Entropy** (per particle): $\varphi = \mathrm{E}[-\ln(f(\underline{\boldsymbol{z}})/h)]$
  where $\underline{\boldsymbol{z}}$ is the vector of random variables describing the random position and velocity of the particle; $f(\boldsymbol{z})$ is the probability density function of the random vector $\underline{\boldsymbol{z}}$; and h is a standardizing constant with units identical to those of f.
- **Temperature**: $1/\theta := \partial\varphi/\partial\varepsilon$
  where ε is the internal energy of the particle.
- **Pressure**: $p := \mathrm{E}[\mathrm{d}\underline{F}/\mathrm{d}s] = 2\,\mathrm{E}[\mathrm{d}\underline{q}]/\mathrm{d}t\,\mathrm{d}s$
  where $\mathrm{d}\underline{F}$ represents force perpendicular to the areal element $\mathrm{d}s$ due to collision of molecules and $\mathrm{d}\underline{q}$ is the momentum of molecules colliding at the areal element $\mathrm{d}s$ during time $\mathrm{d}t$
- Note: Natural units are used, i.e. entropy is dimensionless and temperature is in energy units (Koutsoyiannis 2011).

Some deterministic laws resulting from a stochastic framework

- **Basic principle**: Entropy is as high as possible (principle of maximum entropy).
- **Energy distribution**: the total kinetic energy (thermal energy) is equally distributed among different molecules (whether of the same kind or different kinds).
- **Equipartition**: the kinetic energy is equally distributed among the different degrees of freedom of a molecule.
- **Law of ideal gases**: $p\,v = \theta$
- **Law of adiabatic change in gases**: ${\theta_A}^{1+\beta/2}/p_A = {\theta_B}^{1+\beta/2}/p_B$ where β is the number of degrees of freedom of the motion of a particle.
- **Law of phase change** (relationship between the equilibrium partial pressure p of and temperature θ—the Clausius-Clapeyron equation): $p = p_0\, e^{\xi/\theta_0\,(1-\theta_0/\theta)}\,(\theta_0/\theta)^{(\beta_L/2-\beta_G/2-1)}$ where β_L and β_G are the degrees of freedom in the liquid and gaseous phase, respectively, ξ is the amount of energy per molecule required to break the bonds between molecules of the liquid phase and $(\theta_0, p_0 := p(\theta_0))$ is an anchor point (Koutsoyiannis, 2012).

Phase change of water: Impressive agreement of observations with theoretical results based on entropy maximization

Standard version of textbooks:
$p = p_0 \, e^{\xi/\theta_0 \, (1 - \theta_0/\theta)}$
Error up to 7%



$p = p_0 \, e^{\xi/\theta_0 \, (1 - \theta_0/\theta)} \times (\theta_0/\theta)^{(\beta_L/2 - \beta_G/2 - 1)}$
(Koutsoyiannis, 2012)
Error ≈ 0

A note about expected values and accuracy

- Strictly speaking, expected values are abstract quantities referring to random variables and corresponding to ensembles or realizations.
- The number of particles in macroscopic quantities of gases is tremendous, $N \approx 10^{27}$ per kmol; in addition, they are identical to each other (or of a few kinds) and the random variables describing their states can be regarded independent to each other.
- According to the law of large numbers, the expected values can be estimated with negligible error from typical statistical estimators for a single realization of the N particles.
- This make the expected values observable macroscopic quantities.
- However, this does not apply to small N, particularly when there is dependence among the different random variables. In such a case, expected values remain abstract and non-observable quantities.
- For small N, we rather measure realizations rather than expected values and we cannot anticipate high accuracy of estimates.

Mesoscopic world: flow in porous medium

Flow paths through a saturated porous medium (soil)

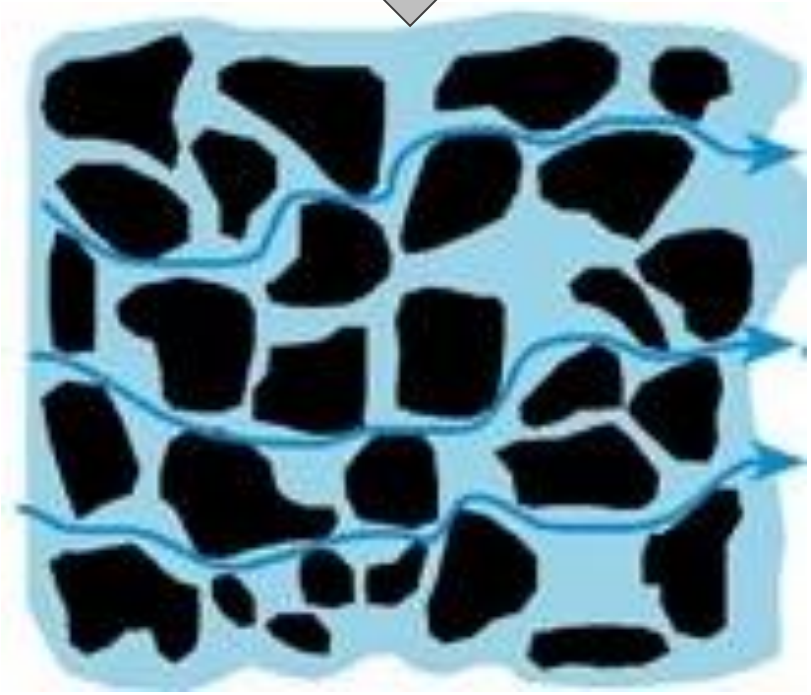
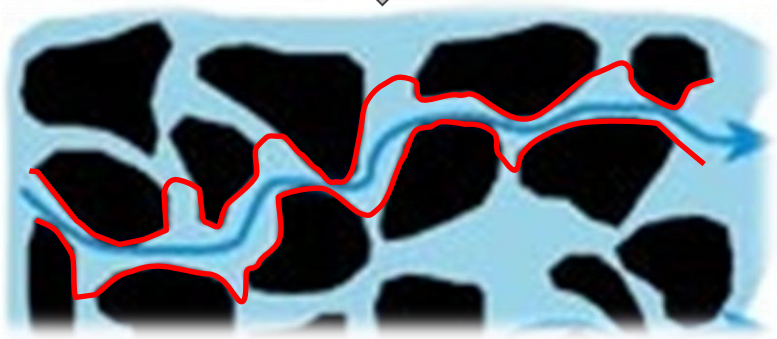
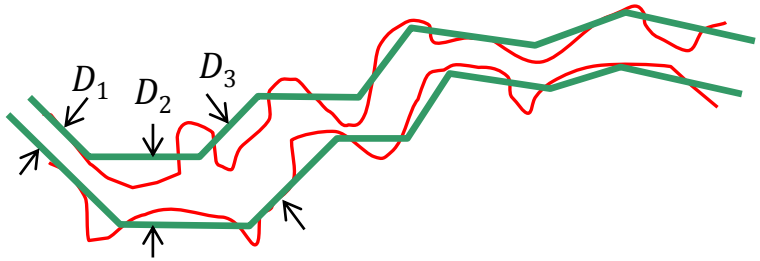


Image adapted from www.terragis.bees.unsw.edu.au/terraGIS_soil/sp_water-water_flow.html

Delineation of an irregular pipe of varying cross section



Approximation of the irregular pipe by cylindrical pipe segments

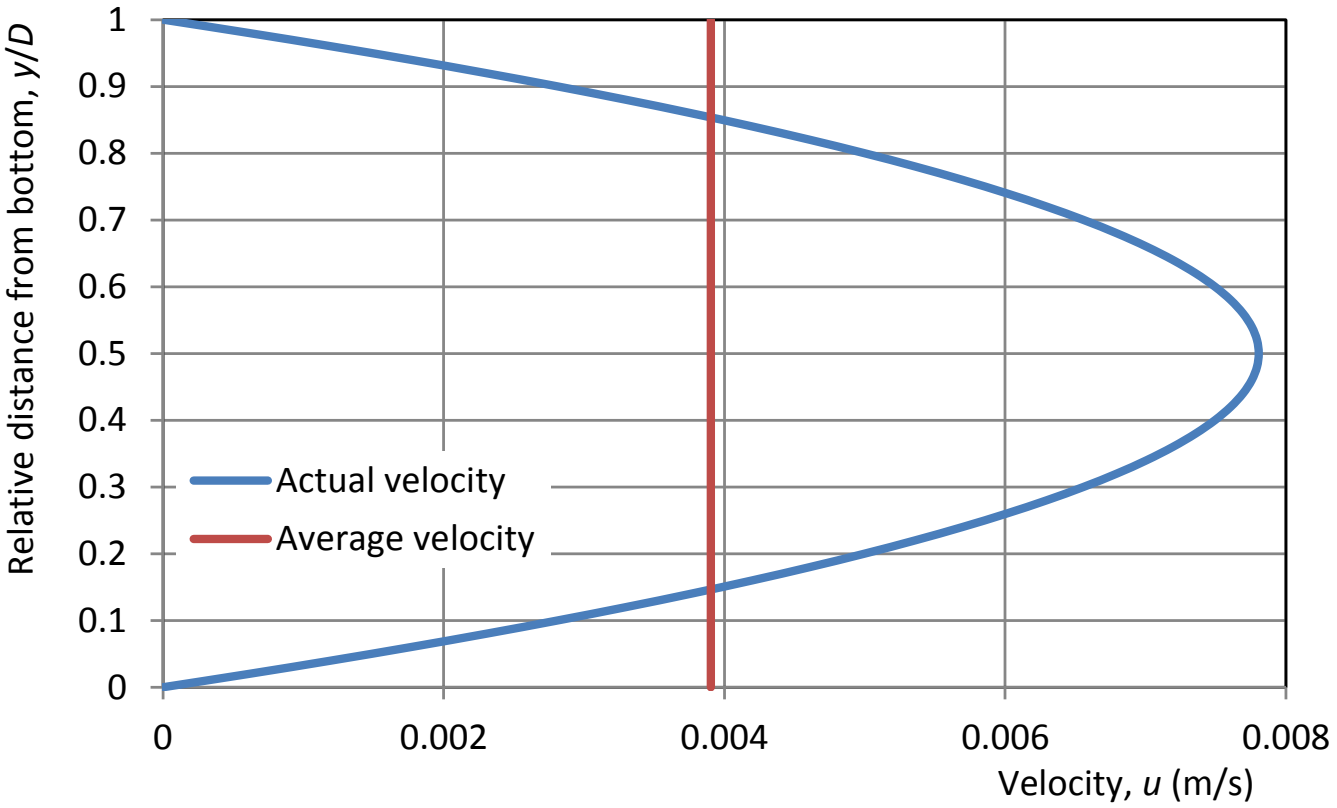


A detailed toy model of an irregular pipe

- A pipe with length L = 10 m and variable cross section is considered; this length is divided into segments of equal length δL = 10 mm, each one assumed to be cylindrical pipe with constant area A.
- In each segment, the quantity $\sqrt{A}$ is assumed to be a realization of a log-normally distributed variable $\underline{w}$ with mean μ = 1 mm and standard deviation σ = 0.5 mm; variables $\underline{w}_i$ in consecutive segments i are linearly correlated with a Hurst-Kolmogorov structure with Hurst coefficient H = 0.75.
- The discharge is Q = 10 mm^3/s, which for a characteristic area of 1 mm^2 gives a characteristic mean velocity of 10 mm/s.
- Under these conditions, a characteristic Reynolds number is Re = 10, which indicates that the flow is laminar.

The velocity profile in laminar flow in a cylindrical pipe

- Laminar flow has a velocity profile with steep gradients (strongly non-uniform).
- Yet it is convenient to use in applications the average velocity V.



Model of head losses in the irregular pipe

- Major losses in laminar flow (theoretically derived equation):

$$h_{\mathrm{f}} = \frac{8\pi\nu L}{gA} V = \frac{8\pi\nu L}{gA^2} Q$$

where π = 3.14…, ν is the kinematic viscosity, g is the gravity acceleration, L and A are the length and cross-section area of the pipe, respectively, and V is the flow velocity, spatially averaged over the cross-section.

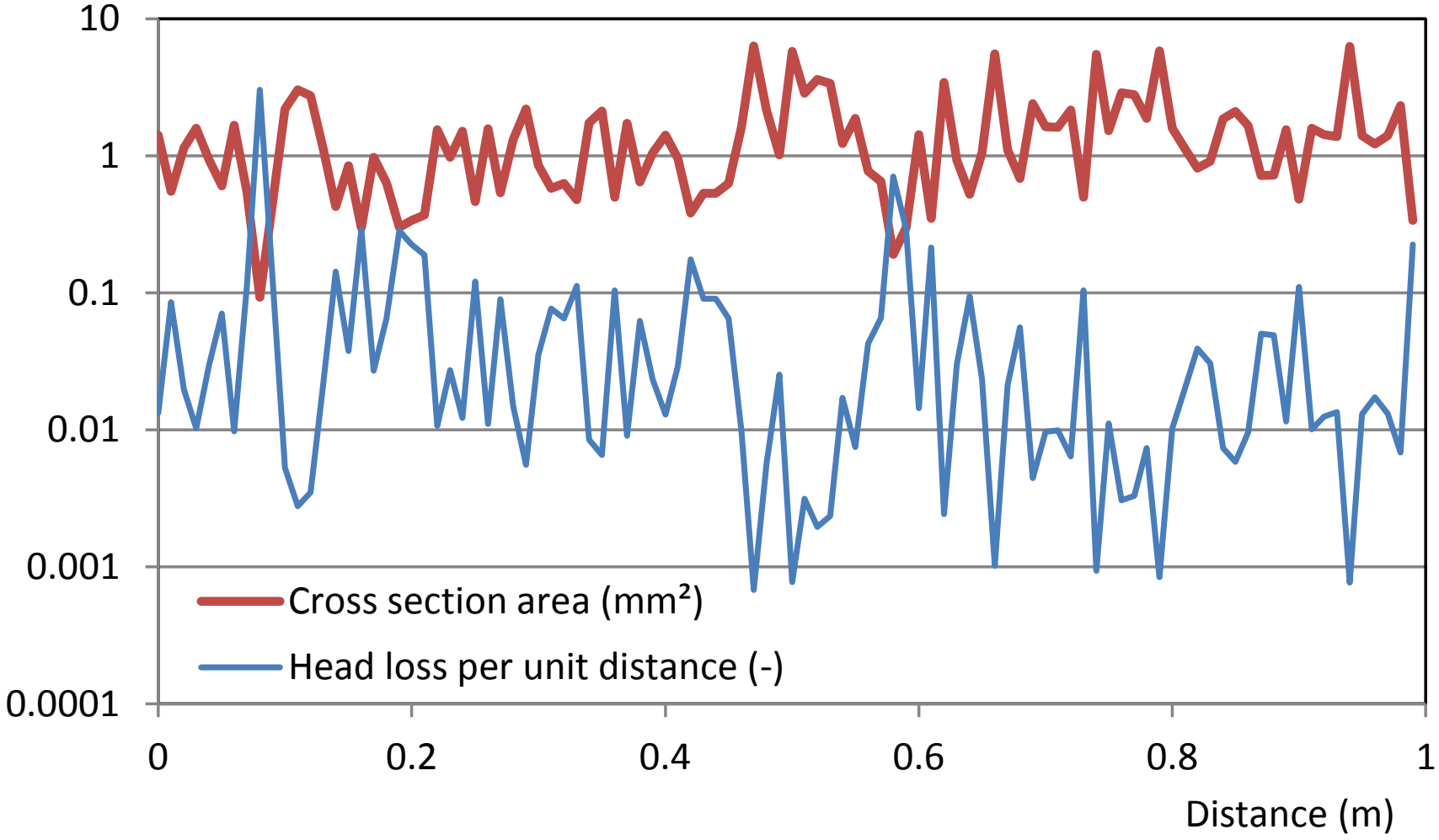
- Minor losses:

$$h_{\mathrm{L}} = \min\left(\frac{1}{2} + \frac{V_{\mathrm{u}}}{V_{\mathrm{d}}}, 1\right) \frac{(V_{\mathrm{u}} - V_{\mathrm{d}})^2}{2g}$$

where V_{u} and V_{d} is the velocity upstream and downstream, respectively.

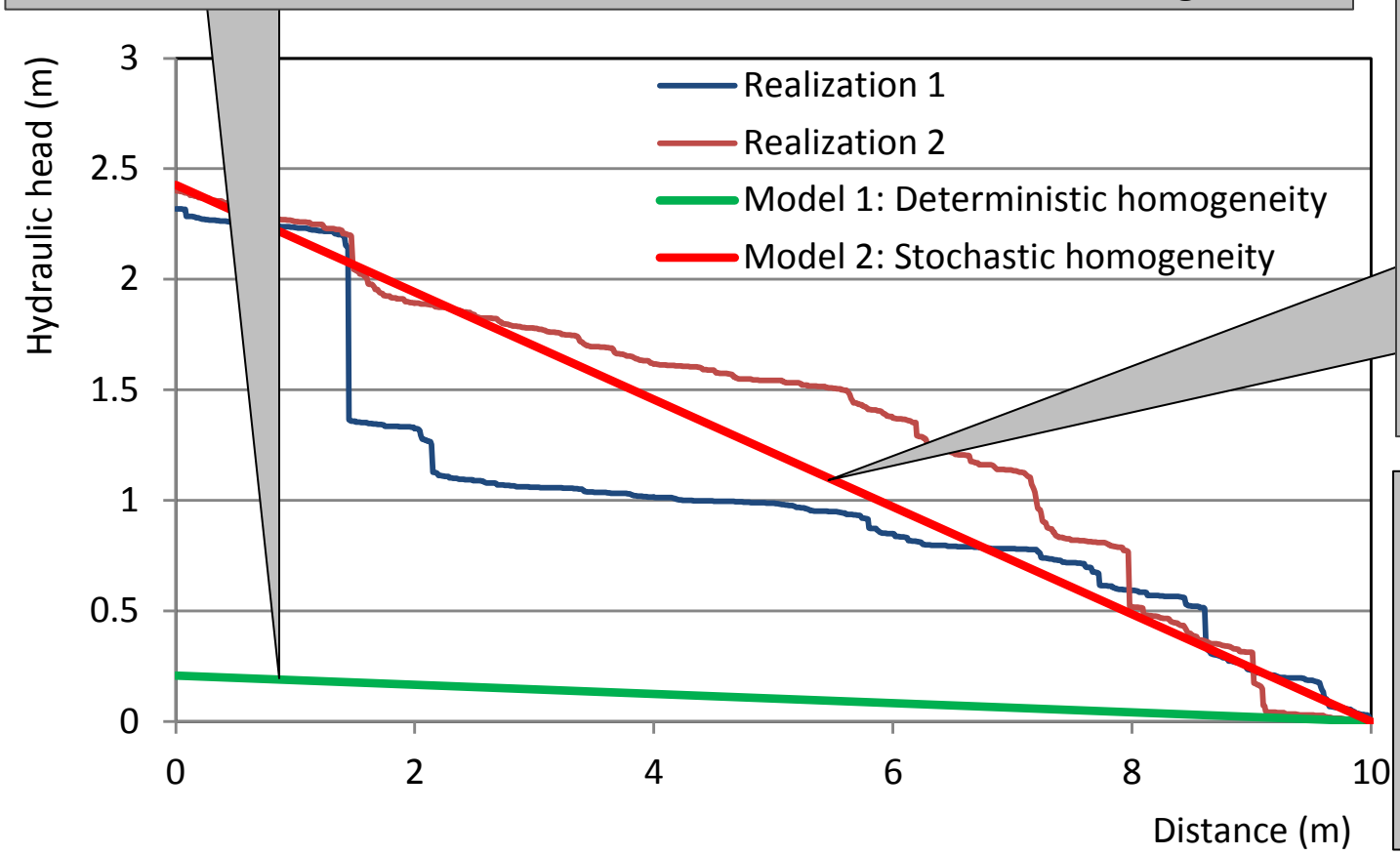
Note: this simple equation constitutes a satisfactory approximation of graphical or tabulated results commonly referred to in textbooks

One realization of (part of) the pipe and the resulting variation of head loss



Two realizations of the heterogeneous toy model and two homogeneous models

A daring determinist would "homogenize" the entire pipe using as A the average area of the pipe, so that $h_f = \frac{8\pi\nu L}{gA^2} Q$.

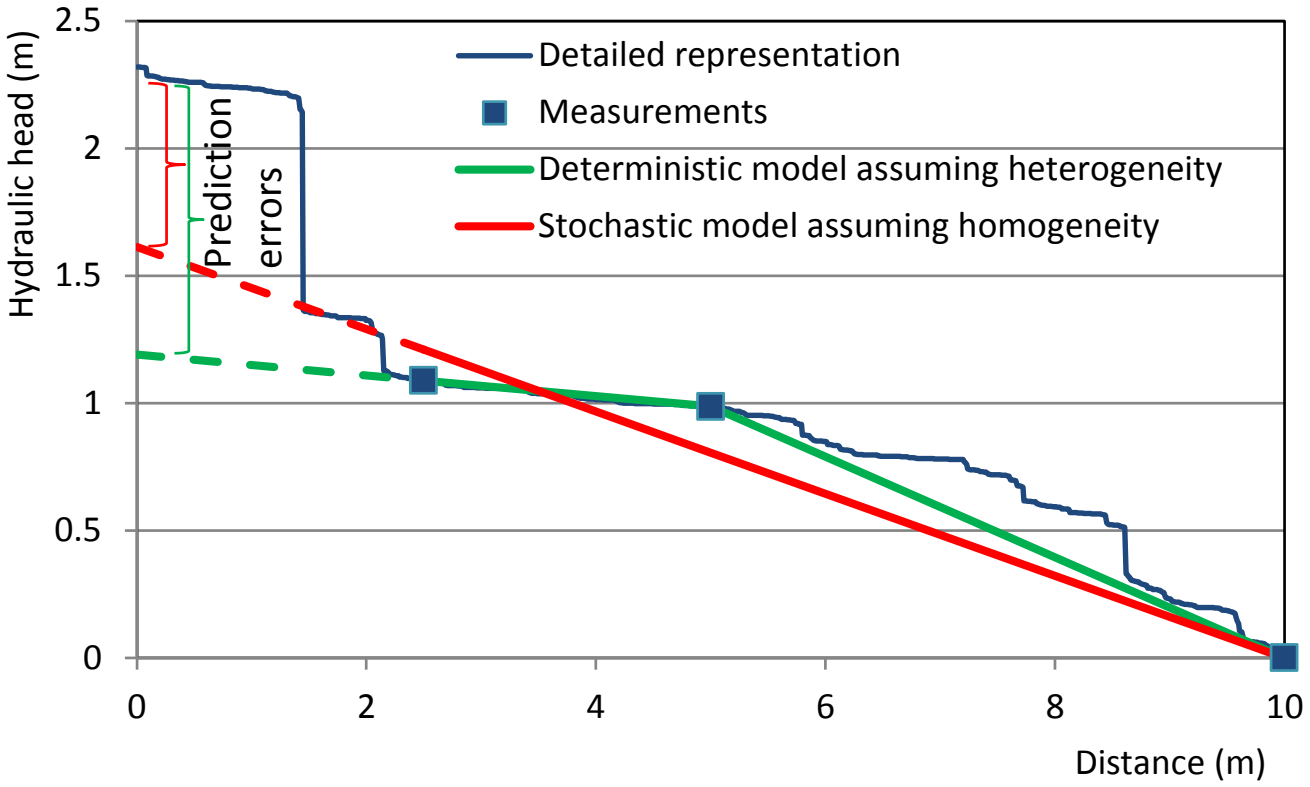


A stochastics-zealot will always dare homogenize, obtaining the correct equation $\mathrm{E}[\underline{h_f}] = \frac{8\pi\nu L}{g} \mathrm{E}[\underline{A}^{-2}]\, Q$

Conclusion: Stochastic homogeneity works perfect; details are useless.

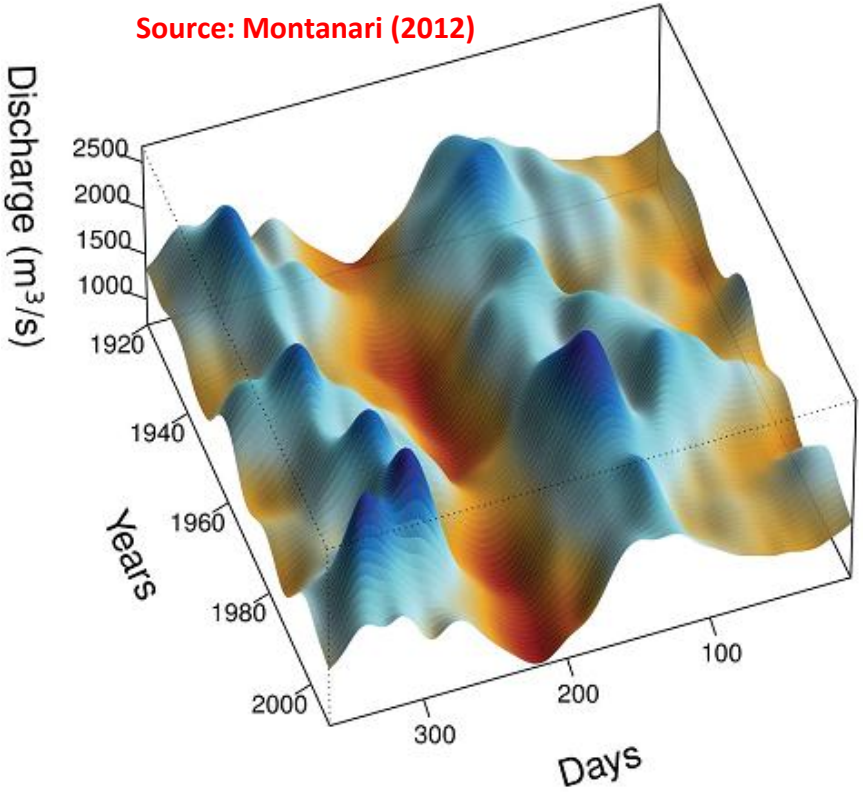
Predictions based on observations

- Realization 1 is assumed to be the true representation.
- Three measurements are assumed in points shown in figure.
- A prediction at point $x = 0$ is sought.
- In a deterministic approach the three measurements allow the assumption of different properties for the left and the right halves (heterogeneity)
- Stochastic homogeneity gives a better prediction than deterministic inhomogeneity.
- In the stochastic approach the error is expected and quantifiable.



Macro- and mega-scales

Source: Montanari (2012)



Po river

Q1: Would it ever be possible to model all details in the river basin and would this have any value for prediction?

Depiction of the Po river discharge time series

Q2: Could any model predict these patterns if the basin were ungauged?

All errors and uncertainties taken into account: stochastic process-based hydrology modelling

WATER RESOURCES RESEARCH, VOL. 48, W09555, doi:10.1029/2011WR011412, 2012

A blueprint for process-based modeling of uncertain hydrological systems

Alberto Montanari[1] and Demetris Koutsoyiannis[2]

Received 16 September 2011; revised 19 August 2012; accepted 20 August 2012; published 29 September 2012.

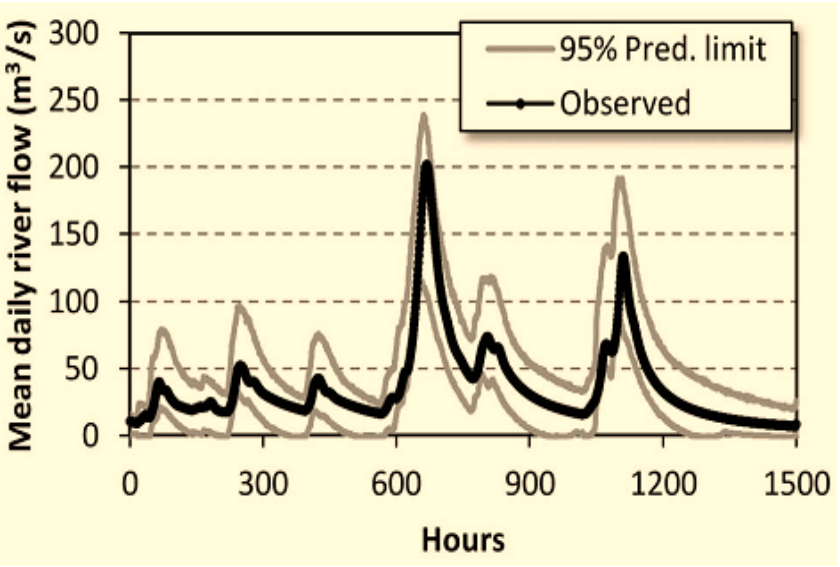
[31] To summarize the whole set of analytical derivations expressed by equations (1)–(6) we may see that we passed from the deterministic formulation of the hydrological model expressed by equation (1), i.e., (to replicate it for clarity),

$$Q = S(\boldsymbol{\Theta}, \mathbf{X}) \tag{7}$$

to the stochastic formulation expressed by

$$f_Q(Q) = \int_{\boldsymbol{\Theta}}\int_{\mathbf{X}} f_e(Q - S(\boldsymbol{\Theta}, \mathbf{X})|\boldsymbol{\Theta}, \mathbf{X})\, f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}) f_{\mathbf{X}}(\mathbf{X})\, d\boldsymbol{\Theta} d\mathbf{X} \tag{8}$$

with the following meaning of the symbols:
$f_Q(Q)$, probability density function of the true value of



Conclusions and discussion

- The real world is characterized by diversity—from micro- to mega-scales.
- Models of the real world can be homogeneous and need not describe the diversity of the details (heterogeneity).
- Probability theory and stochastics provide the tools to tackle the non-conformity between the homogeneous models with the heterogeneous reality.
- Only trivially simple systems (e.g. two bodies, harmonic oscillators) can be studied purely deterministically.
- Only in microscopic scales is deduction possible and requires physical laws both deterministic (Newton's) and stochastic (maximum entropy).
- In meso-, macro- and mega-scales inference relies on induction, based on data, and its theory is none other than stochastics.
- In models in these scales, error is inescapable and uncertainty is impossible to exterminate.
- However it is possible to quantify uncertainty through probability distribution functions of predictors and predictands.
- The only way to make **reliable** predictions in ungauged basins is to convert them to gauged basins (sorry about that!).

References

Hemelrijk, J., Underlining random variables, *Statistica Neerlandica*, 20 (1), 1966.

Koutsoyiannis, D., *Lecture Notes on Hydrometeorology: A probability-based introduction to atmospheric thermodynamics*, 45 pages, Department of Water Resources and Environmental Engineering – National Technical University of Athens, Athens, 2011 (itia.ntua.gr/1167/).

Koutsoyiannis, D., Clausius-Clapeyron equation and saturation vapour pressure: simple theory reconciled with practice, *European Journal of Physics*, 33 (2), 295–305, 2012.

Montanari, A., Hydrology of the Po River: looking for changing patterns in river discharge, *Hydrol. Earth Syst. Sci.,* 16, 3739-3747, doi: 10.5194/hess-16-3739-2012, 2012.

Montanari, A., and D. Koutsoyiannis, A blueprint for process-based modeling of uncertain hydrological systems, *Water Resources Research*, 48, W09555, doi:10.1029/2011WR011412, 2012.