Tokunaga self-similarity arises naturally from time invariance

Yevgeniy Kovchegov
Oregon State University

joint work with
Ilya Zaliapin of U. Nevada Reno
Horton–Strahler ordering.

The **Horton–Strahler ordering** of the vertices of a finite rooted labeled binary tree is performed in a hierarchical fashion, from leaves to the root:

(i) each leaf has order $r(\text{leaf}) = 1$;

(ii) when both children, $c_1, c_2$, of a parent vertex $p$ have orders $i$ and $j$, the vertex $p$ is assigned order

$$r = \left\lfloor \log_2(2^i + 2^j) \right\rfloor = \begin{cases} \max\{i, j\} & \text{if } i \neq j \\ i + 1 & \text{if } i = j \end{cases}$$
Horton-Strahler ordering and Tokunaga indexing.

Example: (a) Horton-Strahler ordering  
(b) Tokunaga indexing.

Two order-2 branches are depicted by heavy lines in both panels. The Horton-Strahler orders refer, interchangeably, to the tree nodes or to their parent links. The Tokunaga indices refer to entire branches, and not to individual vertices.
Horton pruning of a tree mod series reduction

The order of the tree is $k(T) = 3$ with $N_1 = 10$, $N_2 = 3$, $N_3 = 1$, and $N_{1,2} = 3$, $N_{1,3} = 1$, $N_{2,3} = 1$. 
Tree self-similarity

$N_j$ – the number of branches of order $j$

$N_{ij}$, $i < j$ – the number of side branches of order $\{ij\}$, i.e. instances when an order-$i$ branch merges with an order-$j$ branch in a finite tree $T$.

The average number of branches of order $i$ in a single branch of order $j$ can be traced with

$$T_{ij} = \frac{\mathbb{E}[N_{ij}]}{\mathbb{E}[N_j]}$$

Tree self-similarity: $T_{ij} = T_{j-i}$ for a sequence $\{T_k\}_{k \geq 1}$.

Tokunaga self-similarity: $T_k = ac^{k-1}$, $k \geq 1$, $a, c > 0$. 
**Geometric branching process**

\[ X \overset{d}{=} \text{Geom}(r) \text{ if } \text{Prob}(X = k) = r(1 - r)^k, \ k = 0, 1, \ldots \]

Given a non-negative sequence \( \{T_k\}_{k \geq 1} \). Let

\[ S_K := 1 + T_1 + \cdots + T_K \]

for \( K \geq 0 \) by assuming \( T_0 = 0 \).

**Geometric branching process:**

- Markovian, where the numbers of side branches are independent.
- The number \( m_{j,i} \) of side branches of order \( i < j \) in a branch of order \( j \) is distributed as

\[ m_{j,i} \overset{d}{=} \text{Geom}\left(\left[1 + T_{j-i}\right]^{-1}\right). \]

Thus, \( E[m_{j,i}] = T_{j-i} \).
- Branches of order 1 have no side branches.
Geometric branching process

- Markovian, where the numbers of side branches are independent.

- The number $m_{j,i}$ of side branches of order $i < j$ in a branch of order $j$ is distributed as

$$m_{j,i} \overset{d}{=} \text{Geom} \left( [1 + T_{j-i}]^{-1} \right).$$

Thus, $\mathbb{E}[m_{j,i}] = T_{j-i}$.

- Branches of order 1 have no side branches.

The independence of branches implies $\mathbb{E}[N_{ij}] = \mathbb{E}[N_j] \mathbb{E}[m_{j,i}]$ and hence

$$T_{ij} = \frac{\mathbb{E}[N_{ij}]}{\mathbb{E}[N_j]} = \frac{\mathbb{E}[N_j] \mathbb{E}[m_{j,i}]}{\mathbb{E}[N_j]} = T_{j-i}.$$
Geometric branching process: formally

(i) The process starts with \( \text{ord}(T) - 1 \overset{d}{=} \text{Geom}(p) \).

(ii) At every time instant \( s > 0 \), each population member of order \( K \) terminates with probability \( S_{K-1}^{-1} \), independently of other members.

At termination, a member of order \( K > 1 \) produces two offspring of order \((K-1)\); and a member of order \( K = 1 \) leaves no offsprings.

(iii) A population member of order \( K \) survives (with probability \( 1 - S_{K-1}^{-1} \), and produces a single offspring (side branch) of order \( i \) (\( 1 \leq i < K \)) drawn from the distribution

\[
p_{K,i} = \frac{T_{K-i}}{T_1 + \cdots + T_{K-1}}.
\]
Geometric branching process

Properties:

• The geometric branching process with \( p = 1/2 \) and \( T_k = 2^{k-1} \) is the critical binary Galton-Watson tree.

• **Prune invariance:** Given an arbitrary sequence \( \{T_k \geq 0\}_{k \geq 1} \) and \( 0 < p < 1 \), the probability measure for the geometric branching process is invariant with respect to Horton pruning.

• **Easy to simulate.** Generation of geometric trees for arbitrary parameters \( (p, \{T_k\}) \) is easily implemented on a computer.

May facilitate the analysis in a range of simulation-heavy problems, from structure and transport on river networks to phylogenetic trees.
Time shift operator $S$ advances the process time by unity. It can be applied to individual trees and forests.

A consecutive applications of $d$ time shifts to a tree $T$ is equivalent to removing the vertices/edges at depth less than $d$ from the root.
Difference equation for the state vector

Let $x_i(s), i \geq 1$, denote the average number of vertices of order $i$ at time $s$ in a geometric branching process, and

$$x(s) = (x_1(s), x_2(s), \ldots)^T$$

be the state vector. Then

$$x(s + 1) - x(s) = GS^{-1}x(s),$$

where $x(0) = \pi := \sum_{K=1}^{\infty} p(1 - p)^{K-1}e_K,$

$$G := \begin{bmatrix} -1 & T_1 + 2 & T_2 & T_3 & \cdots \\ 0 & -1 & T_1 + 2 & T_2 & \cdots \\ 0 & 0 & -1 & T_1 + 2 & \cdots \\ 0 & 0 & 0 & -1 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}, \quad \text{and} \quad S = \text{diag}\{S_0, S_1, \ldots\}.$$
Time invariance

Geometric branching process is time invariant if and only if the state vector $x(s)$ is invariant with respect to a unit time shift $S$:

$$x(s) = x(0) \equiv \pi \quad \forall s \iff GS^{-1}\pi = 0.$$
Time invariance

Geometric branching process is time invariant if and only if the state vector $x(s)$ is invariant with respect to a unit time shift $S$:

$$x(s) = x(0) \equiv \pi \quad \forall s \iff GS^{-1}\pi = 0.$$ 

**Theorem (YK and I. Zaliapin, Chaos 2018).** A geometric branching process is time invariant if and only if

$$p = 1/2 \quad \text{and} \quad T_k = (c - 1)c^{k-1} \quad \text{for} \ c \geq 1.$$ 

We will call this a critical Tokunaga process, and the respective trees – critical Tokunaga trees.
**Time invariance**

Recall that $S_0 = 1$, and for $K \geq 1$,

$$S_K := 1 + T_1 + \cdots + T_K$$

First, we established the following.

**Lemma 1.** A geometric branching process is time invariant if and only if $p = 1/2$ and the sequence $\{T_k\}$ solves the following (nonlinear) system of equations:

$$\frac{S_0}{S_k} = \sum_{i=1}^{\infty} 2^{-i} \frac{S_i}{S_{k+i}}$$

for all $k \geq 1$. 
**Time invariance**

Let $a_k = S_k / S_{k+1} \leq 1$ for all $k \geq 0$. Then, for any $i \geq 0$ and any $k > 0$ we have $S_i / S_{k+i} = a_i a_{i+1} \ldots a_{i+k-1}$. The system

$$\frac{S_0}{S_k} = \sum_{i=1}^{\infty} 2^{-i} \frac{S_i}{S_{k+i}} \quad \text{for all } k \geq 1$$

rewrites in terms of $a_i$ as

$$\frac{1}{2} a_1 + \frac{1}{4} a_2 + \frac{1}{8} a_3 + \ldots = a_0,$$

$$\frac{1}{2} a_1 a_2 + \frac{1}{4} a_2 a_3 + \frac{1}{8} a_3 a_4 + \ldots = a_0 a_1,$$

$$\frac{1}{2} a_1 a_2 a_3 + \frac{1}{4} a_2 a_3 a_4 + \frac{1}{8} a_3 a_4 a_5 + \ldots = a_0 a_1 a_2,$$

and so on ...
Time invariance

Lemma 1. A geometric branching process is time invariant if and only if $p = 1/2$ and the sequence $\{T_k\}$ solves the following (nonlinear) system of equations:

$$\frac{S_0}{S_k} = \sum_{i=1}^{\infty} 2^{-i} \frac{S_i}{S_{k+i}} \quad \text{for all } k \geq 1.$$

Lemma 2. The system

$$\sum_{j=1}^{\infty} \frac{1}{2^j} \prod_{k=j}^{n+j-1} a_k = \prod_{k=0}^{n-1} a_k, \quad \text{for all } n \in \mathbb{N}$$

with $a_0 = \frac{1}{c}$ ($c > 0$) has a unique solution $a_0 = a_1 = a_2 = \ldots = 1/c$.

Once established, Lemma 1 and Lemma 2 imply our main result (Theorem).
• **Zaliapin and YK (CSF 2012):** Extreme values and level-set trees of time series via Horton pruning.

• **YK and Zaliapin (Fractals 2016):** Tree self-similarity with \( \lim_{j \to \infty} \sup T_j^{1/j} < \infty \) implies (strong) Horton law.


• **YK and I. Zaliapin (2017) arXiv:1608.05032**
  A novel multi-type branching processes is considered: the Hierarchical Branching Processes.

  Generalized dynamical pruning with applications in continuum ballistic annihilation.

• **YK and I. Zaliapin (Chaos 2018):** This talk.