Supplementary Information to: "Understanding the variability of daily travel-time expenditures using GPS trajectory data".

Riccardo Gallotti*
Institut de Physique Théorique, CEA-Saclay, Gif-sur-Yvette, France

Armando Bazzani and Sandro Rambaldi
Department of Physics and Astronomy, University of Bologna, INFN Bologna section

LIST OF ANALYZED CITIES

#	Name	Pop.	Area	v	$\langle t\rangle$	$\langle n\rangle$	$\langle T\rangle$	α	β	R^2
1	Roma	2786034	1285.3	26.87	0.253	5.87	1.52	0.69±0.01	1.00±0.01	0.99
2	Palermo	653522	158.9	23.00	0.170	6.80	1.39	0.60±0.01	0.92±0.01	0.99
3	Genova	606978	243.6	23.17	0.211	5.34	1.29	0.50±0.02	0.88±0.02	0.98
4	Napoli	957012	117.3	24.47	0.232	6.35	1.58	0.61±0.01	1.11±0.01	0.99
5	Bari	340355	116.0	31.03	0.172	6.77	1.52	0.58±0.02	1.04±0.02	0.99
6	Milano	1345890	181.8	24.35	0.230	5.52	1.44	0.48±0.01	1.02±0.01	0.99
7	Torino	905780	130.3	24.17	0.196	5.68	1.36	0.57±0.02	0.88±0.01	0.97
8	Bologna	383949	140.7	29.22	0.180	5.94	1.37	0.46±0.02	0.95±0.01	0.97
9	Firenze	373446	102.4	25.93	0.209	5.76	1.40	0.52±0.02	0.98±0.02	0.95
10	Reggio Calabria	186273	236.0	24.79	0.123	8.32	1.42	0.66±0.02	0.90±0.01	0.99
11	Perugia	169290	449.9	35.22	0.149	7.05	1.36	0.67±0.04	0.83±0.02	0.98
12	Catania	289971	180.9	23.95	0.170	7.12	1.48	0.61±0.02	0.99±0.01	0.99
13	Foggia	152181	507.8	24.20	0.149	7.61	1.45	0.55±0.03	0.98±0.02	0.99
14	Forlì	118968	228.2	33.30	0.186	6.61	1.27	0.45±0.02	0.87±0.02	0.99
15	Grosseto	82616	474.3	31.80	0.154	6.98	1.19	0.39±0.02	0.83±0.02	0.99
16	Latina	120526	277.8	31.32	0.156	7.19	1.40	0.53±0.03	0.95±0.02	0.99
17	Lecce	96274	238.4	31.46	0.155	8.33	1.46	0.65±0.04	0.94±0.02	0.98
18	Messina	240858	211.2	23.13	0.185	6.89	1.43	0.74±0.03	0.89±0.02	0.98
19	Modena	186289	183.2	32.09	0.158	6.65	1.30	0.53±0.03	0.83±0.02	0.97
20	Parma	189833	260.8	33.87	0.166	6.45	1.37	0.48±0.02	0.93±0.02	0.98
21	Ravenna	159404	244.2	38.18	0.161	6.58	1.31	0.52±0.03	0.86±0.02	0.98
22	Reggio Emilia	172317	231.6	32.00	0.158	6.48	1.29	0.57±0.03	0.81±0.02	0.97
23	Salerno	138284	59.0	28.16	0.194	6.55	1.49	0.57±0.03	1.03±0.02	0.99
24	Siracusa	123517	204.1	24.43	0.136	7.46	1.30	0.66±0.03	0.82±0.02	0.99

TABLE S1: The different columns are: # = id number in the figures, Name = municipality where the driver spent most of his parking time, Pop. = Municipality Population, Area = Municipality Area (km^2), v = speed (km/h), $\langle t\rangle$ = average travel-time per trip (h), $\langle n\rangle$ = average number of daily trips, $\langle T\rangle$ = average daily travel time expenditure (h), α = accessibility measure (h), β = travel-time budget (h), R^2 = coefficient of determination. Errors represent the 95% confidence intervals for the fit the parameters, estimated using a bootstrap method with 100 repetitions.

Duration Model

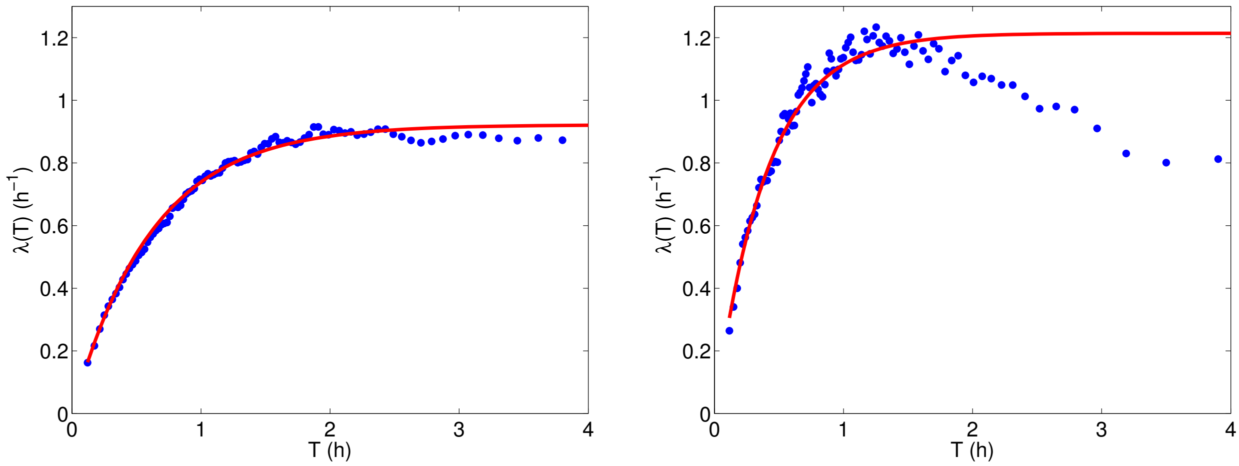

FIG. S1: **(Left) Hazard function in Naples.** The empirical hazard function $\lambda(T)$ (dots) is found to be exponentially converging to a constant value. **(Right) Hazard function in Grosseto.** We remark that the lack of data in the TTE distribution tail does not allow an accurate numerical evaluation of the hazard function. Therefore, in this case to estimate of α and β in some cities, we exclude the decreasing behaviour of the empirical hazard function. In both figures, the values of the parameters obtained by the exponential fit of $\lambda(T)$(solid line) define the solid line in Fig. 1.

The proposed TTE model is based on the Markov property for the evolution of the survival distribution $S(T)$ (see def. (2) in the main paper). Assuming a regular character for the decision stochastic process, we have

$$S(T + \Delta T) = \pi(T + \Delta T|T)S(T) + o(\Delta T)\,, \tag{S1}$$

where the likelihood $\pi(T + \Delta T|T)$ defines the conditioned probability for the representative individual of performing a daily mobility $T + \Delta T$ given the elapsed time T from the beginning of the daily mobility. Then, we introduce the *hazard function* $\lambda(T)$ (i.e. the probability of ending the mobility after a time the interval ΔT knowing that the user was still driving at time T) according to

$$\lambda(T) = \lim_{\Delta T \to 0} \frac{1 - \pi(T + \Delta T|T)}{\Delta T}\,, \tag{S2}$$

However the definition (S2) cannot be used if we consider a microscopic stochastic dynamics that mimics the individual decisions, so in the paper , we propose an alternative definition which link the individual behaviour with the average property of the population

$$\lambda(T) = \left\langle \frac{1 - \hat{\pi}(T + \Delta T/T)}{\Delta T} \right\rangle_{\Delta T} \tag{S3}$$

where $\hat{\pi}(T + \Delta T/T)$ is the conditional probability to observe a TTE $T + \Delta T$ of the individual dynamics and the average value is computed over the distribution of the possible increments ΔT in the considered population.
In the limit $\Delta T \to 0$ from (S1) we get the differential equation:

$$dS(T)/dT = -\lambda(T)S(T)\,. \tag{S4}$$

In a stationary situation, the hazard function $\lambda(T)$ is constant ($\lambda(T) = h_0$) and the differential equation (S4) leads to an exponential solution $S(T) = h_0 \exp(-h_0 T)$ which corresponds to an exponential probability density ($p(T) = -dS(T)/dT$) for the TTE. Under this point of view the exponential tail of the empirical distribution $p(T)$

can be associated to a constant probability of ending the daily mobility, independently from the elapsed daily travel-time T as expected for a Poisson process. However, the observed under-expressed short travel-times implies an increasing trend for the hazard function $\lambda(T)$: i.e. when the TTE is short, the probability to stop the daily mobility is lower than the asymptotic value. This observation suggests that it is unlikely for an individual to consider his daily mobility concluded after a very short cumulative travel-time T, because some daily duties have still to be accomplished. A more suitable shape of the function $\lambda(T)$ can be extrapolated from GPS data (see Supplementary Fig. S1). We perform an analytical interpolation of the empirical data by

$$\lambda(T) = -\frac{dS(T)/dT}{S(T)} = \beta^{-1}\left(1 - \exp(-T/\alpha)\right) , \tag{S5}$$

where the parameters β is the TTB, characteristic of a particular city, and α is the timescale of the short travel time expenditures under-expression. According to our point of view α could be interpreted as the average time necessary to satisfy the daily mobility demand using private cars. After a time α, the choice of going back home is only due to the limited TTB constraint, quantified by the time scale β. Given $\lambda(T)$, we integrate analytically the differential equation (S4) obtaining an analytical form for the survival function

$$S(T) = C_N \exp\left(-\alpha\beta^{-1}\exp(-T/\alpha) - T/\beta\right) , \tag{S6}$$

Imposing $S(0) = 1$ (thus neglecting null TTEs), the normalisation constant can be fixed at $C_N = \exp\left(\alpha\beta^{-1}\right)$. Consequently, the probability density function for the TTE distribution reads

$$p(T) = \beta^{-1}\exp\left(\alpha\beta^{-1}\right)\left(1 - \exp(-T/\alpha)\right)\exp\left(-\alpha\beta^{-1}\exp(-T/\alpha) - T/\beta\right) . \tag{S7}$$

Analytical solution for the time consumption model

Let us consider a driver who has carried out a daily mobility T and he has to decide if to perform or not a further trip whose duration is ΔT. According to our Statistical Mechanics point of view, the exponential decay of the empirical TTE distribution (see eq. (1) and Fig. 1 in the paper), suggests that the mobility time plays the role of the *energy*. As a consequence we expect that the probability to accept a TTE T is

$$P(T) = \exp\left(-\frac{T}{\bar\beta}\right) , \tag{S8}$$

where $\bar\beta$ is the expected value of TTE. In the model, to evaluate the probability of performing a new trip the drivers considers the possibility to accept the cost ΔT of the new trip using a threshold function

$$\theta_{x_{max}/\langle n\rangle}\left(\frac{\Delta T}{T}\right) = \begin{cases} 1 & \text{if } \frac{\Delta T}{T} < x_{max}/\langle n\rangle , \\ 0 & \text{otherwise} , \end{cases} \tag{S9}$$

where $\langle n\rangle$ is the average number of performed daily activities and x_{max} is a universal threshold (see Fig. 5 left). Then according to empirical observations (see. eq. (9) in the text) we introduce the conditional distribution for the time cost ΔT of the single trips

$$p(\Delta T/T) \approx \langle t\rangle^{-1}\exp(-\Delta T/\langle t\rangle)\theta_{x_{max}/\langle n\rangle}\left(\frac{\Delta T}{T}\right) \tag{1}$$

The presence of the threshold function $\theta_{x_{max}/\langle n\rangle}$ means that, as the TTE increases, the individual gets used to be driving and he has a propensity to accept longer trips (compatibly with his TTB and the number of activities he has to perform). Since the perceived cost of a new trip in the model is set $\propto \Delta T/T$, we correlate this choice with the existence of a *log-time* perception. To reproduce the macroscopic statistical laws of human mobility, the drivers are considered as *independent particles* and we average on the cost ΔT of the individual trips using the empirical distribution (1) (Supplementary Fig. S2 (a)). According to the TTB existence assumption, a rational driver evaluates the probability to perform the new trip after having used a TTE T, as

$$\pi(T+\Delta T|T) = \frac{P(T+\Delta T)}{P(T)} = \exp\left(-\frac{\Delta T}{\bar{\beta}}\right) . \tag{S10}$$

and using the definition (S3) of the hazard function, we set

$$\lambda(T) = \left\langle \frac{1-\pi(T+\Delta T|T)}{\Delta T} \right\rangle = \int_0^\infty \left(\frac{1-\exp(-\Delta T/\bar{\beta})}{\Delta T} \right) \theta_{x_{max}/\langle n\rangle} \left(\frac{\Delta T}{T}\right) d\Delta T \,, \tag{S11}$$

and using the definition (S9) we derive an analytical expression for the hazard function

$$\lambda(T) = \frac{1}{\alpha}\int_0^{x_{max}} \exp\left(-\frac{Tx}{\langle n\rangle\alpha}\right) \frac{1-\exp(-Tx/(\langle n\rangle\bar{\beta}))}{x} dx \,, \tag{S12}$$

where we introduce the timescale α (see eq. (S5))

$$\alpha \simeq \langle n\rangle\langle t\rangle/x_{max} \,. \tag{S13}$$

The time scale (S13) is consistent with the empirically evaluated timescale for the short TTE under-expression (see Fig. 5 left) with $x_{max} \simeq 2$. A numerical integration of eq. (S12) provides a hazard function which has the same behaviour as the interpolation (S5) derived from the empirical GPS data. In the Fig. 5 right, we show a comparison between the integral (S12) and the empirical hazard function where $\beta = 1.08\bar{\beta}$ and α computed from the previous relation. We remark the presence of a scaling factor between the empirical evaluated β and the theoretical expected value $\bar{\beta}$. More precisely β proves to be an overestimate of $\bar{\beta}$ since the incremental ratio in the integral (S11) decreases as the cost ΔT becomes large. In other words, according to the time consumption model, the empirical data bestow a greater TTB to individuals with respect to the theoretical value, due to the reduced perception of the trip cost when the TTE increases.

Properties of the average trips' duration $\langle t\rangle$ and average number of trips $\langle n\rangle$.

We consider the correlation between the parameters α and β with the average travel-time $\langle t\rangle$ for a single trip. The results show that $\langle t\rangle$ has a positive correlation of 0.57 with β and no correlation with α. But $\langle t\rangle$ is strongly correlated with the average house costs per square meter (Supplementary Fig. S2 (b)) and negatively correlated with the municipalities surface (Supplementary Fig. 3 (b)). This empirical observation could be a consequence of the activities sprawling in the larger cities, whereas they are concentrated inside the historical center for the smaller cities. The relationship between $\langle t\rangle$ and the average trip's speed seems instead not trivial (Supplementary Fig. (d)). As a matter of fact, an almost constant average trip length of ≈ 5.3 km is observed in the majority the cities, independently by the municipality area. Therefore one expects a relation $\langle t\rangle v = const$ among the cities, where v is the average travel speed characteristic of the different road networks. Indeed, if we exclude Rome, whose spatial scale is much larger than that of all the other cities, the cities with an average speed greater than 25 km/h verify this relation, whereas we observe a strong deviation from the theoretical curve in the cities with average speed < 25 km/h. We interpret this effect as the result of a different dynamic regime in the road network: when the average travel speed is low the stochastic effects due to the stops at crossings or to congestion effects could strongly influence the vehicle dynamics, so that the proportionality between covered distance and time is lost. On the contrary a high average travel speed suggests that the free flow is dominant and the vehicle dynamics can be described in a deterministic way.

If one computes the number of daily trips n, whose empirical distribution $p(n)$ shows an exponential tail [1], we see that the limiting average values of $\langle n\rangle$ are strongly anti-correlated (correlation coefficient -0.78) with the average trip length $\langle t\rangle$, suggesting a tradeoff consistent with the concept of TTB. This seems confirmed by the negligible correlation (correlation coefficient -0.17) between $\langle n\rangle$ and β, whereas between $\langle n\rangle$ and α the correlation is 0.40, reflecting the role of α as a measure of the time needed for the necessary mobility. Finally, we have a remarkably low correlation (correlation coefficient 0.13) between the average number of daily trips $\langle n\rangle$ and the average trip's speed, which confirms that mobility induced by travel-time savings is not due to a larger number of trips but to longer trips [2].

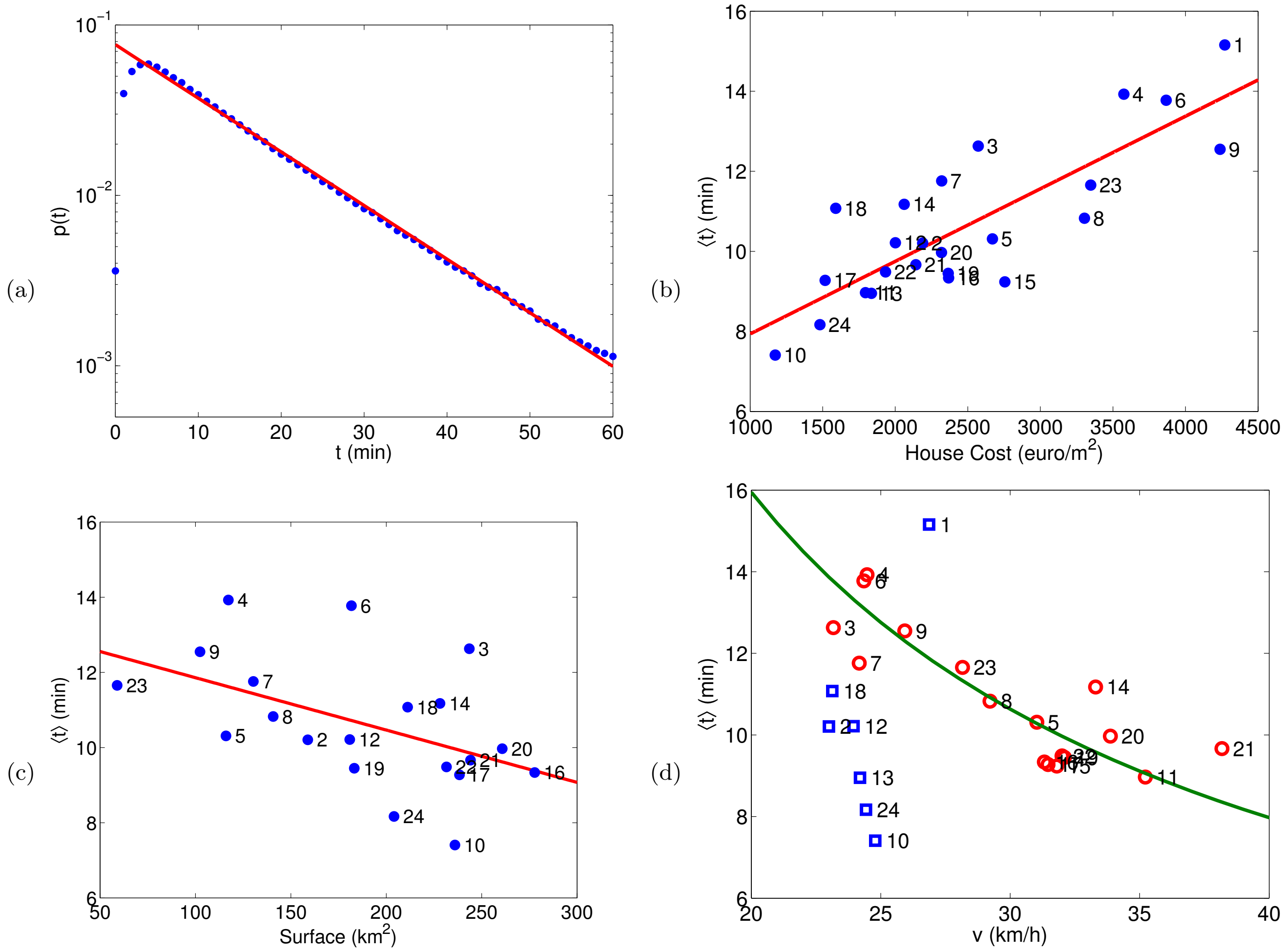

FIG. S2: **Properties of the average trips' duration** $\langle t \rangle$ **(a) Exponential distribution.** The $p(t)$ distribution (dots) for Milan and exponential interpolation of its tail (solid line) with $\langle t \rangle = 13.8$min; **(b) Growth with house prices**. Travel-times grow significantly in cities where housing is more expensive (correlation coefficient 0.83, source: www.immobiliare.com); **(c) Decrease with the city surface.** Travel-times tend to be reduces in ider cities (correlation coefficient -0.49); **(d) Decrease with travel speed.** A part of the cities lie approximatively on an hyperbole (solid line), representing a constant average length of $\approx 5.3 km$.

Disaggregated analysis for the city of Milan

To study the effect of individual heterogeneity we have disaggregated the empirical data into different classes of drivers. This analysis has been performed for the city of Milan. Due to the absence of any metadata the features to characterise the individuals have been extracted from the GPS data according to:

- home location, identified by the parking place where the cumulative parking time is the longest one [3];
- number of days in which the individual have used the car during the month;
- structure of the mobility network: mono-centric or polycentric [4].

To identify differences in TTE dependence from home location we have divided the Milan municipality in three concentric areas, according to the central structure of the city. We have chosen circular boundaries that we can approximatively associate with:

i) the area within the inner ring road (Cerchia dei Bastioni) identified as the *Zona C*, the name that identifies the congestion charge area;

ii) the area between the inner and the outer ring road (Cerchia dei Navigli), that we call *city center* in Fig. 3 (b);

iii) the *periphery*, outside the outer ring road.

Among the drivers identified as citizens of Milan, 7% live in the Zona C, 27% live in the city center and 63% live in the periphery. The remaining 3% are individuals whose home locations we found outside the city area and they were excluded from the analysis. To point out differences in the home's role, we take into account all the mobility performed in and out the municipality area of Milan, evaluating the percentage rt of round trips involving home as origin or destination. When $rt > 75\%$ we define the individual mobility network as mono-centric: 58% of the drivers in Milan have this property. Conversely, if $rt < 75\%$, we can introduce a second hub in the individual mobility network [4], which has a significative role in the organisation of the individual mobility.

Numerical formulation of the time consumption model

Each individual accumulates progressively the travel times into the total travel-time $T_n = \sum_{i=1}^{n} t_i$, where n is the number of daily trips. From the other hand each trip is associated to a performed activity and it is possible to introduce an utility function U [5], representing in some preference scale the satisfaction and/or the advantages derived by performing that activity. Without any further information, our null hypothesis is that the activity utility U is a random variable uniformly distributed in the interval $(0, 1)$ (arbitrary units). To define a behavioural model we assume that this utility is counter-balanced by a cost due to the time already spent driving until that moment.

Each trip represents an increment of total travel time $\Delta T = t$. Travel-times t are distributed exponentially with average fixed at the experimental values of table S1. The total cost associated to travel is not to be quantified not proportionally to T, but to its logarithm $\log(T)$ plus a certain constant to exclude negative values. Then the cumulative utility $\mathcal{U}$ is given by the linear combination:

$$\mathcal{U} = c_1(c_2 U - \log(T)) \tag{2}$$

The parameter c_2 represents the cost/benefit ratio, which could be associated to a value of time), while c_1 is the unit measure of the utility scale which is associated to the shape of the logistic threshold. T is the individual TTE distributed according to eq. (S7).

If we evaluate the probability of performing a daily activity according to the logistic model [6]

$$p(U) = \exp(\mathcal{U})/(1 + \exp(\mathcal{U})) \tag{3}$$

using Monte-Carlo simulations, the related TTE distribution turns out to be very similar to the empirical one (see Fig. 3 right). For all considered cities, the best fits have a $R^2 > 0.986$. Similarly to what observed in the main text, this correspondence does not happen assuming costs proportional to T.

---

* Electronic address: rgallotti@gmail.com

[1] Gallotti R, Bazzani A, Rambaldi S (2012) Toward a statistical physics of human mobility. Int J Mod Phys C 23: 1250061.
[2] Metz D (2008) The myth of travel time saving. Transp Rev 28: 321–336.
[3] Bazzani A, Giorgini B, Rambaldi S, Gallotti R, Giovannini L (2010) Statistical laws in urban mobility from microscopic GPS data in the area of Florence. J Stat Mech P05001.
[4] Gallotti R (2013) Statistical physics and modeling of human mobility. Ph.D Thesis, University of Bologna, Italy, 52–53.
[5] Helbing D (2010) Quantitative Sociodynamics. Springer-Verlag Berlin Heidelberg.
[6] Domencich T, McFadden D (1975) Urban travel demand: A behavioral analysis. North-Holland, Amsterdam.