A simple model for drifting buoy life-times, and a method for estimating evolution of a network's size

Etienne Charpentier[1], Mathieu Belbéoch[1], Julien Bourcier[2]

Summary

Using JCOMMOPS database, and series of actual drifter life-times, it was shown that survivability of a drifting buoy network fits very well with an exponential model $N(t)=N_0\ e^{-\lambda t}$. Using regression method, λ coefficient can be estimated. This coefficient is directly linked to the theoretical network half life $L_{1/2}$: $\lambda = \ln(2)/L_{1/2}$. Assuming a constant deployment rate of R_x drifters per day, it is therefore possible to estimate evolution with time of the size of a network of consistent buoys: $N(t)= (N_0 - R_x/\ \alpha)\ e^{-\lambda t}+R_x/\ \lambda$. Network size will eventually tend towards theoretical limit of $R_x/\ \lambda$ buoys. In order to maintain a buoy network at a target level of N_t units, it is therefore recommended to deploy buoys at a rate of $N_t\ \alpha$ units per day. A higher deployment rate will be needed in case the initial number of buoys is substantially lower than the target. New JCOMMOPS web page permits to query the database in order to make quick estimation of λ coefficient and to simulate network evolution based on provided criteria. This works, of course provided that buoy deployment information, and particularly deployment date is made available to JCOMMOPS. Recently developed buoy metadata collection system and deployment notification scheme permits to collect required information and to assist buoy operators in managing their networks.

[1] *JCOMMOPS*
[2] *:IUP, University Paul Sabatier, Toulouse*

1) The exponential model

A set of consistent observational platforms (units) deployed at the same time are supposed to follow a probability density function f(t) of exponential form: f(t) = $\lambda . e^{-\lambda . t}$. Justification for the exponential model is given in paragraph 4. The following characteristic functions are derived (*NIST/SEMATECH, 2005*):

- ***Probability density function*** : f(t), t>0.

  $$f(t) = \lambda . e^{-\lambda . t}$$

- ***Cumulative distribution function*** : F(t) = P(T < t) = $\int_0^t f(x).dx$. It gives the probability that a randomly selected unit dies before t (i.e. will fail by time t). $F(t_2)-F(t_1)$ represents the probability that a randomly selected unit fails between t_1 and t_2 (i.e. will survive to time t_1 but fail before time t_2). It also provides for the proportion of the population that fails by time t.

  $$F(t) = 1 - e^{-\lambda . t}$$

- ***Reliability function or survival function*** : R(t) = P(T ≥ t) = 1 - F(t) = $\int_t^{\infty} f(x).dx$. It measures probability that a randomly selected unit survives beyond time t. It also provides for the proportion of the population that survives beyond time t.

  R(t) = $e^{-\lambda.t}$

  So statistical evolution of network size is given by the formula:

  $$\boxed{N(t) = N_0.e^{-\lambda.t}}$$

  With N_0 = number of units deployed at time t=0, assuming all units have been deployed at t=0.

- ***Instantaneous failure rate*** : h(t) = $\frac{f(t)}{R(t)} = \frac{-d[\ln(R(t))]}{dt}$. This is also called hazard function or conditional failure rate. It measures instantaneous rate of failure for the survivors at time t.

  h(t) = λ

- C***umulative hazard function*** : H(t) = $\int_0^t h(x).dx$ = -ln (R(t)).

  H(t) = λ.t

- ***Average failure rate between T_1 and T_2*** : AFT(T_1,T_2) = $\frac{H(T_2)-H(T_1)}{T_2-T_1}$

  AFT(T_1,T_2)= λ

- ***Life expectancy = mean life-time =*** $\int_0^{\infty} x.f(x).dx = \int_0^{\infty} x.\lambda.e^{-\lambda.x}.dx = \frac{1}{\lambda}$

- ***Half life*** : $R(t) = e^{-\lambda.t} = \frac{1}{2} \Rightarrow half-life = \frac{\ln(2)}{\lambda}$

2) Evolution after time t=0 of a network of units which had been deployed at different times but with no further deployment after time t=0

We'll assume that we start with a network of M_0 units all deployed at time $-t_0$.
Between time $-t_0$ and 0 we'll assume a constant deployment rate R_x.
Units deployed before and after time t=0 are assumed to be consistent in terms of reliability (i.e. same coefficient λ).
We'll consider no further deployment for $t \geq 0$. This scenario (a) is described in figure 1 below.
We are interested in the evolution of the network size after t=0.

For t between $-t_0$ and 0 : Let g(t) be the number of units that had been deployed at time t.
Deployment rate at time t is given by g'(t). Here $g'(t)=R_x$.

- For $t=-t_0$: M_0 units.
- For $-t0 \leq t_1 < 0$, for units deployed between time t_1 and t_1+dt number of remaining units at time $t \geq 0$ is given by $g'(t_1).e^{-\lambda.(t-t_1)}.dt = R_x.e^{-\lambda.(t-t_1)}.dt$.
- For $t \geq 0$: No further deployment, units deployed before t=0 just continue their lives and eventually die.

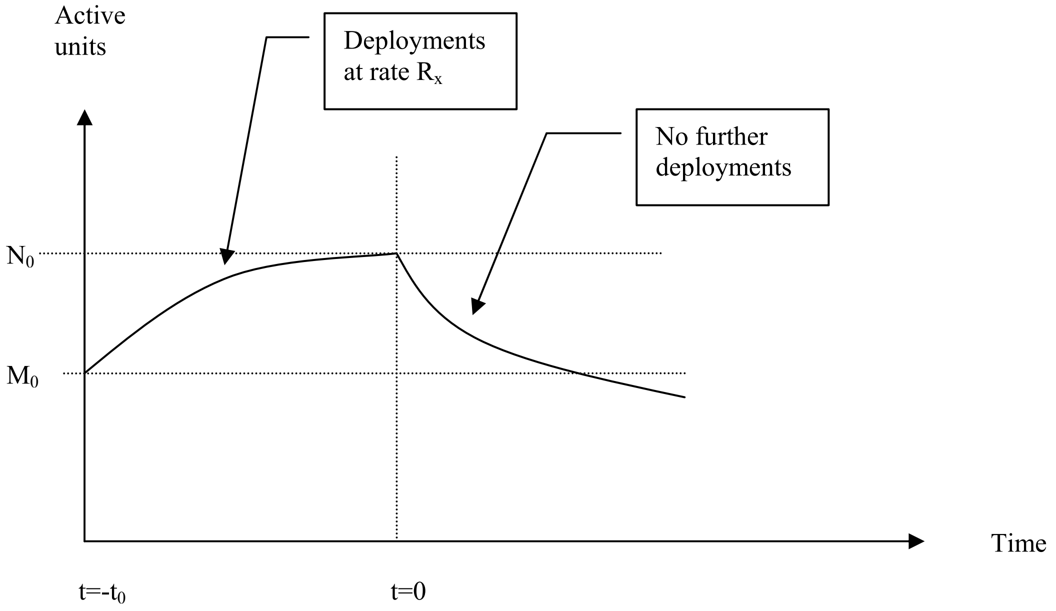

Figure 1: Number of active units as a function of time for scenario (a).

At time $t \geq -t_0$, for the M_0 units deployed at time $t=-t_0$, we have $M_0.e^{-\lambda.(t+t_0)}$ left in the network.

Hence, considering all units deployed between time $-t_0$ and 0, number of remaining units at time $t \geq 0$ is given by

$$N(t) = M_0.e^{-\lambda.(t+t_0)} + \int_{-t_0}^{0} R_x.e^{-\lambda.(t-x)}.dx$$

$$N(t) = [(M_0 - \frac{R_x}{\lambda}).e^{-\lambda.t_0} + \frac{R_x}{\lambda}].e^{-\lambda.t}$$

If we define $N_0 = (M_0 - \frac{R_x}{\lambda}).e^{-\lambda.t_0} + \frac{R_x}{\lambda})$

Hence,

$$\boxed{N(t) = N_0.e^{-\lambda.t}}$$

N_0 is also the number of active units at time t=0.

When we stop deployments at a constant deployment rate, the network size evolution therefore follows the same law as for units all deployed at the same time.

3) Evolution of a network of units deployed at different times with continued deployments.

We'll assume that at time t=0 we start with a network of units which had been deployed according to the scheme defined in paragraph 2 above but with an unknown deployment rate R_{1x}. However, we'll assume we know the network size at time t=0, i.e. N_0.
For $t \geq 0$, we'll assume a constant deployment rate R_x. Units deployed before and after time t=0 are assumed to be consistent in terms of reliability (i.e. same coefficient λ). This scenario (b) is described in figure 2 below. We are interested in the network size evolution after t=0. An approximation of this is to look at a network size evolution as of time t=0 with a constant deployment rate R_x after that, but where the deployment history before t=0 is unknown.

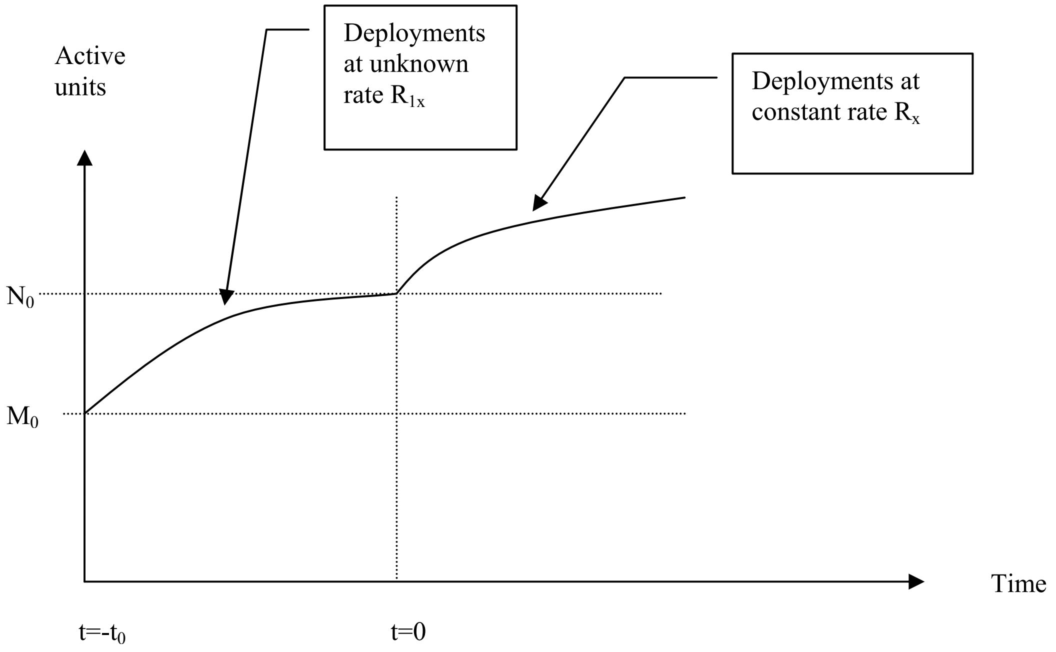

Figure 2: Number of active units as a function of time for scenario (b).

We've seen in paragraph 2 above that units deployed before t=0 follow the law for $t \geq 0$:

$$N(t) = N_0.e^{-\lambda.t}$$

Let g(t) be the number of units that had been deployed at time t. Deployment rate at time t is given by $g'(t)=R_x$.

For units deployed between time t_i and t_i+dt number of remaining units at time $t \geq 0$ is given by $g'(t_i).e^{-\lambda.(t-t_i)}.dt = R_x.e^{-\lambda.(t-t_i)}.dt$.

Hence, considering all units deployed between time 0 and t, number of remaining units at time t is given by

$$N(t) = N_0 . e^{-\lambda . t} + \int_0^t R_x . e^{-\lambda .(t-x)} . dx$$

Hence

$$N(t) = N_0 . e^{-\lambda . t} + \frac{R_x}{\lambda} .(1 - e^{-\lambda . t})$$

$$\boxed{N(t) = (N_0 - \frac{R_x}{\lambda}) . e^{-\lambda . t} + \frac{R_x}{\lambda}}$$

4) Justification for the exponential model

JCOMMOPS database contains information about actual drifting buoy and Argo profiling float life-times. This is based on information collected through:

- Global Drifter Programme log file
- EGOS historical database
- Argo database (AIC) and Argo float deployment notification scheme
- New web based drifting buoy data collection scheme implemented at JCOMMOPS
- Argos locations from Service Argos database

However, some information is still missing for many observational platforms, e.g. date when GTS transmission stopped, dates when individual sensors failed, dates when the drogue went off, date when a platform got ashore. On the other hand, location information is available for almost all of the platforms so we are suggesting in this study to define life-time as the time between the deployment date (or first location) and the time of the last location before a platform stopped being located. This approach is relatively optimistic as for example beached platforms, un-drogued ones, or platforms which data had been deleted from GTS distribution are still considered active as long as they are located.

To validate the model, we are using NOAA/AOML drifter programme Argos #6325 and considering all drifters deployed between 1 January 2002 and 30 June 2004. These can be considered as a relatively consistent set of instruments with same statistical survivability. At the time of writing this report (i.e. August 2005) many drifters were still alive so their eventual life-time is unknown. This brings a little bit of uncertainty in the study but this can be circumvented as explained below.

During that period, we had 452 drifters deployed, 172 of them still being active at the time of the study (figure 3). Observed average age is 444.9 days, and maximum age is 961 days. As more than 50% of the drifters were still alive at the time of the study, actual half life cannot be estimated. However, two approaches can be considered, i.e. an optimistic one which would assume that all alive drifters will survive forever, and a pessimistic one which would assume that they would all die immediately after the time of the study. Evolution of number of drifters with age, including optimistic and pessimistic approaches is given in figure 4. Quartiles are given in figure 5. This shows that 25% of the drifters died before 275 days, and that actual half life is between 465 and 522 days. Last quartile cannot be resolved because more that 25% of the drifters would survive forever in the optimistic approach.

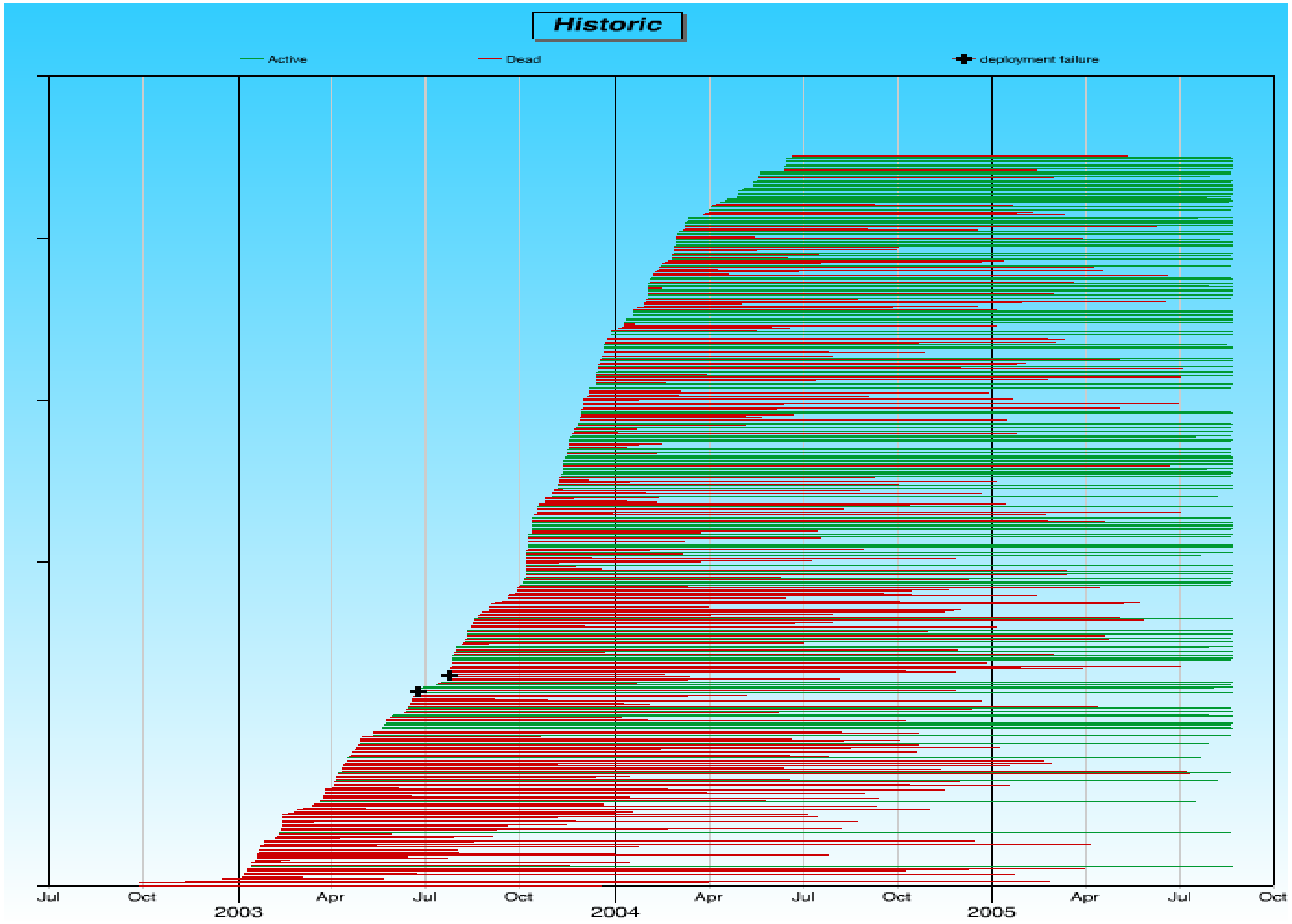

Figure 3: deployment and life history for the considered set of drifters. In red are drifters which died, in green those which were still active at the time of the study.

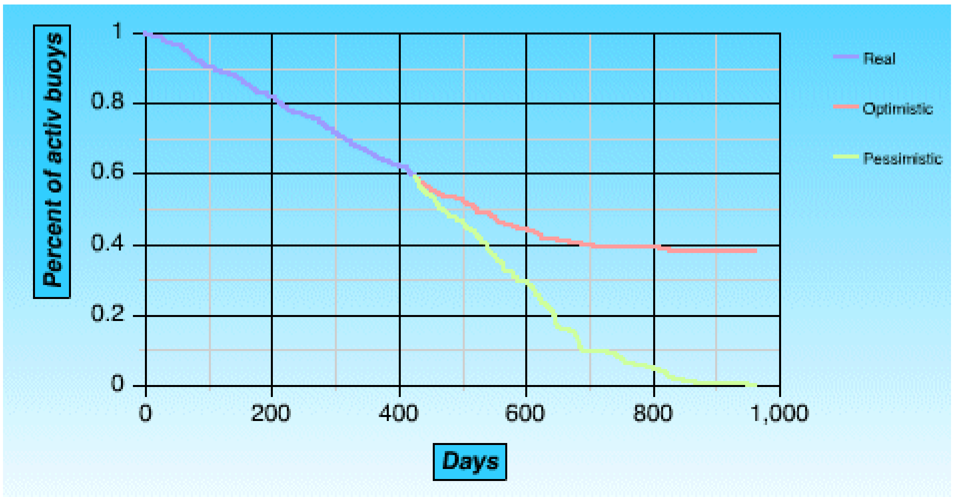

Figure 4: **Evolution of the number of remaining drifters as a function of age.**

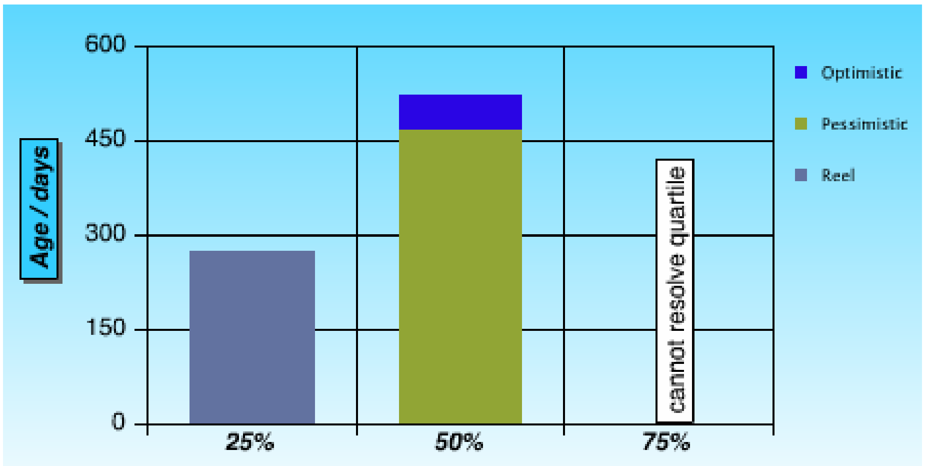

Figure 5: **quartiles for the considered dataset.**

Considering the total number of drifters N_0, for those which died before the time of the study, if we imagine that they have all been deployed at the same time, the life-time series for these drifters, i.e. the number of drifters that survived a given number of days as a function of time, N(t), would provide for the reliability function R(t)= $\int_t^\infty f(x).dx = \frac{N(t)}{N_0}$.

Hence, validating the exponential model is equivalent to validating $N(t) = N_0.e^{-\lambda.t}$. We propose using regression method using least mean squares. In order to have a consistent set of drifters, all dead, we considered only life-times that were below 414 days, as 414 days correspond to the time period between the last deployment (30/06/2004) and the date of the study (18/08/2005). For life-times beyond 414 we would have a mixture of dead and alive floats and an inconsistent dataset. This is also where the optimistic and pessimistic curves in figure 4 diverge.

Figure 6 shows the fit with the exponential model for ages below 414 days. Obtained value for λ is 0.0012301438, with a correlation coefficient of 0.9965. Fit is excellent. Model half-life is $L_{1/2}=\ln(2)/\lambda=563$ days which is a bit beyond the upper optimistic limit for actual half-life: even though in the optimistic approach active units are assumed to survive forever, other units which had their life-times exceeding 414 days but died before the date of the study (and that had not been considered for the regression computation) had a failure rate greater than the model. The model is therefore more optimistic than the reality for life times exceeding 414 days.

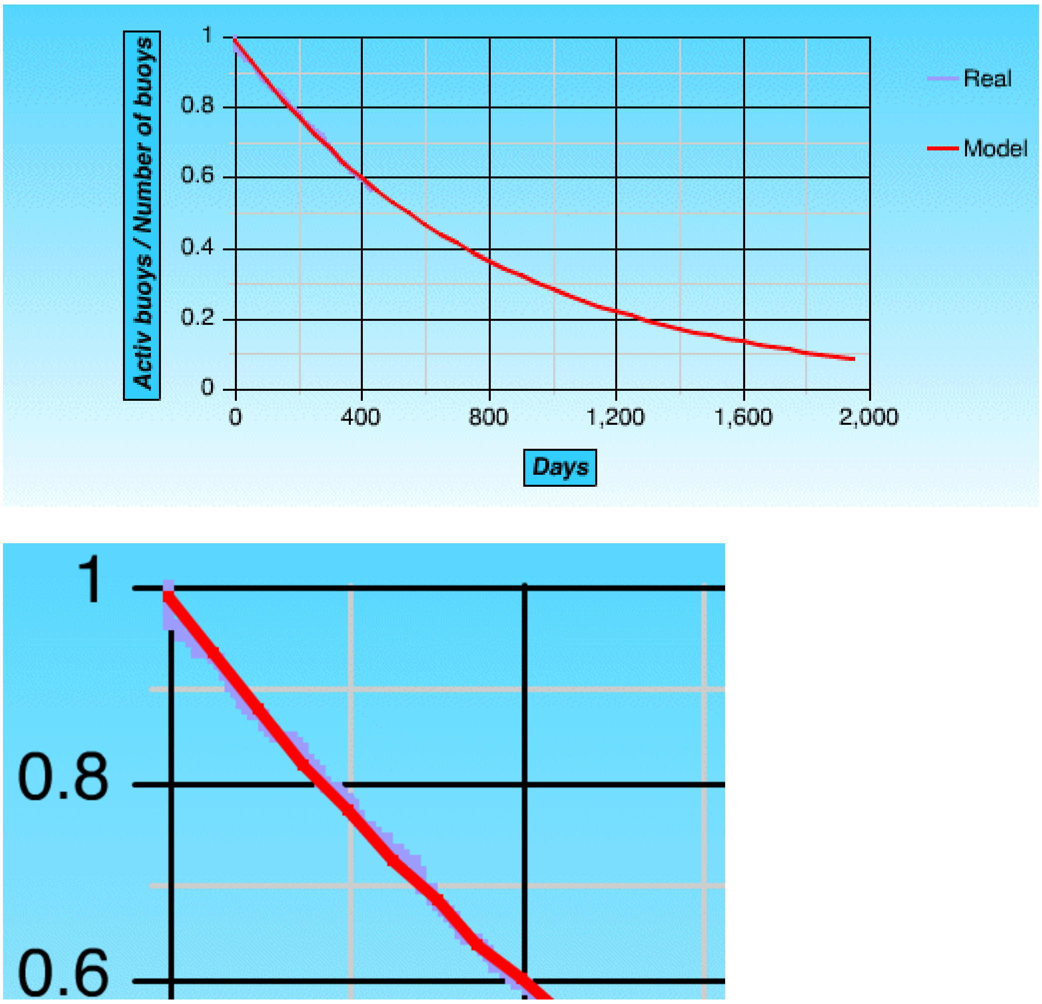

Figure 6: Fit with exponential model (top) and close up for period 0-414 days(bottom).

Figure 7 below proposes to validate our model using hindcast. In other words, one curve (blue) shows the evolution of the network size with time while considering real deployments and deaths, while the other curve (red) shows evolution of the network size while considering real deployments but simulating drifter death according to the model. Model provides for a probability to die at any given age. Random number generator is used to decide whether any given drifter alive at a given time should die using model probability. After June 2004, no further deployments are being considered in our dataset so the two curves drop. The two curves remain pretty close one from another.

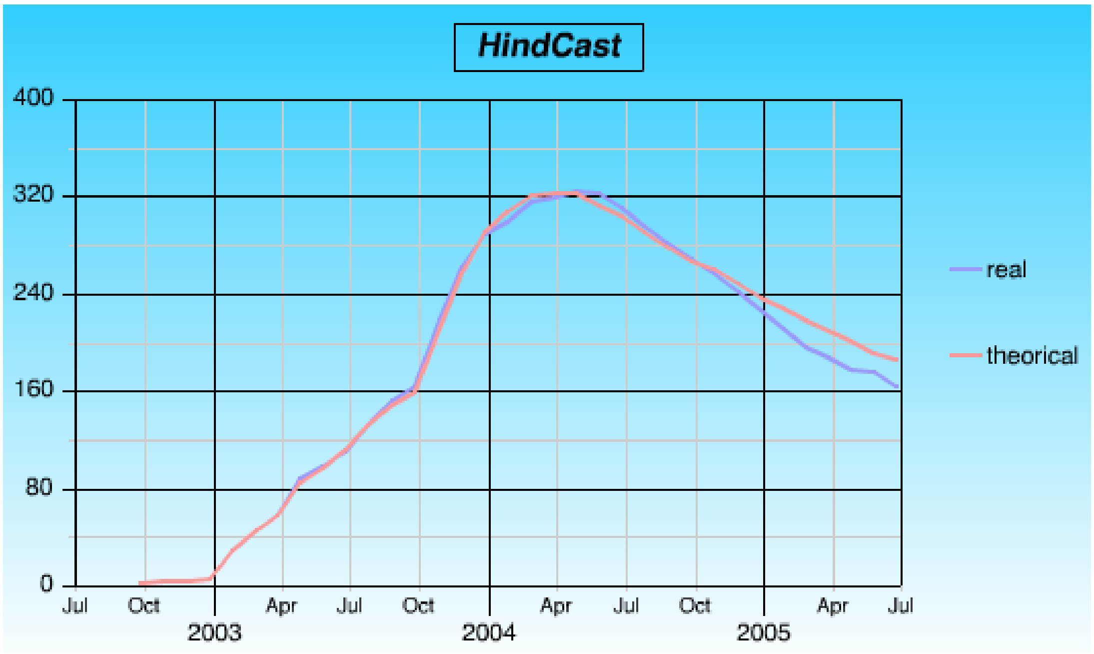

Figure 7: Hindcast of proposed model compared to the reality. Number of active buoys versus time.

Now that our model has been validated, we can start playing with it, and simulating future buoy deployments. We will work under the assumptions made under paragraph 3 and therefore use following formula for network size evolution:

$$N(t) = (N_0 - \frac{R_x}{\lambda}).e^{-\lambda.t} + \frac{R_x}{\lambda}$$

Figure 8 (or resolving following equation) shows that if we start with a network of 1000 units at time t=0 and want to achieve a target of 1250 operational units in our network in 1 year (365 days), then the equation tell us that we should deploy units at a rate R_x of 760 units per year. Once the target has been reached, in order to maintain the network at the target level, then we should deploy units at a rate Rx=1250.λ=562 units per year.

- Equation to reach target in 1 year: $1250 = (1000 - \frac{R_x}{\lambda}).e^{-\lambda.365} + \frac{R_x}{\lambda}$

⇒ R_x=2.08 units/day = 760 units/year

- Equation to maintain network at target level (t=∞): $1250 = \frac{R_x}{\lambda}$

⇒ R_x=1.54 units/day = 562 unis/year

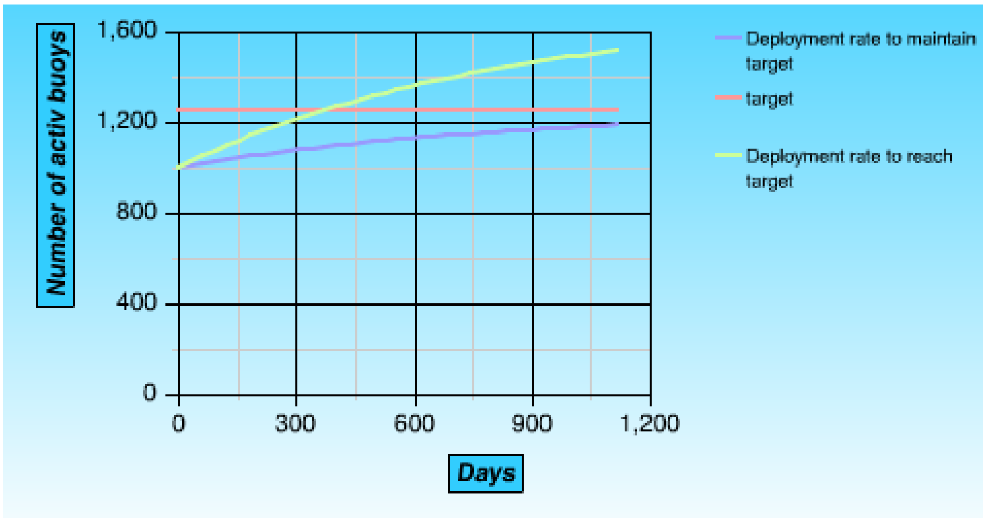

Figure 8: Reaching the target at high deployment rate (yellow), and then maintaining it at the target level using reduced deployment rate (blue).

Figures 1 through 8, and model simulation can be obtained for any platform data-set using JCOMMOPS dynamic web site; go to http://www.jcommops.org, click on “PLATFORMS” on the right menu, then enter query fields to select the platforms you are interested in, and click on “Show life-time stats” button.

Remark: Validity of the exponential model can also be explained by a particular shape of the typical bathtube curve for non-repairable system (*NIST/SEMATECH, 2005*) for h(t) which considers three typical periods, i.e. (i) the early failure period, (ii) the intrinsic failure period, and (iii) the wearout failure period. For drifting buoys, early failures can be considered separately, e.g. if we observe a 10% early failure rate, then we can first work on the assumption that there is no early failure, and as a second step add 10% to the obtained deployment rate. Drifters are non-repairable. Also, for drifters, because most of the failure causes are due to either harsh marine environment or battery failures, most buoys die before ever reaching the wearout failure period (electronics, sensors). So for our applications we are typically on the flat part of the bathtub curve. Assuming a flat, constant, function h(t)=λ leads by definition to the exponential model.

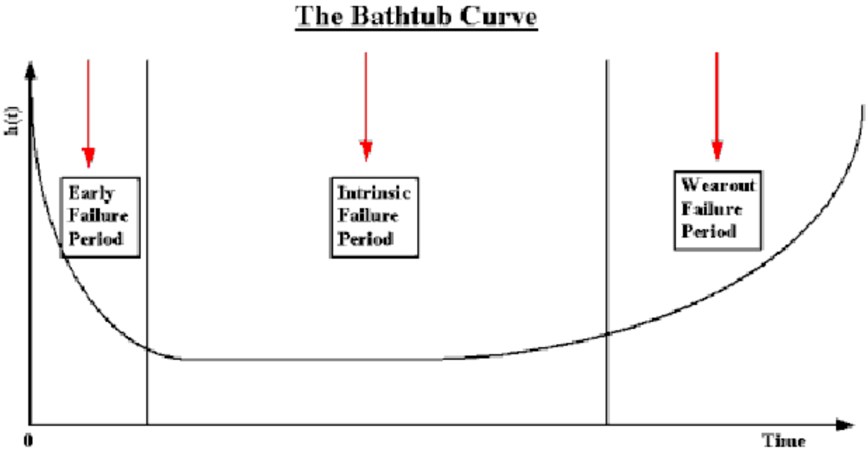

Figure 9: **The typical bathtub curve h(t).**

References:

US National Institute of Standards and Technology (NIST), NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/ , August 2005.