Advances in Complex Systems

AN EMPIRICAL MODEL FOR ELECTRONIC SUBMISSIONS TO CONFERENCES

PATRICK FLANDRIN

Université de Lyon, École Normale Supérieure de Lyon
Laboratoire de Physique (UMR 5672 CNRS)
46, Allée d'Italie, 69364 Lyon Cedex 07, France.
flandrin@ens-lyon.fr

Received (received date)
Revised (revised date)

Electronic submission to a conference is a process that is known to evolve nonlinearly in time, with a dramatic increase when approaching the deadline. A model has recently been proposed by Alfi *et al.* (Nature Physics, 2007) for such a process, and the question of its universality has been raised. This problem is revisited here from a data analysis and modeling point of view, on the basis of a larger data set. A new model is proposed that better describes the total evolution of the process (including saturation) and allows for a running prediction of the total number of submissions.

Keywords: Deadline; publication; growth phenomena.

1. Introduction

For anyone who ever organized a conference, it is well-known that most of the electronic submissions are to be expected during the very last days (or even hours) before the deadline. This folklore observation has recently received a more quantitative attention [1], ending up eventually with a possible model (referred to in the following as "APP", for Alfi-Parisi-Pietronero) for the submission process prior the deadline and suggesting its possible universality. The analysis proposed in [1] was based on a very simple assumption (the probability of submission was assumed to be inversely proportional to the lasting time before the deadline) and it resulted in a model that fitted remarkably well the analyzed experimental data but, at this point, two remarks can however be made:

(1) The first one is related to the APP model itself which, strictly speaking, predicts a finite time singularity in the form of a logarithmic divergence when approaching the deadline (see Sect. 3.1 below), a property that has to be regularized in some way since, in practice, the total number of submissions has always to be finite. As a corollary, the prediction of this number is not easy, and the rule of thumb proposed in [1] was purely *ad hoc* and not any by-product of the model.

(2) The second remark concerns the supporting data which essentially consisted in only one major conference and one more local workshop. Whereas the proposed modeling was clearly relevant for such data, it is obvious that more examples would be needed for assessing some form of universality in the reported behaviour. Moreover, it is likely that the results might also depend on the type of conference considered and/or on the concerned scientific community.

In order to address those issues, another (somehow larger) set of data is envisioned here. This data set is first described in Sect. 2, whereas a visual inspection of the corresponding time series will suggest a refined model (as compared to the APP one) that is discussed in Sect. 3, in both terms of fit and prediction of the total number of submissions. Finally, interpretation issues and still open questions about universality will then be addressed in Sect. 4.

2. Data

2.1. *Data set*

The data set considered in this paper consists in the timestamps of electronic submissions to seven conferences that play a major role in the signal and image processing (SIP) community:

- Four of them are organized at a national level (the GRETSI symposia which are, every two years and under the auspices of the GRETSI association [2], the largest French-speaking event);
- One at the European level (the European Signal Processing Conference EUSIPCO that takes place every year and is organized by the EURASIP association [3]);
- Two at a worldwide level (a more specialized biennal workshop: the IEEE Statistical Signal Processing Workshop SSP, and a very big annual convention: the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP, which is by far the most largely attended conference in the field, both events being organized by the IEEE-SP Society [4]).

Locations and dates of the conferences analyzed here are as follows:

- GRETSI-03, Paris (F), Sept. 8-11, 2003
- GRETSI-05, Louvain-la-Neuve (B), Sept. 6-9, 2005
- GRETSI-07, Troyes (F), Sept. 11-14, 2007
- GRETSI-09, Dijon, Sept. 8-11, 2009
- EUSIPCO-4-03, Vienna (A), Sept. 6-10, 2004
- SSP-05, Bordeaux (F), July 17-20, 2005
- ICASSP-06, Toulouse (F), May 14-19, 2006

and details of the corresponding total numbers of submissions are given in Table 1.

Table 1. Actual numbers of submissions of the considered conferences, estimated parameters of the proposed model (see Sect. 3.2 and 3.3) and predicted total number of submissions (see Sect. 3.4). In some cases, the initial deadline has been officially postponed and results corresponding to both deadlines are reported.

conf.	G-03	G-05	G-07	G-09	E-04	S-05	I-06
# subm.	423	462	481	402	876	438	3903
β	2.6	2.5	1.6	2.5	1.6	1.3	1.7
	2	-	-	-	1	1.3	-
T_* (days)	0.25	-0.1	-2	-1.3	2	1	-0.4
	-1	-	-	-	-0.2	-1.5	-
1 week pred.	410	451	487	395	894	456	4078

2.2. *Time series*

In all cases, the data have been recorded from the opening of the submission websites to their effective closing, this latter taking always place later than the deadline initially announced. Timestamps are given in UNIX time and thus known with a precision of one second, but their analysis has been carried out with no noticeable difference either on the basis of all timestamps with their actual time locations or on a smaller subset corresponding to the situation of submissions at regularly spaced intervals (typically one day). For a sake of simplicity and readability, the results reported here will correspond to such uniformly sampled time series, with about fifty samples in each case, corresponding to a submission period that is generally of the order of two months.

2.3. *Specificity*

In terms of specificity, all the chosen events can be considered as attended by essentially the same (SIP) scientific community, which certainly differs from that (of Statistical Physics) attending the conference StatPhys 23 considered in [1]. Moreover, whereas submission to StatPhys meetings consists in sending a short abstract only, submission to those SIP conferences implies the sending of a much more complete material, namely the full paper (4 or 5 pages, double column) in the case of ICASSP or EUSIPCO, and an extended summary (3 or 5 pages, single column) for GRETSI or SSP. This is likely to impact on the timing of submissions, with a tendency to skew the distribution towards shorter time distances from the deadline.

2.4. *Empirical renormalization*

While diverse in size, the considered conferences can however be viewed as representative of a common situation in terms of field, community and style, and the eyeballing of the different submission histories reveals indeed striking resemblances.

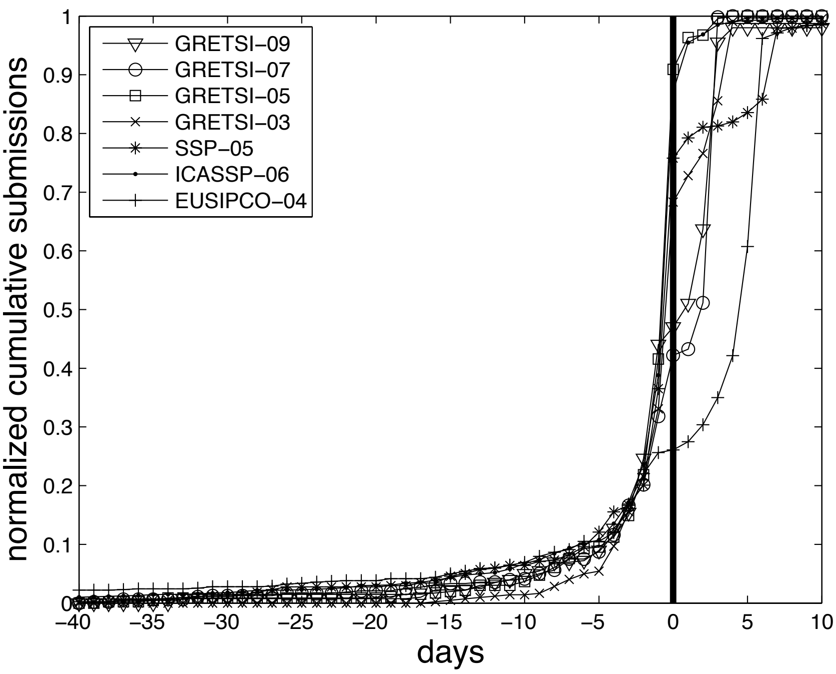

Fig. 1. Cumulative number of submissions as a function of time. All data have been renormalized by their corresponding total number of submissions so as to saturate at unity, and shifted relatively to a deadline arbitrarily fixed at $T = 0$ so as to superimpose in the regime prior the first announced deadline. Variations after $T = 0$ correspond to postponed deadlines.

This is reported in Fig. 1 where, after a renormalization by the total number of submissions in each case (and some possible deadline shift that will be commented later), all data points are found to fairly well superimpose on a common curve prior the deadline, here referred as time $T = 0$ (an even better superimposition could be obtained by adjusting a scale factor on the time axis as proposed in Sect. 3.2, but this has not been done here). The dissimilarities occurring after $T = 0$ are due to deadline extensions (3 days or 1 week, announced just before the initial deadline), and it striking to observe that, when taking into account the corresponding shifts, the same type of behaviour is recovered.

In short, the submission process can be depicted by a nonlinear evolution that is highly accelerated when approaching the deadline and slowly relaxed for late submissions (when still accepted).

3. Models

3.1. *APP and other divergent models*

Up to the saturation effect, evolutions such as those in Fig. 1 were supposed to be captured by the APP model. Indeed, this latter assumes that the "probability" of submission at time t is inversely proportional to the remaining time up to the deadline T, thus reading:

$$p(t) = \frac{C}{T - t}, \tag{1}$$

where C is some constant related to the total number of submissions.

It thus follows that the number of cumulative submissions $N(t)$ at time t is expected to be of the form

$$N(t) = \int_0^t p(s)\, ds = C \, \log \left(\frac{T}{T - t} \right), \tag{2}$$

predicting a logarithmic divergence at $t = T$.

Even if ignoring the necessary regularization of this divergence, such a model does not account properly for the present data, as evidenced in Fig. 2 where a fit according to eq. (2) cannot adequately reproduce the global evolution from early submissions to late ones (ICASSP-06 has been chosen as an example because it corresponds to the largest data set, but a similar behaviour has been observed with the other conferences). Moreover, applying the APP rule of thumb for prediction—i.e., extending the initial linear increase of the early submissions and multiply by three the intersect at the deadline [1]—would lead in all cases to a total number of submissions that is largely under-estimated (e.g., by a factor of almost 3 in the case of ICASSP-06).

As suggested in [1] for payments as opposed to registrations, an exponential "utility function" could be added to the model in order to take into account a "pressure" to postpone the submission. While certainly improving the situation in terms of fit prior the deadline (see Fig. 2), this however does not solve the divergence issue when approaching T nor it offers an easy way of making predictions about the total number of submissions.

Finally, it should be remarked that the APP model is based on the *a priori* arbitrary assumption that the probability of submission is inversely proportional to the time remaining before the deadline. Having such a probability which increases when the deadline becomes closer certainly makes sense, but the choice of an inverse proportionality is not imposed by any argument (except that, *in fine*, it permits a good data fit), and it could be replaced by other types of increasing functions. In this respect, it turns out that, if we restrict the analysis to a time span $\Theta := \{-\infty \leq t \leq T - 1\}$ that does not go further than one day prior the deadline, a

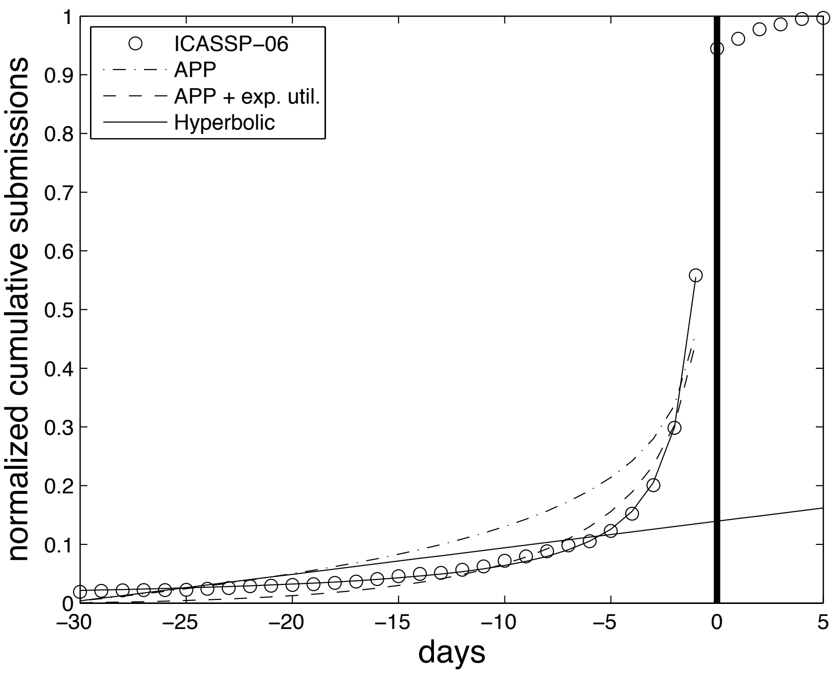

Fig. 2. Data and optimal least-squares fits with divergent models in the case of ICASSP-06. All three models are compared on a time span that does not extend beyond one day prior the deadline referred to as $T = 0$ (thick line)). The APP model with and without an exponential utility function (whose fitted time scale happens to be of about 14 days) is compared to a simple hyperbolic model. The thin straight line corresponds to the extension of a linear fit on the early part of the APP model, its intersect at $T = 0$ corresponding, when multiplied by 3, to the APP rule of thumb for the prediction of the total number of submissions.

simple hyperbolic function

$$N(t) = \frac{C'}{T' - t}; t \in \Theta \tag{3}$$

(with an actual time instant $T' \geq T$ for the singularity that might slightly differ from T) fits remarkably well the data, see Fig. 2. In such a model, the probability of submission is implicitly considered as inversely proportional to the *square* of the time prior the deadline. Again, choosing a quadratic dependence rather than a linear one can be questioned.

3.2. *An empirical model*

All those observations prompt to look for some more effective model, capable of reproducing at once the global behaviour, including the saturation. The main characteristics that is looked for is therefore some nonlinear, sigmoid-type evolution, as

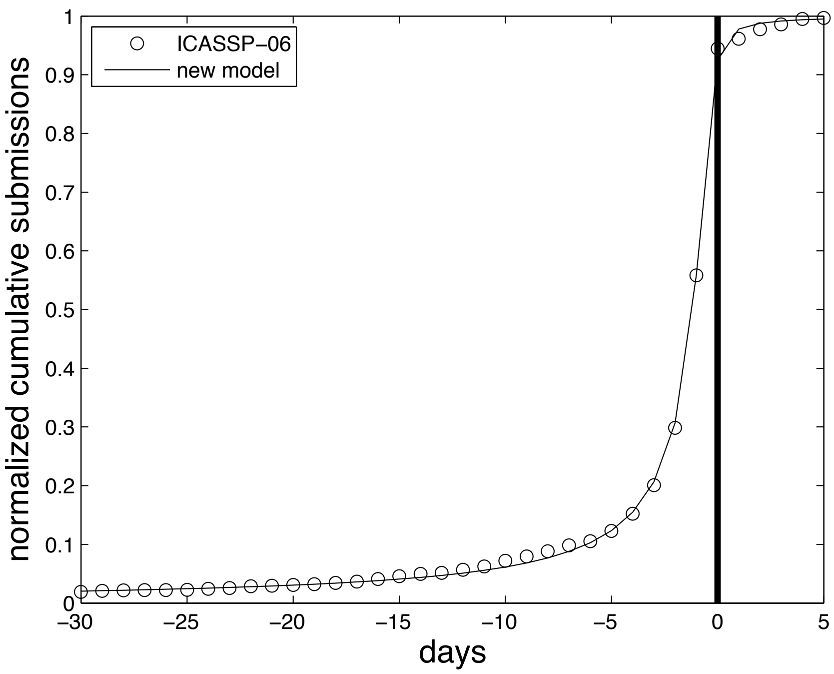

Fig. 3. Data and optimal least-squares fit with the new model in the case of ICASSP-06. In this case, the model is compared to the actual data on the total time span, before as well as after the deadline referred to as $T = 0$ (thick line)).

commonly observed in many growth phenomena that are often described thanks to variations around the logistic equation [5]. Warped sigmoid-based models (such as, e.g., those derived from the Richards growth equation [6]) turn however not to be adapted whereas better candidates (see, e.g., [7]) suffer for not having any analytic expression for the integrated growth equation.

From a pragmatic point of view, a very simple and convenient model for the cumulative number of submissions $N(t)$ as a function of time t turns out to be:

$$N(t) = \frac{1 - \exp\{-\tan^{-1}[\beta(t-T)] - \pi/2\}}{1 - \exp\{-\pi\}}\, N_{tot}, \tag{4}$$

where N_{tot} stands for the total number of submissions and β is some time scale factor. As evidenced in Fig. 3, fitting (in a least-squares sense) the model of eq. (4) to the largest data set (ICASSP-06) is very satisfactory. Moreover, it is easy to show that the model (4) admits an asymptotic expansion that is hyperbolic. Whereas such an approximation is supposed to hold for large negative times, it turns out that is still accurate up to one day prior the deadline, in accordance with the result plotted in Fig. 2.

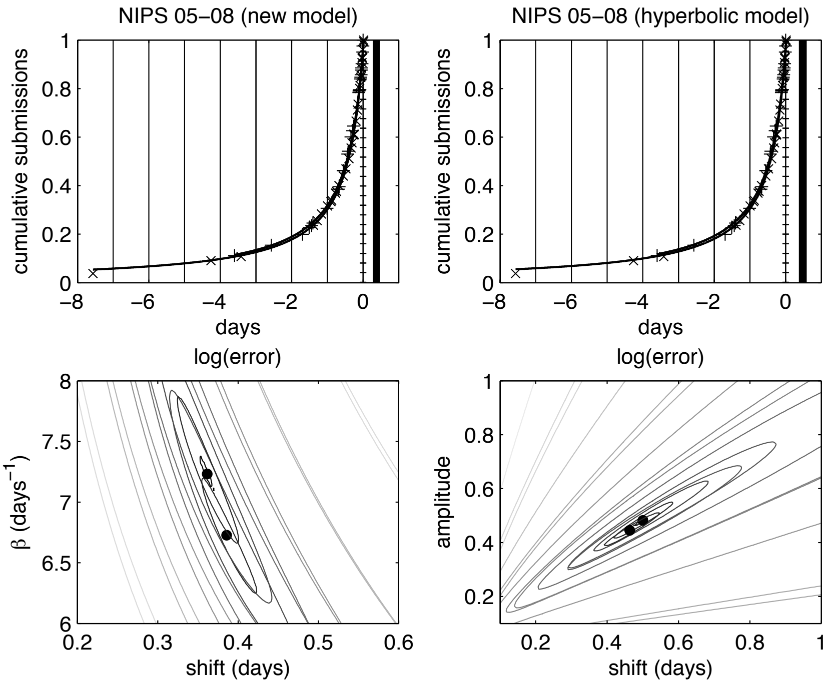

Fig. 4. Data and optimal least-squares fits in the cases of NIPS-05 (+) and NIPS-08 (x). Fits (thin lines, first row) are considered with respect to both the new model (left column) and its hyperbolic approximation (right column), whereas isocontours and minima of the corresponding surface errors with respect to the models parameters are reported (and superimposed to emphasize the low variability in the estimates) in the graphs of the second row. In both cases, the surface errors are displayed on a logarithmic scale, with a dynamic range of 30 dB for the new model and 50 dB in the case of the hyperbolic one. Data of both conferences have been normalized so as to culminate at unity, the official deadline is referred to as $T = 0$ (dotted lines) and the estimated singularity times are plotted as solid lines.

Remark. As far as the hyperbolic approximation is concerned, it can be remarked that some complementary results can been obtained, thanks to a new data set made recently available [8]. For those NSIP conferences in computer science, the situation is somehow different from the SIP ones considered so far since strictly no deadline extension is allowed, ending up with no relaxation in the growing of submission numbers. As illustrated in Fig. 4, the new model is however still relevant for this kind of data, though with some increased complexity since it now calls for the fit of a third parameter, namely the total number of submissions. In accordance with the remark made in Sect. 3.1, the restriction of the new model to the time span where the evolution has essentially an upward concavity admits an accurate approximation by means of the hyperbolic model previously defined in eq. (3). This is confirmed in Fig. 4 and it could even suggest a way of predicting the total number of submissions which, if we trust the hyperbolic approximation, is directly given by the ratio C'/T' in this specific situation where no relaxation related to late acceptances is observed after the theoretical deadline. In practice, such an estimate is however not very robust and analyzing data this way would require some further work.

3.3. *More on fits*

Turning back to the modeling based on the general model of eq. (4), one could think of the actual deadline $T = 0$ as a fixed quantity, but it is in fact more natural to consider it as a free parameter. This allows for an increased flexibility that may take into account possible deadline extensions as well as (personal) "targeted" deadlines that may differ from the actual one. Fitting this way the proposed model (4) with 2 degrees of freedom (the scale factor β and the effective deadline T_*, the total number of submissions N_{tot} being assumed to be known) to the whole data set corroborates the results obtained in Fig. 3 for ICASSP-06. This is evidenced in Fig. 5, with the corresponding numerical values of the estimated scale factors β and effective deadlines T_* reported in Table 1 (in the three cases where the initial deadline has been officially postponed, two models have been adjusted, with an estimated deadline fitted in the vicinity of each actual one).

One can remark that the values of the scale factor β have some variability, ranging from 1 to 2.6. However, in the case of SSP-05, it is interesting to point out that the same value ($\beta = 1.3$) is obtained whatever the considered deadline (initial or postponed), supporting the interpretation that the one week deadline extension resulted mostly in a pause of the same amount of time in the submission process. Similar (even if less striking) observations can be made for the two other conferences (GRETSI-03 and EUSIPCO-04) whose deadline has been postponed with an official announcement.

3.4. *Prediction*

In practice, the total number of submissions N_{tot} is of course one extra degree of freedom in any model. In this respect, (4) is of particular relevance in the prediction problem for which, given the knowledge of past submissions from the opening time of submissions t_0 until the current time t, i.e., $\{N(s|t) := N(s), t_0 \leq s \leq t\}$, the corresponding model fit $N_{opt}(t)$ permits a running average prediction $\hat{N}_{tot}$ according to

$$\hat{N}_{tot}(t) = \frac{1}{t - t_0} \sum_{s=t_0}^{t} N_{opt}(s); t > t_0, \tag{5}$$

Such predictions are plotted in Fig. 6, based on initial estimates starting in each case at $t_0/2$. What can be learned from those curves is that a pretty fair prediction (see Table 1) is possible about one week prior the official deadline, a period of time at which no more that about 10% of the papers are submitted.

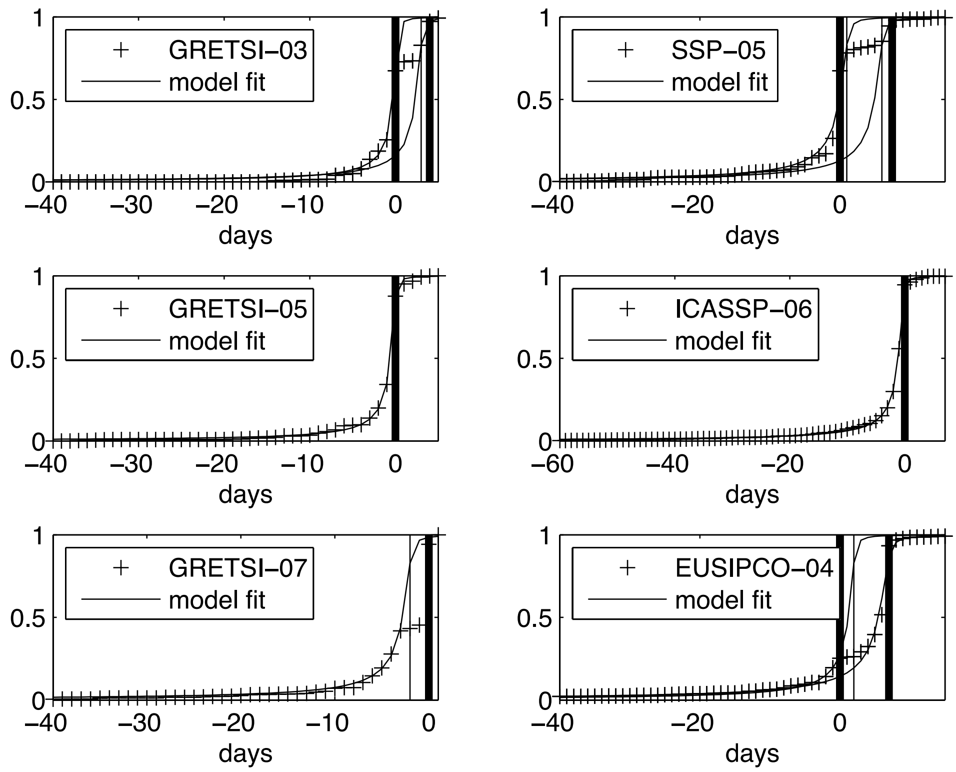

Fig. 5. Normalized cumulative submissions: data and optimal least-squares model fits. Actual (resp. fitted) deadlines are plotted as thick (resp. thin) vertical lines. In three cases (GRETSI-03, SSP-05 and EUSIPCO-04), the initial deadline has been officially postponed, with an announcement a couple of days prior the initial deadline here referred to as $T = 0$. In those cases, two models have been adjusted, with an estimated deadline fitted in the vicinity of each actual one. The corresponding values of the fitted parameters are reported in Table 1.

Since all predictions have been made under the assumption that the effective deadline should lie in the vicinity of the actual first one (no deadline extension taken into account at this point), a companion outcome of the analysis is that extending the deadline does not really increase the total number of submissions, but rather shifts the usual rush of late ones, as already suggested by Figs. 1 and 5.

4. Conclusion

A new model has been proposed for the time history of electronic submissions to conferences. It differs from the first (APP) model proposed in [1] in the sense that it includes explicitly a saturation that better describes the observed evolutions and

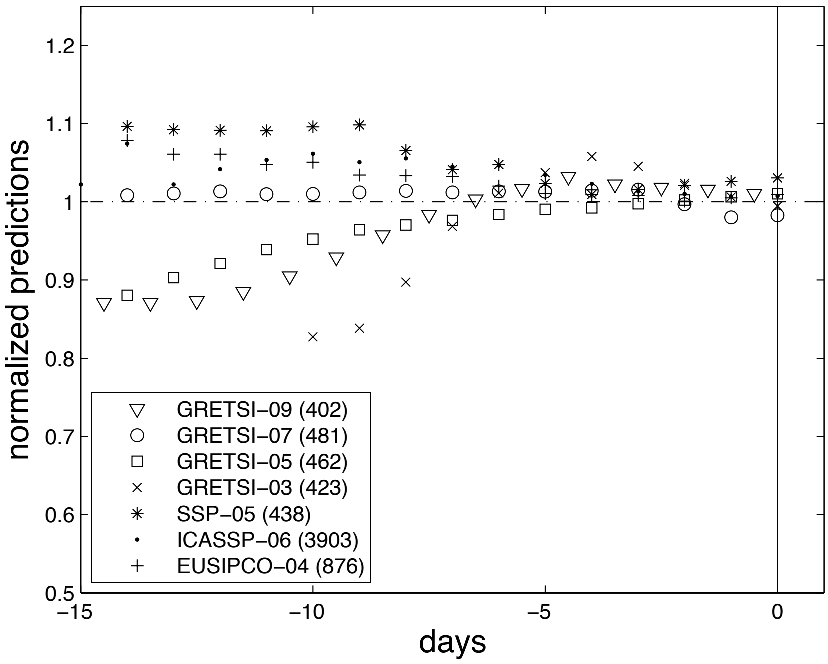

Fig. 6. Running predictions of the total number of submissions. Results are plotted as a function of time prior the official deadline (vertical line at $T = 0$), after a normalization by the actual total numbers of submissions recalled in the box.

makes easier a quantitative prediction of the total number of submissions.

From the point of view of conference organizers, the two lessons learned from the conducted analysis are that:

(1) They don't have to worry too early about a small number of submissions since, on the average, half of them will be done during the last day (given the model, $N_{tot}/2$ is attained at time $T - 1.31/\beta$, and β is typically in between 1 and 3), and a reasonable prediction can be made one week earlier.
(2) There is no real point in extending the deadline (at least in terms of number of submissions), unless it is believed that offering more time will end up with better written papers.

While the APP model was clearly relevant for the data considered in [1], the new model turned out to be much better suited to the new data set analyzed here. It is suggested that this is basically due to the different natures of the considered conferences and of the corresponding scientific communities, but this in turn prompts to reconsider from a more critical perspective the issue of universality in people's reaction to a deadline.

More fundamentally, the proposed model is essentially *ad hoc*, thus calling for a more constructive approach that would derive it (or another one, with similar description and prediction power) from well-established basic principles. This is the purpose of current investigations that will be reported elsewhere.

Acknowledgments

Thanks to GRETSI, EURASIP, CMS and Iain Murray (Univ. of Toronto) for accessing the data, and to Eric Bertin (ENS Lyon) for discussions and suggestions.

References

[1] A. Alfi, G. Parisi and L. Pietronero, "Conference registration: how people react to a deadline," *Nature Physics* **3** (2007) 746.
[2] `www.gretsi.org`
[3] `www.eurasip.org`
[4] `www.ieee-sps.org`
[5] P.-F. Verhulst, "Recherches mathmatiques sur la loi d'accroissement de la population," *Nouv. Mm. de l'Acad. Royale des Sci. et Belles-Lettres de Bruxelles* **18** (1845) 1-41.
[6] F.J. Richards, "A flexible growth function for empirical use," *J. Exp. Botany* **10** (1959) 290–300.
[7] C.P.D. Birch, "A new generalized logistic sigmoid growth equation compared with the Richards growth equation," *Ann. Botany* **83** (1999) 713–723.
[8] `www.cs.toronto.edu/~murray/nips_submission_times/`