Société de Calcul Mathématique, S. A.
Algorithmes et Optimisation

Probabilistic methods for stock management

by

Olga Zeydina and Charline Carlier
Société de Calcul Mathématique S. A.
111 Faubourg Saint Honoré, 75008 Paris, France

October 2007

Abstract

The stock is, by definition, the difference between the inputs and the outputs of a good. Both may vary according to complex, non deterministic, laws. We take here the example of the French supply of natural gas, and we show how probabilistic methods may help take a decision about the size of the stock. Both imports and consumption are represented by rather complicated probability laws, depending upon temperature, but the deterministic conclusion is very simple and clear : if the stock falls below a certain level at the end of winter (April 1st), there is a high risk of insufficient supply of gas during the whole year. Neither the inputs nor the outputs are critical : our study shows that the important factor is the level of the stock at that precise date.

The theory presented here is based upon a practical study we made for the French "Direction Générale de l'Energie et des Matières Premières, Observatoire de l'Energie", in 2006. Concrete results are not given (they belong to DGEMP), but this is of no importance here : the theory would apply to other situations as well.

The work presented here is part of the PhD of the first author, prepared at the University of South Brittany (Vannes) under the supervision of Prof. Emile Le Page and Prof. Bernard Beauzamy.

I. Description of the problem

On one hand, we have gas consumption (by the population and the industry) and, on the other hand, imports are made from various countries (the French production of natural gas is now very modest). Legal stocks are maintained, which should cover several days, even in the case of insufficient supplies, but the question arises : are these stocks sufficient ?

The risk occurs if, during a sufficiently long period, the consumption exceeds the imports. Several circumstances may occur : interruption of supplies, excess of consumption due to unusual weather conditions, and so on.

The consumption and import depend on several common factors ; for example, both are in a strong correlation with the temperature. Here, the temperature will occur as an intermediate parameter which will not appear in the final result : this is an interesting point in our method ; see the discussion below.

There is, however, a difference between consumption and imports. The consumption varies from day to day, according to the needs. Conversely, most of the imports are governed by long term contracts, with fixed amounts, and a possible extra import may be decided on the "spot" market, but this has a high price and should be avoided.

II. Mathematical approach

A. General description

Let C_j be the random variable which indicates the French consumption of natural gas on the $j-$th day of the year ($j=1$ is April 1st in our situation : the year starts at the end of the winter season). The law of this random variable is unknown. Similarly, let I_j be random variable which indicates the import on the $j-$th day. The stock on the $j-$th day is by definition :

$$S_j = \left(I_1 + ... + I_j\right) - \left(C_1 + ... + C_j\right)$$

Because of lack of pertinent data, neither the law of the C_j nor the law of the I_j can be approached directly. Only the relationship with temperature can be approached, in a non-precise way. This means that if T is the temperature, the conditional laws of $C_j \mid T$ and $I_j \mid T$ can be computed. After that, the law of T_j, temperature on the $j-$th day of the year can be computed, using historical data. But this leads to a mathematical difficulty, which we now describe.

B. Propagating conditional laws

Assume that T,U,V are random variables, and that the conditional law of $V \mid T$ is unknown, but conversely the laws of both $V \mid U$ and $U \mid T$ are known. This is our case here : we do not know the behavior of the consumption with respect to the day, but we know the consumption with respect to temperature, and we know the temperature for each day (as probability laws).

Then, a natural idea is to define, for discrete variables :

$$P\left\{V=v \mid T=t\right\}=\sum_{u} P\left\{V=v \mid U=u\right\} \times P\left\{U=u \mid T=t\right\} \tag{1}$$

and in the case of laws given by a density :

$$\varphi_{V|T=t}(v)=\int_{u} \varphi_{V|U=u}(v)\, \varphi_{U|T=t}(u)\, du$$

But this "natural" choice implies some hidden assumptions, as we now see.

Indeed, clearly :

$$\begin{aligned} P\left\{V=v \cap T=t\right\} &=\sum_{u} P\left\{V=v \cap U=u \cap T=t\right\} \\ &=\sum_{u} P\left\{V=v \mid U=u \text{ and } T=t\right\} P\left\{U=u \text{ and } T=t\right\} \\ &=\sum_{u} P\left\{V=v \mid U=u \text{ and } T=t\right\} P\left\{U=u \mid T=t\right\} P\left\{T=t\right\} \end{aligned}$$

and therefore :

$$P\left\{V=v \mid T=t\right\}=\sum_{u} P\left\{V=v \mid U=u \text{ and } T=t\right\} P\left\{U=u \mid T=t\right\} \tag{2}$$

Comparing this to (1) shows that the relation (1) will be satisfied only if :

$$P\left\{V=v \mid U=u \text{ and } T=t\right\}=P\left\{V=v \mid U=u\right\} \tag{3}$$

for all t, u, v, which means, in other terms, that the information coming from T, when applied to V, is useless : it is entirely contained in the information coming from U.

In our case, for instance, it means that the consumption depends upon the temperature, but not on the particular day. Two days with similar temperatures (one in spring, one in autumn, for instance) would have the same probability laws for consumption.

This assumption should be checked in practice, upon real data. We insist that the choice of a model such as (1), though it appears as quite natural, in fact contains a hidden assumption which is not so obviously satisfied.

C. End of the construction

For a given day of the year, the laws of all $C_1, \ldots, C_j$, $I_1, \ldots, I_j$ are computed, using the conditional laws deduced from the temperature. But to write explicitly the law of S_j would be quite complicated, so we use a simulation procedure, based upon a Monte-Carlo type method : we produce a large number of scenarios, for each day, both for the inputs and the outputs, using the corresponding laws. For each of these scenarios, we compute the corresponding stock and this gives us the expression of the law of the stock.

III. Activity data and temperature

We use the daily data of consumption of gas for France for 12 months in 2005 and the 7 first months of 2006. We want to link it with the temperature, but which temperature ? Temperature is not uniform all over the French territory. In order to solve this preliminary question, we take the daily temperature for different French cities : Lille, Brest, Strasbourg, Paris, Bordeaux et Marseille. We compute the national average and we compare it with the temperature of Paris. The comparison, which was made over several years, shows that the temperature in Paris is representative of the national average, so we will use the temperature in Paris as a reference.

So we have only 19 months of daily consumption for gas ; if we had more, it would not help us a lot, because habits change permanently.

IV. Dependence between consumption and temperature

A. Deterministic approach

We could try to express the dependence between consumption and temperature by means of a certain function :

consumption = function (temperature),

Here is a general curve, with the temperature in X-axis and with the consumption in Y-axis, for the period which interests us (19 months, 2005-2006).

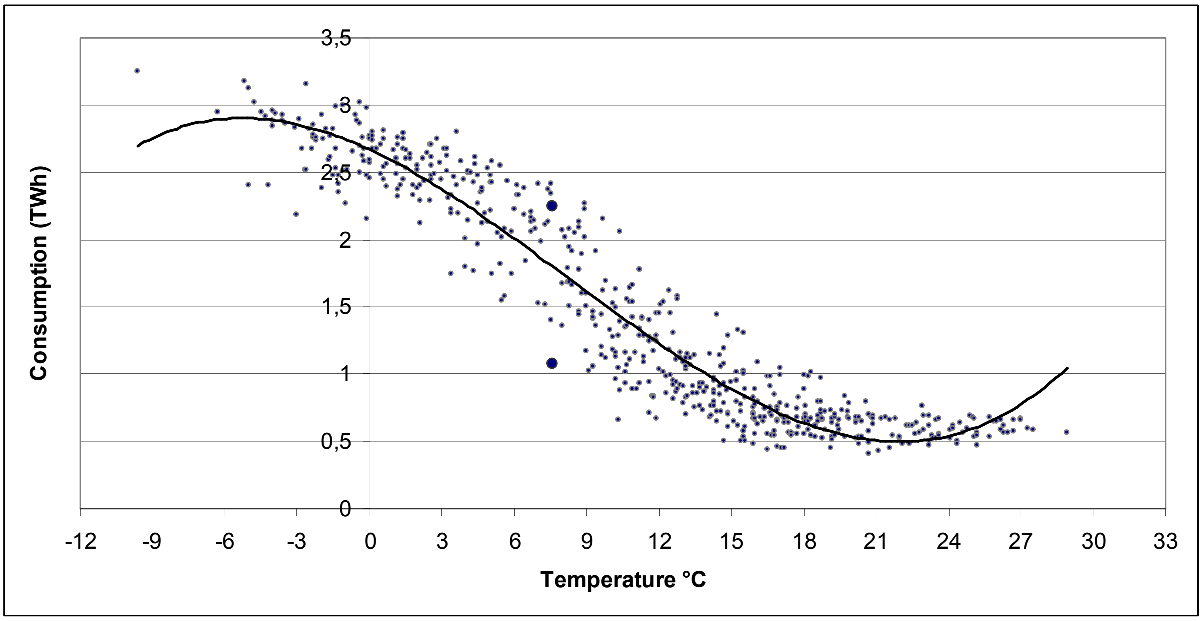

Figure 1 : Adjustment of consumption with temperature

The adjustment is not satisfactory, because of high dispersion of the consumption points, for a given temperature. For example, for the temperature 7.6°C, we see values of consumptions 1.07 TWh to 2.24 TWh and even above.

This remark shows that a deterministic approach, based upon adjustments, is not appropriate and cannot be used in order to substantiate a decision. Much of the information is lost, when we pass from the raw data to the adjustment.

The use of a probabilistic approach, conversely, is more promising, because it allows to retain all available information. The result will be given as a probabilistic law.

B. Probabilistic approach

This approach requires computing tables of conditional probabilities. See [1] for a general introduction to these methods.

As we said, we have 577 data : daily data of consumption and temperatures during 19 months.

The temperatures lie in the interval [-11.7°C; 31.6°C]. We divide this interval into subintervals with width 0.5°C. In this way, we obtain 88 subintervals.

The possible values of consumption lie in the interval [0.406 TWh; 3.250 TWh]. We divide this interval into subintervals with width 0.1 TWh. In this way we obtain 33 subintervals.

We take the first interval of the temperature, [-12°C; - 11.5°C[, and count how many times the recorded temperatures fell in it. Let us call this number T_1.

Then, among these T_1 times, we count the number of times, when the consumption belongs to the first interval [0.4 TWh; 0.5 TWh [, then to the second interval [0.5 TWh; 0.6 TWh [and so on. Let us call these numbers $C_{1,1}, \ldots, C_{1,33}$. This way, we construct a "table of hits", for the couples temperature x consumption. The table of conditional probabilities is obtained by dividing all $C_{1,1}, \ldots, C_{1,33}$ by T_1, so $\sum_j \frac{C_{1,j}}{T_1} = 1$

This way, we obtain the probabilistic distribution of gas consumption when the temperature lies in the interval [-12°C; - 11.5°C[.

We do the same for the second interval of temperature [-12.5°C, - 11°C [and obtain the same way the probability distribution of gas consumption.

Finally, all probability distributions of gas consumptions, conditioned by the temperature, are presented as a 88 x 33 matrix.

Here is an example of probability law for gas consumption, assuming that the temperature belongs to the interval [10°C; 10.5°C[:

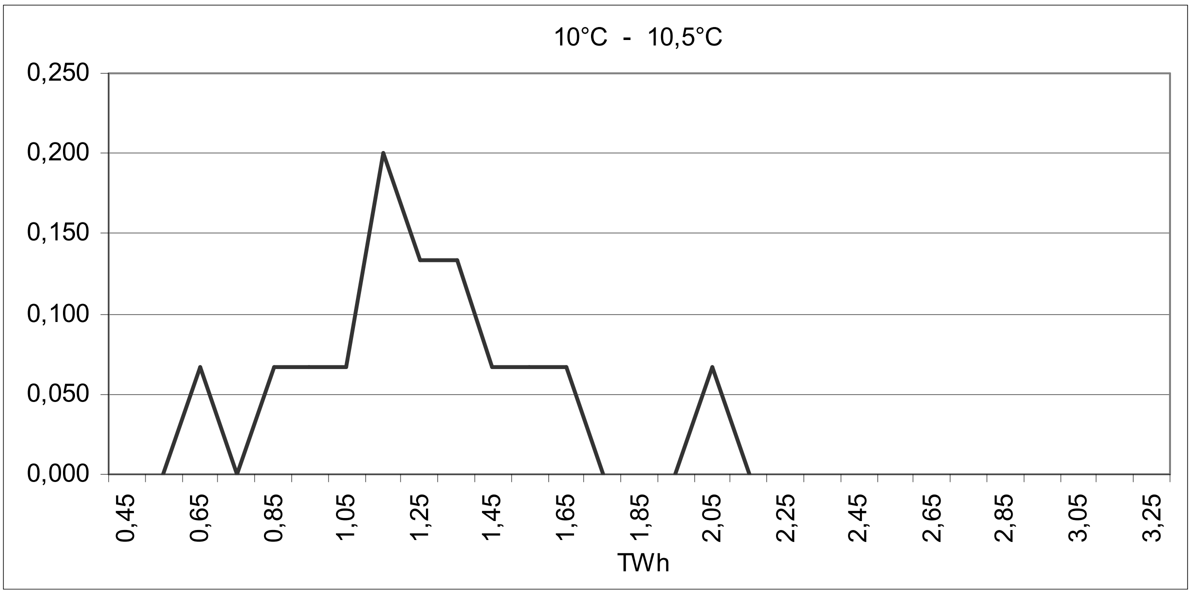

Figure 2 : Distribution of probability for the consumption, when the temperature is in the interval 10°C - 10.5°C

At this point, we have linked consumption with temperature. We need now to build probability laws for temperature. This is much more stable, and can be done using historical data over a long period, in fact from 1900 until now.

V. Probability laws for temperature

We group all data of temperature according to days : we put all data for the 1st of January in the first column and so on. Here is the beginning of this table :

January	1	2	3	4	5	6
1900	7,3	10	7,7	7,2	4,1	2
1901	5,1	5,1	4	-1,1	-3,8	-8,3
1902	7,3	11	10,9	9,1	6,7	5
1903	2,4	8,4	10,8	10	9,9	6,6
1904	-3	-0,9	2,7	4,3	2,2	-0,5
1905	-0,8	-7,2	-5,5	2,7	5,5	7,6
1906	-0,1	2,8	6,3	9,7	11,4	9,6
1907	6,8	6,5	4,8	2,9	2,4	7,1
1908	-0,6	-4,4	-7,2	-5,1	-4,6	-6
1909	-2,5	1,5	2,6	0,4	-3	-2,8
1910	4,5	5,9	5,1	6	4,6	2,8
1911	4,6	1,3	1	0,7	1,4	1,1

Table 3 : Temperature data per day of the year

Using this table, we can construct a probability law of temperature, for each day of the year.

Now we have at our disposal two kinds of probabilistic information : for each day, a probability of consumption as a function of temperature and a probability of temperature. Combining both, we will get the probability of consumption for each day.

Here is the final result for the 1st of January :

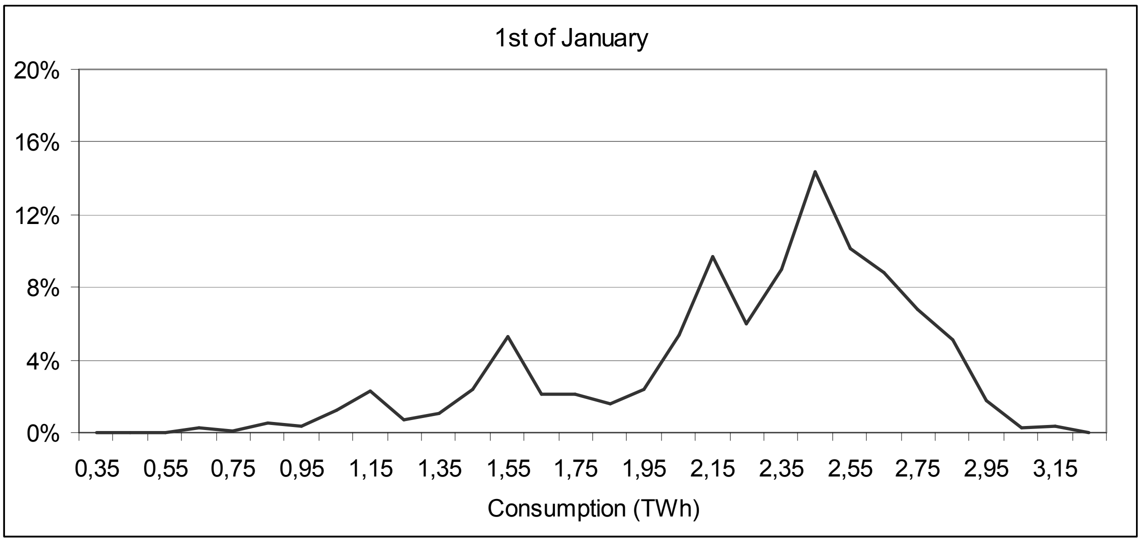

Figure 4 : Probability law for consumption, January 1st

From this graph we conclude that the most probable consumption of gas, for January 1st, is in the range 2.3 – 2.7 TWh. This way of presenting the information is certainly much more efficient than just a single value, or even a confidence interval : no information is lost.

This probability law was established using temperature, but this parameter disappeared from the final outcome ; the temperature was an intermediate help which gave us the opportunity to use "long series", which exist for temperature and not for consumption.

VI. Probability laws for imports

France buys its gas mostly from four countries, plus the possibility to buy each day on the spot market. We take the same period of observation as for consumption (19 months). We build a probability law for each day for all of them using the same principle as before. Here each supplier is considered separately because of the difference in their behaviors. The result is here, for each of the 5 supplies (4 countries and "other"), for each day of the year, a probability law indicating what we buy, on this day, from this supplier. These probability laws are built using the same strategy as before : linking the imports with temperature on 19 months, using long series of temperature, and combining both.

VII. Stocks

The reference starting day is April 1st (end of winter), which represents the end of the peak period for consumption. Starting from the value of the stock (amount of gas we have) on this day, the value of the stock on every day of the year is determined by the difference between the consumption and the imports, between the initial date and this day.

Our computation deals only with the "practical" stock and does not take into account the legal stock which must be kept by law.

For each day, we have probability laws for consumption and import ; we simulate the daily level of the stock combining both.

For the first day of cycle (1st of April) the volume of stock reflects the difference between the import and consumption in that day.

For any day in the year k $(k = 1,\ldots,365)$ the volume of stock S_k for the k-th day is:

$$S_k = \left(I_1 + \cdots + I_k\right) - \left(C_1 + \cdots + C_k\right) \tag{4}$$

where I_k is the import for the k-th day and C_k is the consumption for the same day.

Here we make an assumption about independence between the data from one day to another ; see the discussion below.

In order to generate a set of possible values for S_k, we use a Monte Carlo method : we throw independently random values for each day for the consumption and the import, according to their respective probability laws.

We simulate 10 000 scenarios, for each day : both consumption and import. By difference, we obtain, for each day, 10 000 scenarios for stock. This allows us to build the probability law for the stock of each day of the year.

The need for a Monte-Carlo method is due to the fact that to build explicitly the probability law for the stock, each day, using formula (1), is impossible.

Here is the example of the probability law of the stock for March 31st, end of the cycle :

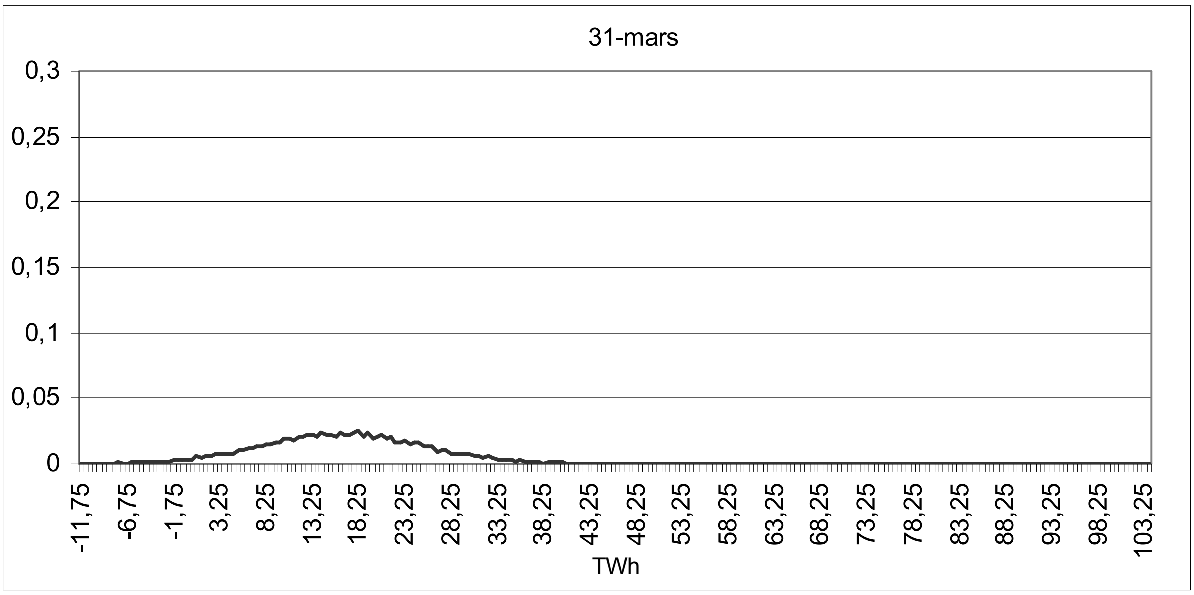

Figure 5 : Probability law for the stock, March 31st

The volume of the stock S_k can be positive or negative. In our case, we are interested by the probability to obtain a negative stock at any time of the year. The most dangerous periods for such an accident are the beginning and the end of the cycle, because these are the times where the stock is naturally low.

This method indicates of course what is the probability to have a negative stock at any time : these are the scenarios, among the 10 000, which meet a negative stock. We found out that if the stocks are low on April 1^{st} (and we gave a precise value), the probability of an accident is high (and we gave a precise value for this probability).

We computed the minimal value of the stock on April 1^{st}, so that the probability of accident remains smaller than 0.05 (for instance). This leads to a concrete recommendation.

We observed that an interesting conclusion is that the total amount of supplies is sufficient : we do not need to buy more, but since the consumption is highly variable, we need to increase our stocks.

VIII. Discussion

The method adopted here is intermediate between two approaches, which are :

– Purely statistical

One does not try to understand the physical phenomena (for instance the influence of temperature, in our case), but relies only upon statistical data, namely the values of the stock for various days. This approach is here completely impossible, since data older than 3 or 4 years have very little value. So we would be left with 2-3 data per day.

– Completely deterministic

One tries to understand as much as one can the underlying laws, both for imports and for consumptions. But usually these laws are quite complicated, and even after careful study a part of randomness remains : this is typically the case here, since temperature cannot be predicted.

So our approach is in between : it cannot be purely statistical, but it includes only a very limited amount of physics (namely, temperature for one day, in our case). Perturbations "around" the chosen physical process are treated as random variables : one does not try to go further, in terms of explanation.

But in fact, one might go a little further. In our treatment, the probability laws for consumption, each day, are built independently, and so are the probability laws for temperature. This is not completely correct : the temperature of a given day is conditioned by the temperature of the previous ; you do not have the same probability to be above 10°C if the day before you were at 5°C and if you were at 12°C. This could be taken into account by building several probability laws, for the temperature of a given day : one, for example, if the temperature on the previous day is between 0 and 3°C, one if it is between 3°C and 6°C, and so on.

This work is part of the "Robust Mathematical Modeling" program, carried over by SCM and several institutions. See http://www.scmsa.com/robust.htm for further details about the program.

IX. Comparison with existing work

There are many situations, such as [3], where one tries to optimize a stock under various constraints. But in this work, the probability laws are assumed to be known. In [2], the same holds in the financial sector : a probabilistic approach is presented, using well-defined laws. In our work, on the contrary, a large part of the task is devoted to the construction of these laws, from real-life data. In the work [4], which deals with stocks of tuna fish, the author uses arbitrary probabilistic laws (log-normal, and so on), but wonders precisely about the influence of such a choice upon the results.

References

[1] Bernard Beauzamy : Méthodes probabilistes pour l'étude des phénomènes réels. Editions de la SCM, 2004.

[2] Manak C. Gupta : Money Supply and Stock Prices: A Probabilistic Approach *The Journal of Financial and Quantitative Analysis*, Vol. 9, No. 1 (Jan., 1974), pp. 57-68

[3] Warren H. Hausman, L. Joseph Thomas : Inventory Control with Probabilistic Demand and Periodic Withdrawals *Management Science*, Vol. 18, No. 5, Theory Series, Part 1 (Jan., 1972), pp. 265-275

[4] Tom Polacheck : Quantification of Uncertainty in Fishery Stock Assessments: Statistical Challenges in the Provision of Management Advice for Southern Bluefin Tuna, CSIRO Marine Research, GPO Box 1538, Hobart, Tasmania 7001, Australia (preprint).