Research Department

Technical Memorandum No. 336

End to End Ensemble Forecasting: Towards Evaluating the Economic Value of the Ensemble Prediction System

by

L.A. Smith[1,2], M.S. Roulston[1,2] and J. von Hardenberg[1,2]

[1] Centre for the Analysis of Time Series, London School of Economics, U.K.
[2] Pembroke College, Oxford University, U.K.

April 2001

Abstract

This report assesses the value of "end to end" forecasting of economic variables that have a weather dependent component using the ensemble prediction system (EPS) of the European centre for medium range weather forecasting (ECMWF) [*Molteni et al.* 1996]. Decision making scenarios in which the outcome is a continuous variable (*e.g.* demand) and in which the decision is also a continous variable (*e.g.* production) are considered. Making binary choices (*e.g.* to salt or not salt potentially icy roads) is well described by cost-loss analysis [*Richardson* 2000]. The current report aims to move away from the cost-loss ratio toward questions of "how much?" rather than questions of "whether or not?".

The basic concepts of utility maximization in decision making are outlined with an idealized example of bagel sales. These concepts are then applied to the problem of forecasting electricity demand in 12 cities around the world. They are also applied to forecasting potential wind generation at a site in the U.K. The period from January 1999 to October 2000 was used to test the value of the ECMWF ensemble prediction system to decision makers in these two scenarios. It was found that the ECMWF EPS yields more optimal decisions in both scenarios. The results indicate that it is *essential* to treat all deterministic forecasts (control or ensemble members) as probabilistic forecasts by attempting to estimate the error distribution about each forecast. It was also found that the 51 member ensembles are not distinguishable from a "singleton ensemble" distribution constructed by adding historical errors to the best guess (high resolution control) forecast on approximately two thirds of the days in the period. However, on the days on which the ensembles are distinguishable on average they outperform the best guess forecast combined with the historical errors. A crude method for estimating the size of the ensemble required to increase the number days on which the ensemble would be distinguishable was applied to the forecasts in the test period. It was found that for over 95% of ensemble forecasts to be distinguishable ensembles of *at least* about 100 members would be required.

1 Introduction

In this report we investigate the value of probabilistic weather forecasts to weather dependent businesses. Previous work has studied the value of weather forecasts in binary decision making scenarios (*e.g.* whether to grit roads or harvest crops). The value of ensemble forecasts in binary decision situations has been clearly illustrated by Richardson [2000]. We focus on decision making scenarios in which both the outcome and the decision take on a continuous range of values (*e.g.* the demand for, and production of a particular good).

The paper is structured as follows. In the next section we illustrate our approach with a simple example. We show how probabilistic forecasts of demand of a good can be used with a utility maximizing approach to make production decisions. The importance of treating the best guess forecast as a probability distribution is stressed by this example. In this section, we assume a perfect model and a perfect ensemble, that is the ensemble members are chosen from the same distribution as truth then evolved under a perfect model [*Smith* 1995, *Smith* 2000]. The section concludes by taking the discussion to one showing the expected utility of an ensemble forecast as a function of the profit margin of the user.

Section 3 considers end to end forecasts of electricity demand. Here we use empirical relationships between electricity demand and temperature to determine ensemble forecasts of demand for 12 cities using the operational ECMWF forecast system. The goal of end-to-end ensemble forecasts is to translate uncertainty in the current weather all the way to future uncertainty in the quantity of interest to the user, in this case electricity demand. This example illustrates several additional features, including the fact that, the map between temperature and demand is not deterministic (other things influence electricity demand in addition to temperature). This aspect of the forecast is easily included in the end to end approach. We consider both a pseudo-perfect model scenario and an operational scenario, using station data independent of the forecast for calculating synthetic electricity demand.

Section 4 considers end to end forecasts of electricity generation from a windfarm. In this case the approximately cubic dependence of potential generation on windspeed makes windspeed forecasts very valuable for optimizing decisions.

End-to-end forecasts provide a much more direct measure of the economic value of ensemble forecasts than existing methods. They also provide a framework within which the economics of resource allocation (*e.g.* larger ensembles or higher resolution) can be evaluated. Questions of where these calculations are best performed are beyond the scope of this report, yet exploiting the value of the end-to-end approach will require users, or their agents, to interact with the ensemble forecasts more closely than is currently common.

2 A Perfect Example: Bagel Sales

Economic decision theory centres around the concept of utility (or von Neumann-Morgenstern utility [*von Neumann and Morgenstern* 1944, *Johnson and Holt* 1997]). That is, it is assumed that users can rank possible outcomes in order of preference by assigning a "utility" value to each outcome.

Consider a situation where the outcome is a continuous variable, y. The user's decision can be represented by a continuous variable, x. The user's utility for a particular decision and outcome is $U(x,y)$. If the probability distribution function (PDF) of y is $p(y)$ then the expected utility for a given decision, $E[U](x)$, is

$$E[U](x) = \int_y U(x,y)p(y)dy \tag{1}$$

The user can then make the choice X that maximizes their expected utility

$$E[U](X) = \max_x (E[U](x)) \tag{2}$$

To illustrate the above consider the following simple example. A bagel shop makes fresh bagels every morning for sale that day. Each morning they must decide how many bagels to make. Let B be the number of bagels produced. During the day there will be a demand for D bagels. The cost price of each bagel is CP and the sale price of each bagel is SP. Let the utility of the bagel shop be equal to profits. The utility is thus given by

$$\begin{aligned} U(B,D) &= -B \cdot CP + D \cdot SP \quad D \leq B \\ U(B,D) &= -B \cdot CP + B \cdot SP \quad D > B \end{aligned} \tag{3}$$

If the bagel shop has a probabilistic forecast of bagel demand for a particular day, $p(D)$, it can choose the value of B that maximises its expected utility given by

$$E[U](B) = \int_D U(B,D)p(D)dD \tag{4}$$

Note that in this example it has been assumed that the utility is equal to profits. This is the case for a risk neutral user. In reality most users are risk averse which results in a utility function which is a concave function of gross profits. (*e.g.* the first 10,000 GBP they earn is worth more than the second 10,000 GBP.) This risk aversion is partly psychological but is also a result of nonlinear taxation structures. Net profits or personal income are generally a concave function of gross profits or income. (*e.g.* the tax rate on the first 10,000 GBP earned is lower than the tax rate on the second 10,000 GBP earned.)

In this simple example the aim of the café is to maximise its daily profit. One could also argue that running out of bagels incurs a cost due to lost customer loyalty. Such a cost, if quantifiable, could be built in to the bagel shop's utility function but such complications are not considered in this example. Underproduction penalties are relevant in the next example.

For the purposes of an example, assume that the demand for bagels from the café are determined by the temperature, cloud cover, and precipitation on the day [1]. Given a model relating these weather variables to bagel demand each member of an ensemble of weather forecasts can be converted into a member of an

[1] Empirically, there is a relationship with these meteorological variables, as well as day of the week and other confounding variables which will be suppressed here for the sake of clarity. In practice, forecasts would also be conditioned on these variables if sufficient empirical data is available.

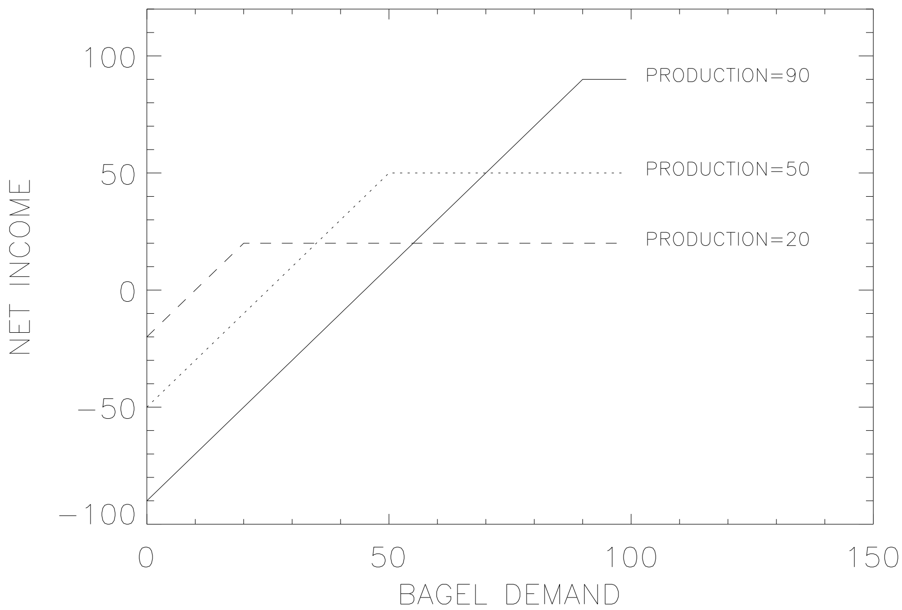

Figure 1: The net income of a bagel seller as a function of demand. Income increases linearly up to the point at which the entire stock is sold after which it is flat. The three curves reflect three differ orders, 20 (dashed), 50 (dotted) and 90 (solid). The profit margin on each bagel is 100%.

ensemble of bagel demand forecasts. That is, each individual weather forecast may correspond to a PDF of bagel demand. The profit (or loss) made on a given day depends on the number of bagels sold, up to a maximum determined by the number of bagels in stock. This is shown in Fig. 1 where the different curves correspond to different levels of stock at the beginning of the day; the larger the number of bagels ordered the larger the number of sales required to breakeven, but also the larger the potential profit due to selling out. Given an ensemble forecast for bagel demand how many bagels should be ordered to maximize expected profit?

Bagel orders based on climatology provide a baseline that any active forecast strategy must beat. In our example, the climatology of weather variables translates to a distribution of bagel sales (a bagel climatology) as shown in Figure 2a.

Given this climatological distribution, each of the production levels shown in Figure 1 then translates to a probability density function (PDF) of profit. These are shown in Figure 4. Obviously, the higher the production, the greater the maximum possible profit (due to selling the entire stock), and also the greater the potential loss. These graphs are for a profit margin of 100%, so that the maximum profit is simply minus the maximum loss. The nonlinearity of the utility function means that the PDFs are not symmetric about zero; the expected value is not the midpoint of the range. If the goal is simply to maximise the expected profit over a season, then the ideal production level corresponds to the distribution which has the maximum expectation.

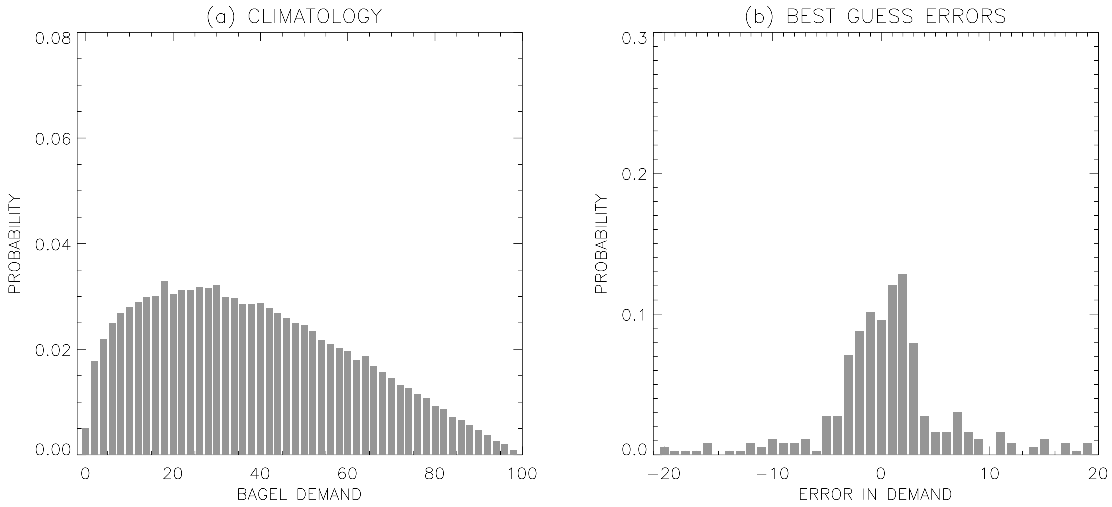

Figure 2: The two probability distributions used in this section (a) climatological distribution of bagel demand, (b) historical errors of the best guess forecast.

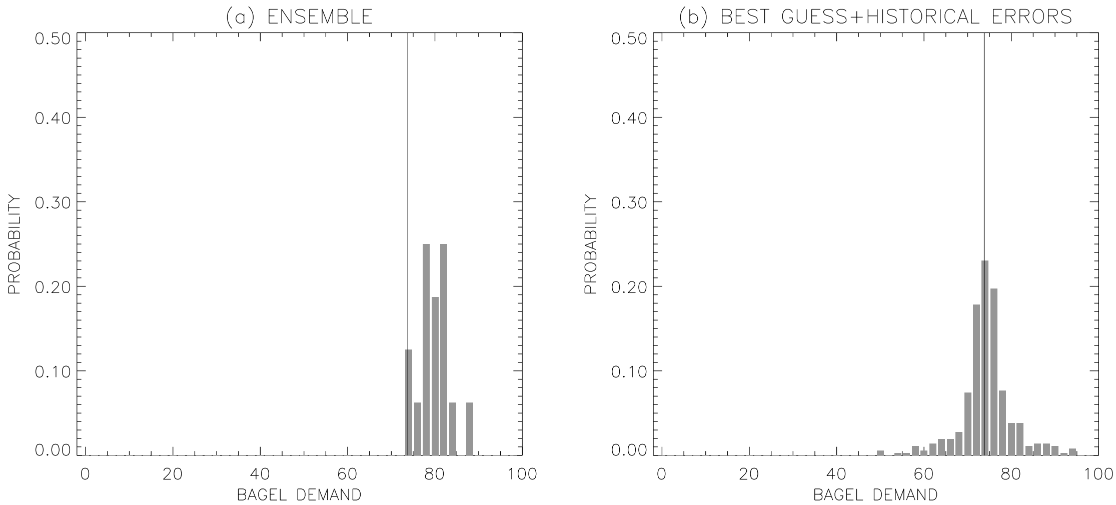

Figure 3: (a) The bare ensemble distribution for one forecast. (b) The best guess forecast with historical error distribution added. This is the "singleton" ensemble forecast. The vertical line represents the best guess forecast in both panels.

The crucial point to note is that if the utility is a nonlinear function of demand then **the expected value of the utility will, in general, not be the utility associated with the expected demand.** It is this fact which makes probabilistic forecasts essential for rational decision making.

The single, best guess forecast, can be converted into a probabilistic forecast by adding the historical distribution of errors of the best guess forecast. This means that the forecast PDF (for a given lead time) always has the same shape but it translates depending on the best guess forecast for that day. The best guess plus historical errors forms a "singleton" ensemble. This type of forecast is compared with a bare ensemble forecast in Fig. 3.

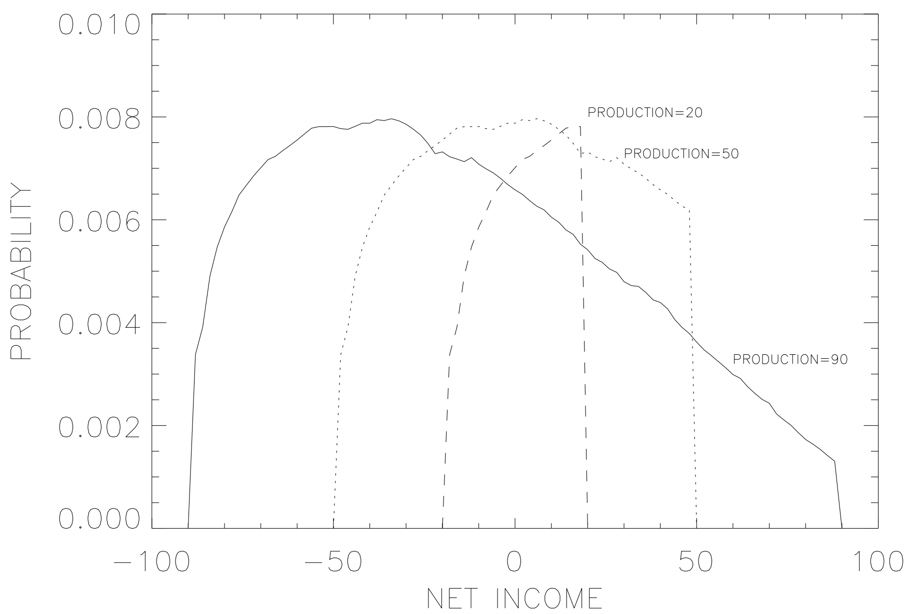

Figure 4: The probability distributions based on climatology for each of the three production levels shown in Figure 1.

Just as one should clothe a single, best guess, forecast to form a singleton ensemble, an ensemble forecast should *not be treated as a sum of delta functions*; instead some uncertainty distribution should be added to clothe each member. For the best guess forecast the relevant distribution is that of historical errors (possibly stratified by season). It is not clear how to best clothe an ensemble forecast with $N > 1$. Options include adding a gaussian of empirically determined variance or adding the error distribution of the best member from the N member ensembles. Each of these two methods is consistent with a perfect ensemble as $N \to \infty$. Several methods are discussed in the next section, but the ideal method and how to generalize to multivariate forecasts remains an open question.

While goods such as bagel typically have high profit margins this scenario could apply to any good with a weather dependent demand. We will, therefore, consider a range of profit margins, where the profit margin is simply the ratio of the profit on each unit to the wholesale cost. We consider only positive profit margins, that is the retail cost must exceed the wholesale cost. Clearly if the profit margin is very large, it pays to overstock while if profit margins are small it's better to err on the side of underproduction.

Figure 5 shows the value of using a forecast system above relative to using the climatology as a function of profit margin. Each line shows the ratio of the profit made using some forecast system to the profit made by always ordering the optimal value based upon climatology. Hence the climatology as a forecast corresponds to the horizontal line at 1 for all values of the profit margin. One method of forecasting sales is simply to place an order equal to the demand corresponding to the best guess forecast. This is the dashed curve in Fig. 5. At profit margins below 5% this strategy underperforms climatology and users with very tight profit margins would incur some pretty spectacular loses. This result again underlines the fact that probabilistic forecasts are essential for rational decision making. The best guess forecast can be converted into a probabilisitic forecast by adding the distribution of historical errors to create a "singleton" ensemble. The performance of this strategy is shown by the dash-dot line in Fig. 5. This

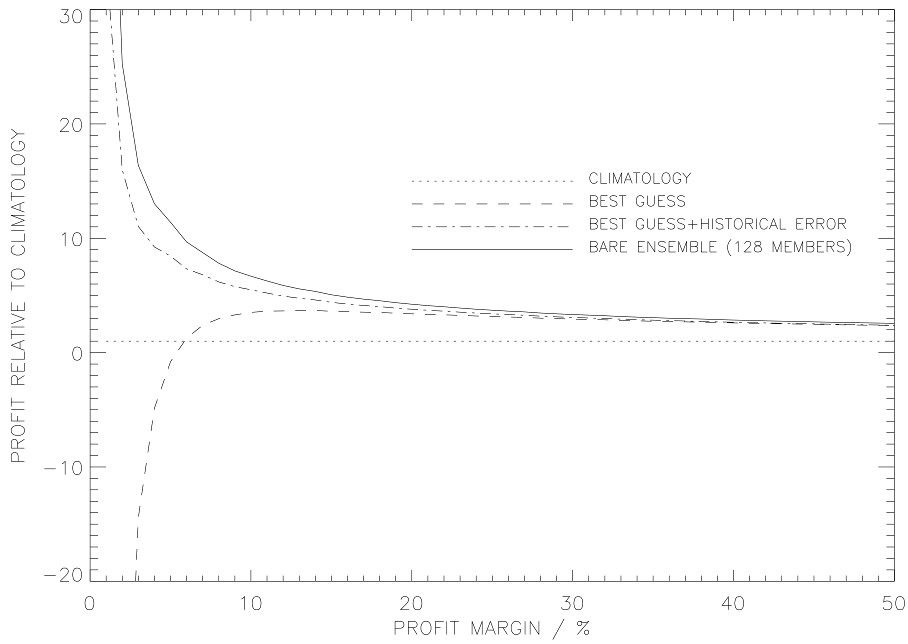

Figure 5: Expected value added from using a forecast rather than climatology as reflected by relative profit as a function of profit margin. The curves show the performance (relative to climatology) of using the best guess forecast as a delta function, the best guess forecast with an historical error distribution and a bare ensemble of 128 forecasts.

immediately leads to a massive improvement over the strictly deterministic decision making approach. This strategy outperforms climatology at all profit margins and indeed, becomes increasingly valuable as the profit margin decreases. While this strategy does require an increase in the sophistication of the decision making approach used it is clearly worth the effort for users on tight profit margins. The solid curve in Fig. 5 is the ratio of the expected profit using a perfect ensemble forecast with 128 members. This strategy outperforms the singleton ensemble. At a profit margin of 5% this strategy would lead to around a 30% increase in profit relative to the best guess forecast with the historical error. This is a substantial improvement in performance, especially since the marginal effort required to switch between using the singleton ensemble and the ensemble prediction system will typically be quite low for end users of forecasts.

Why does using the best guess forecast lead to heavy losses when profit margins are tight? This is seen by comparing the demand with the production when the profit margin is a mere 3%. This comparison is made in Fig. 6. A user using climatology simply produces the same number of bagels each day. A user using the 128 member ensemble produces close to the actual demand but over the period shown always errs on the side of underproduction. This is not suprising since their profit margin is very tight. The user who produces the number of bagels associated with the best guess forecast occasionally overproduces bagels. The impact that this overproduction has on profits can be seen in Fig. 7. Notice that the days on which the bagel production exceed demand correspond to losses. Users basing production decisions on the ensemble forecast suffer no major losses because they rarely overproduce. If the best guess forecast is converted to a probabilistic forecast by the addition of historical errors its performance improves. Losses now become much less frequent although both the best guess forecast with historical errors and the climatology are more likely to lead to overproduction then the bare ensemble.

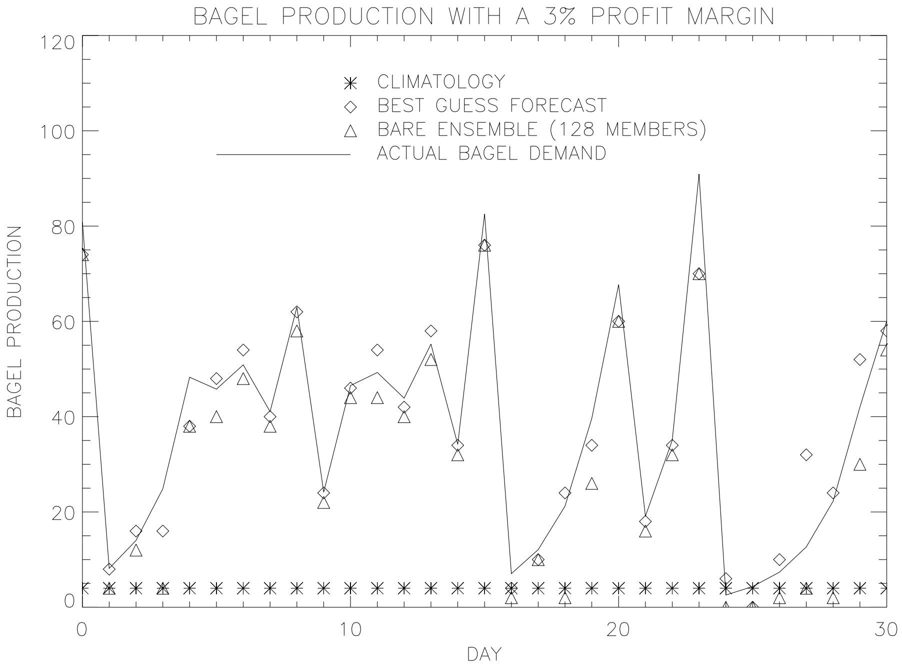

Figure 6: Optimal bagel production for 4 different forecasting schemes compared with observed bagel demand. The profit margin is 3%. The tolerance for overproduction depends on the details of the net income function shown in Fig. 1

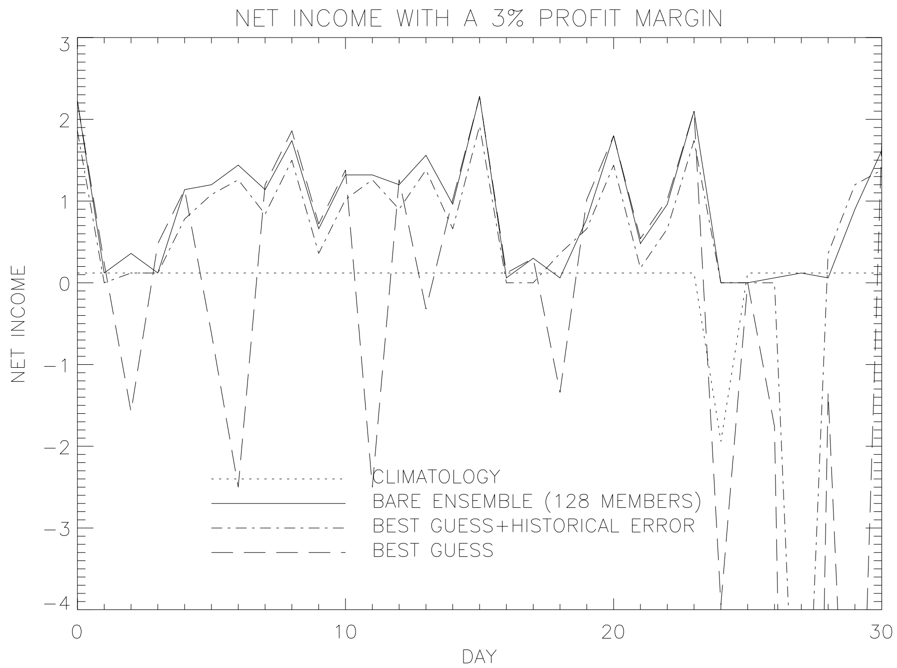

Figure 7: Time series of profits made for each forecast with 4 different forecasting schemes.

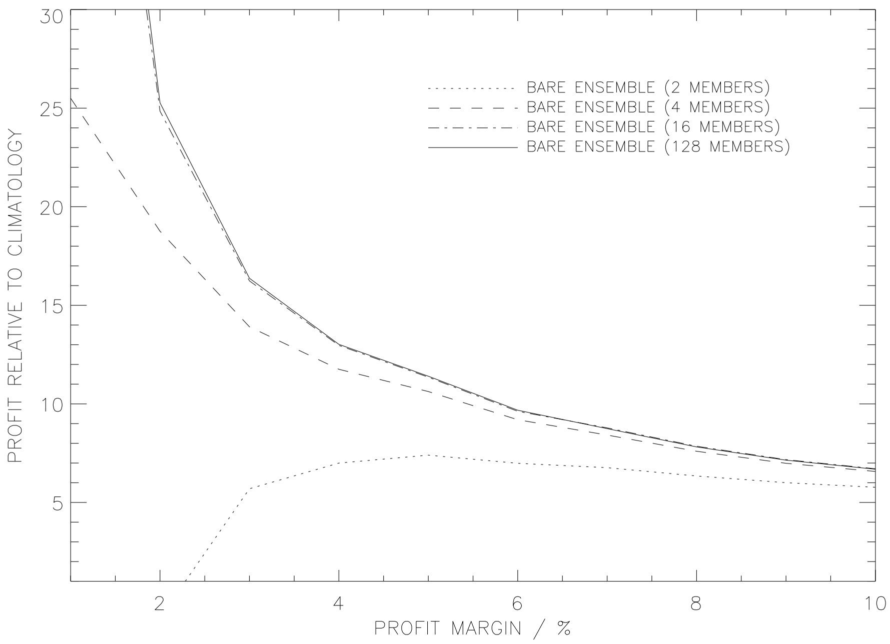

Figure 8: Expected value added from using a forecast rather than climatology as reflected by relative profit as a function of profit margin. The curves reflect using bare ensembles with 2, 4, 16 and 128 members. "Bare" refers to forecasts that have not been "dressed" with individual error distributions.

An examination of Fig. 7 indicates how the use of the forecasts impacts the weather risk to which the bagel shop is exposed. If the climatological forecast is used then the shop incurs heavy losses when bagel demand is low. Using a good forecast the bagel shop can greatly reduce its losses on such days by not making many bagels (*e.g.* days 25 to 28 in Fig. 7). The bagel shop's profits are still greatly reduced however, indeed once the shop's fixed costs are included they could still make a loss on days of low bagel demand, even if production exactly matched demand. Thus good demand forecasts enables the shop to reduce its weather risk exposure but not to eliminate it. Buying weather derivatives would enable the shop to hedge this residual risk but *a bagel shop that did not use the forecasts would have to purchase a greater amount of weather risk cover than one that did.* Only on time scales where end-to-end forecasts have no more skill than climatology is a pure hedging strategy justified. On shorter time scales the forecast should be used unless the price of the forecast is prohibitive. We now investigate how ensemble size affects decision making. Figure 8 shows the performance of ensembles with 2, 4, 16 and 128 members. The performance of the two member ensemble follows a similar pattern to the single best guess forecast. Merely adding another two forecasts, however, leads to a dramatic increase in profits at low profit margins. Increasing the ensemble size to 16 members leads to a further increase but then no improvement is obtained even when the ensemble contains 128 members. Thus, in this particular case, there would seem to be little benefit in generating bare ensembles larger than 16 members.

The information contained in a perfect ensemble increases as the lead time decreases. Figure 9 shows how a the distribution of a perfect ensemble converges to a deterministic forecast as the forecast horizon falls to zero. The details of the collapse of the PDFs illustrated in Fig. 9 will vary from day to day. Here we note that increasing the size of the ensemble will increase the resolution of the PDFs but not reduce their spread. While decreasing uncertainty in the initial condition can reduce the spread. In the perfect model scenario, the end-to-end approach allows us to evaluate the relative value of incresing the ensemble

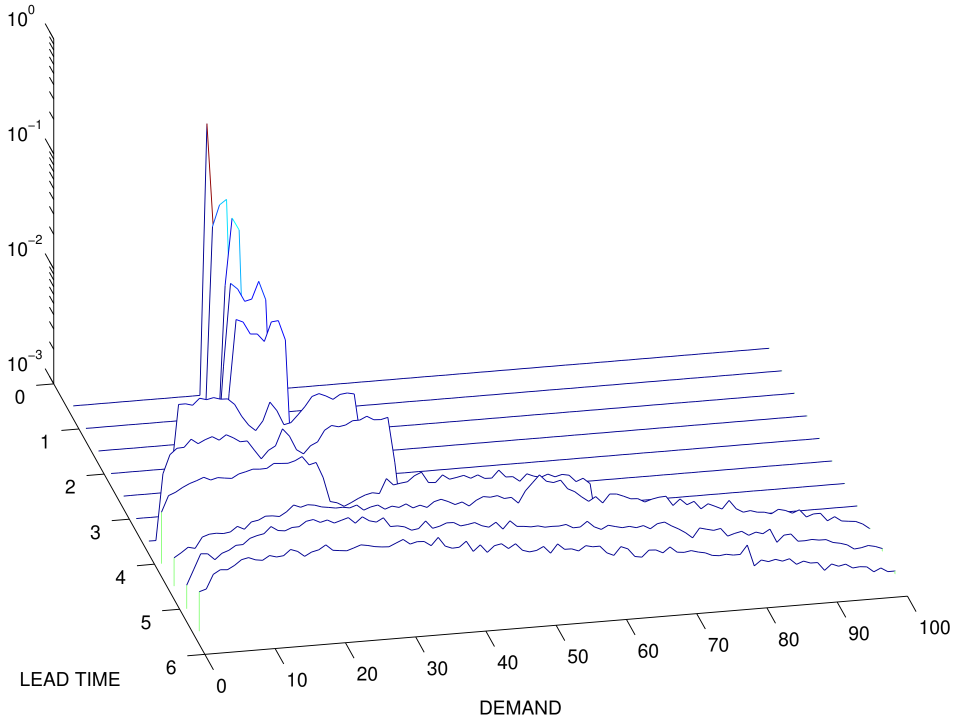

Figure 9: Convergence of a perfect ensemble forecast as a function of lead time.

size versus decreasing the analysis uncertainty. Given an imperfect model (or models) such analysis is complicated by the option of allocating resources to model improvement. In section 3 we return to the question of optimal ensemble size.

3 Electricity Demand

Having outlined the key concepts behind decision theory in a weather dependent industry in the previous section we will apply these ideas to the case of electricity.

The demand for electricity is highly weather dependent. There is increased demand for heating on cold days and for refrigeration and air-conditioning on warm days. Furthermore, electricity is difficult to store so supply and demand must be matched on very short time scales. The lack of storage leads to very volatile prices. For example, U.S. power usually trades at around 30-60 USD/MWh, however, due to high temperatures, during June 1998 the spot price approached 1000 USD/MWh on the East Coast and more than 7000 USD/MWh in the Midwest [*Locke* 1999]. While the use of energy and weather derivatives can help companies deal with such volatility good demand forecasts can be used to improve decision making reducing overall risk.

In the U.K. the New Electricity Trading Arrangements (NETA) will be in place before the end of next year [*DTI* 2000]. Under these arrangements power generators, suppliers and traders will trade electricity on exchanges and through bilateral contracts. These contracts will typically be written far ahead of the time period which they cover. In practice the amount of power exported by generators and imported by suppliers will not exactly match the amounts specified in the contracts. The task of balancing supply and demand in near real-time will falls to the System Operator. At the so-called "gate closure", 3.5 hours before the time of delivery generators and suppliers will have submitted offers and bids for providing power or demand different to that specified in their Final Physical Notificiations (FPN). The System Operator must then accept offers and bids to attempt to balance supply and demand. The details of NETA are complex and the precise way in which electricity market will operate under NETA is yet to be seen. For our study we will use a toy model of an electricity trading arrangement.

Consider a situation where an electricity supplier (a supplier purchases power from generators and supplies it to consumers) can buy a futures contract from a generator to provide a fixed amount of electricity, PP, at a particular price in 4 days time. Let the cost price of this electricity to the supplier be CP per unit. The supplier is obliged to purchase this power from the generator at this price, even if it exceeds the actual demand, AD, during the contract period. The supplier then sells this electricity to its customers at a price of SP per unit. If during the contract period demand, AD, exceeds the amount of pre-purchased power the supplier must buy extra electricity at the *imbalance settlement price*, ISP. The imbalance settlement price can greatly exceed the cost price of pre-purchased electricity thus the supplier would ideally want to match pre-purchased power to actual demand as closely as possible. If we further assume that the power supplier is risk neutral then the supplier's utility function is given by

$$\begin{aligned} U &= -CP \cdot PP + SP \cdot AD \qquad AD \leq PP \\ U &= -CP \cdot PP - ISP \cdot (AD - PP) + SP \cdot AD \qquad AD > PP \end{aligned} \tag{5}$$

In our study we set $CP = 1.0$ and $SP = 1.5$ and considered the effects of changes in ISP. Figure 10 shows the utility as a function of actual demand for three different choices of pre-purchased electricity and $ISP = 5.0$. Having setup our toy electricity market we assumed a temperature dependence of demand. Two forms of the temperature dependence were considered. The first was a deterministic-linear relationship in which demand, AD, in megawatts was described by

$$AD = 5000 - 100T \tag{6}$$

where T is the temperature in degrees Celsius. The second temperature-demand relationship was a more

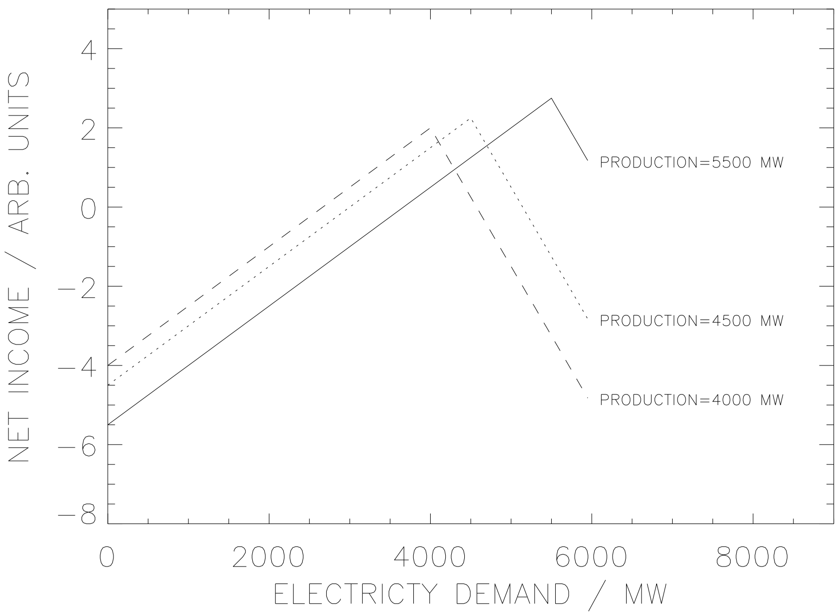

Figure 10: The utility of a risk neutral electricity supplier for an advance purchase price of 1.0, a selling price of 1.5 and an imbalance settlement price of 5.0.

realistic function given by

$$AD = 5945 - 23.12T - 1.47T^2 + 0.0193T^3 + 250 \cdot \mathcal{N}(0,1) \tag{7}$$

where $\mathcal{N}(0,1)$ is a normal random variable with zero mean and variance unity. This temperature dependence is shown by the curve in Fig. 11. The U-shaped dependence shown is the result of increased demand for heating at low temperatures and increased demand for air-conditioning/refrigeration at higher temperatures. Such dependence is observed in the electricity demand for U.S. cities. The points in Fig. 11 are synthetic demand data generated from observed temperatures in Washington DC (WMO station #72405) using Eq. 7.

In the scenario described above how valuable would ECMWF forecasts be in making the decision of how much electricity to pre-purchase?

As in the bagel example, given a probabilistic forecast for demand the user can choose the pre-purchase level so as to maximize their expected utility. In our study we considered the following methods for constructing the probabilistic demand forecast:-

- **Climatology**: A 14 day interval about the forecast date from a 30 year synthetic climatology of daily temperature was used to construct a PDF of demand.
- **Best guess forecast**: The best guess ECMWF forecast was treated as a δ-function, this leads the user to pre-purchase the exact demand associated with the best guess forecast.

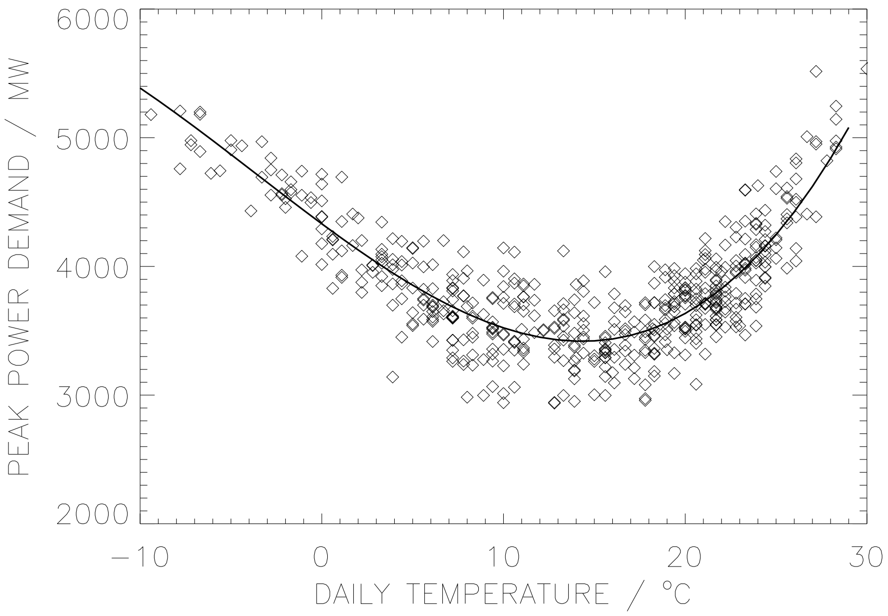

Figure 11: Synthetic demand data constructed using observed temperatures in Washington DC (WMO station #72405).

- **Best guess forecast with historical error**: The forecast errors for the one year were added to the best guess forecast to generate a **"singleton"** ensemble forecast.
- **Bare Ensemble**: The 51 members of the ECMWF ensemble were treated as δ-functions to generate a combined PDF.
- **Ensemble with empirically determined gaussian errors**: A gaussian distribution was convolved with the bare ensemble distribution. This generates a "dressed" ensemble forecast.

The variance, σ^2, of the empirically determined gaussian errors on each the N_{eps} ensemble members was chosen to minimize the information deficit, IGN, of someone in possession of the forecast over a learning period of N_{days} [*Roulston and Smith* 2001]. This deficit is given by

$$IGN = -\sum_{t=1}^{N_{days}} \log\left[\sum_{i=1}^{N_{eps}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(x_{it}-a_t)^2}{2\sigma^2}\right)\right] \tag{8}$$

where x_{it} is value of the relevant variable for the ith member of the ensemble for day t and a_t is the actual value of the variable on that day. Comparison of these and other methods of dressing ensemble forecast is discussed by Roulston and Smith [2001]. We stress that treating the best guess or the ensemble forecasts as delta functions allows a correct assessment of the value of the forecasts.

To estimate the theoretical maximum value of the ECMWF ensemble forecasts a contrived "perfect ensemble" experiment was performed. In this experiment the "true" demand was calculated from a temperature selected at random from the ECMWF ensemble instead of using the observed temperature.

Figure 12 shows the profit of users using different demand forecasts as a function of the imbalance settlement price paid for any shortfall in supply. The electricity demand for each day was calculated by selecting one of the ECMWF ensemble members for Washington D.C. and using the relationship given by Eq. 6 to calculate a demand. Because truth was selected directly from the ensemble and because the temperature-demand relationship had no stochastic terms the estimated ensemble member gaussian error had zero variance—*i.e.* it was the same as the bare ensemble and so is not shown in Fig. 12. The case shown in Fig. 12 can be thought of as showing the potential value of the ensembles under perfect conditions. The temperature ensembles were perfect because truth was picked from the ensemble and furthermore the demand was a linear-determinisitic function of the temperature. Figure 12 demonstrates several points:
(i) As in the bagel example treating the best guess forecast as a deterministic (δ-function) forecast lead to very poor decisions.
(ii) Converting the best guess forecast to a probabilistic forecast by adding historical errors gave a substantial improvement.
(iii) The bare ensemble provided an even larger improvement. When the imbalance settlement price was 20 times the cost price the bare ensembles lead to profits 35% higher than climatology and 20% higher than using the best guess forecast with the historical error.

The perfect ensemble experiment was then repeated using the more realistic temperature-demand relationship given by Eq. 7. The results are shown in Fig. 13. In this case it can be seen that the bare ensemble underperformed relative to just using the best guess forecast with the historical errors. When the estimated ensemble error was added to the ensemble members the ensemble outperformed all the forecasts. Since, in this case, the "true" temperature was selected from the ensemble the error on the ensemble members was solely due to the errors in the model that mapped temperature to demand, *i.e.* the stochastic term in Eq. 7. The empirically determined standard deviation of the ensemble error was 258 MW, which is about what would be expected from Eq. 7. Also, in this case, the assumption that the ensemble errors are gaussian was exact.

Figure 13 indicates that even when the temperature-demand relationship is realistic perfect ensembles can give substantial improvements in the business performance of power suppliers.

Having determined the value of perfect ensembles the value of the actual ensembles were evaluated by basing the synthetic demand on the *observed* temperature.

Figure 14 shows the results when the temperature-demand relationship was linear-deterministic (Eq. 6). The pattern of behaviour in Fig. 14 is similar to that in Fig. 12. Using the best guess forecast deterministically was a recipe for disaster. Adding historical errors to this forecast improved things substantially. The bare ensemble provided yet more improvement and attempting to estimate the error distribution around each ensemble member provided further improvement.

The temperature-demand relationship of Eq. 6 is highly idealized. Figure 15 shows the results obtained using the more realistic relationship of Eq. 7. Both the best guess with the historical errors and the ensemble with estimated errors outperformed the bare ensemble but the ensemble with error outperformed the singleton ensemble. The relative performance of the ensemble with error relative to the bare ensemble indicates how important it is to dress the ensemble forecast. In end to end ensemble forecasting dressing addresses uncertainties both in the weather forecasting model and in the model used to map weather variables to the economic variable of interest.

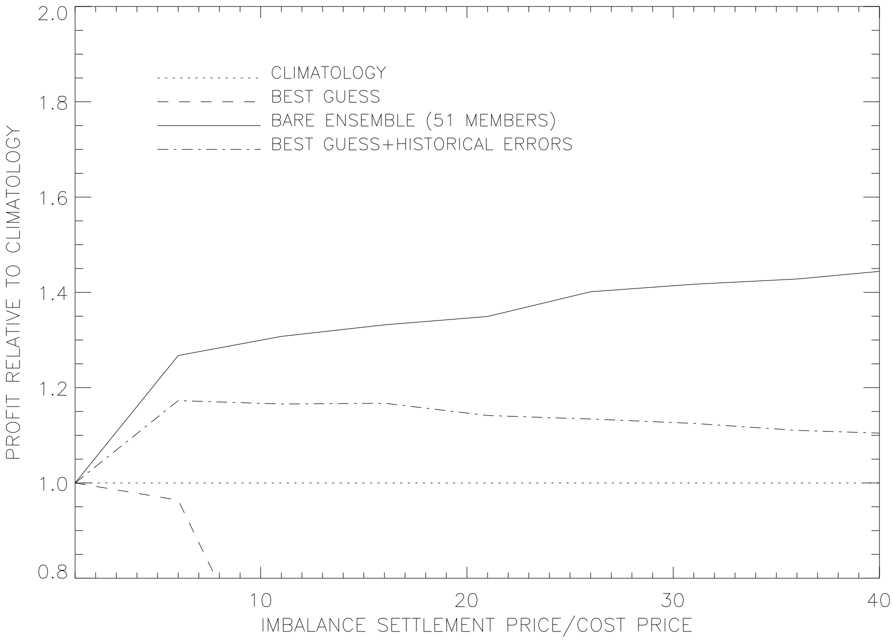

Figure 12: Evaluation of perfect ensemble power demand forecasts for Washington D.C. for the period June 1999 to October 2000. Total profits of users using different forecasts relative to climatology as a function of the imbalance settlement price. The temperature-demand relationship was the linear relationship given by Eq. 6. The true demand was calculated using a temperature picked from the ensemble.

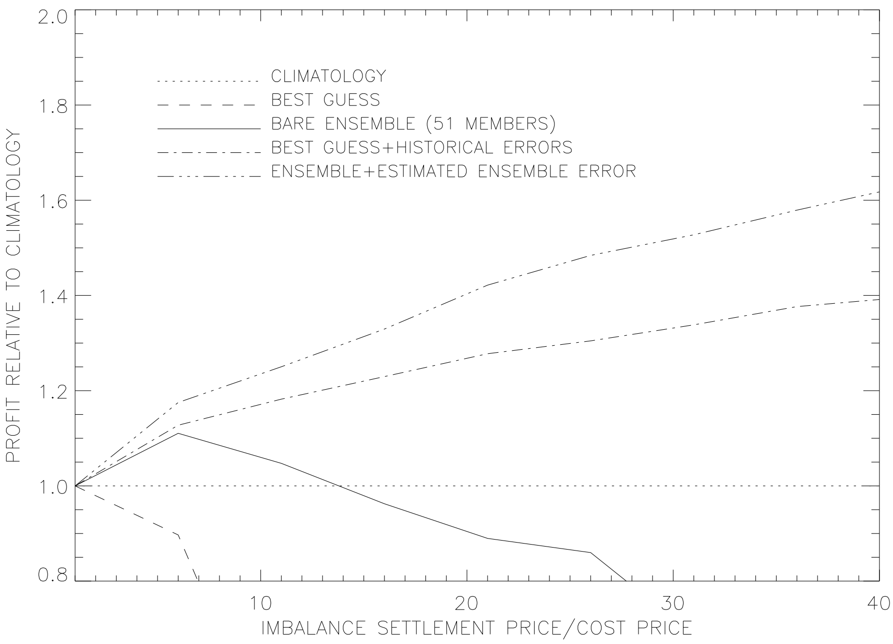

Figure 13: Evaluation of perfect ensemble power demand forecasts for Washington D.C. for the period June 1999 to October 2000. Total profits of users using different forecasts relative to climatology. The temperature-demand relationship was the cubic+stochastic relationship given by Eq. 7. The true demand was calculated using a temperature picked from the ensemble.

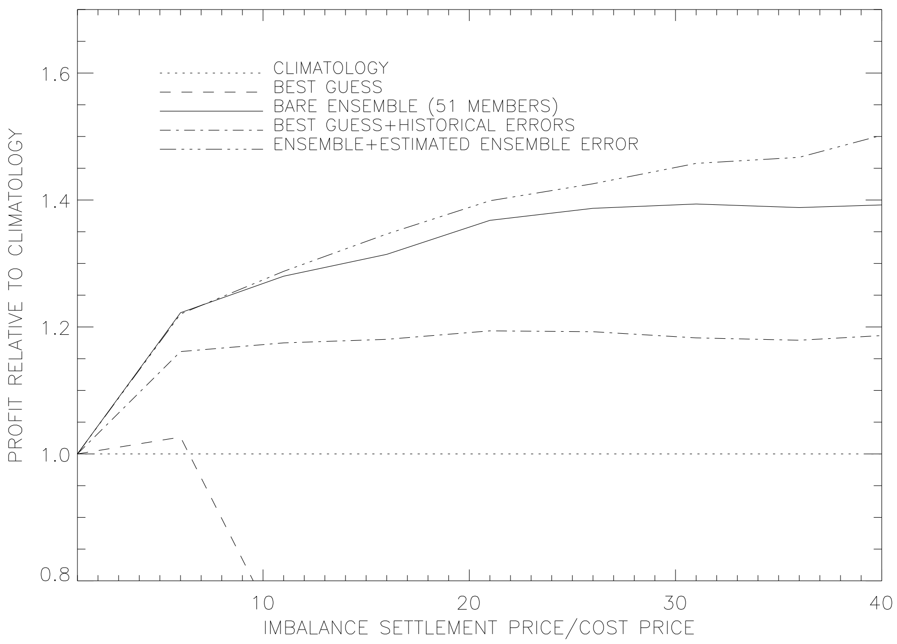

Figure 14: Evaluation of power demand forecasts based on ECMWF 4-day forecasts for Washington D.C. for the period June 1999 to October 2000. Total profits of users using different forecasts relative to climatology. The temperature-demand relationship was the linear-determinisitic relationship given by Eq. 6. The true demand was calculated using observed temperatures.

We have shown that, in this application, in this location, the ensemble forecasts could have offered an advantage over the best guess forecasts provided that an attempt was made to estimate the error distribution on each ensemble member. But what of other locations?

The experiment described above was repeated using ECMWF forecasts for 12 locations around the World. Both the linear-determinisitic demand curve (Eq. 6) and the more realistic, cubic-stochastic, temperature-demand relationship (Eq. 7) were used to calculate the synthetic demand. The results for the linear-determininstic temperature-demand relationship are shown in Figs. 16 and 17. Figure 16 shows the performance of the ECMWF best guess forecast (with historical error) relative to the climatological approach. It can be seen that the ECMWF forecast (with historical error) would have improved decisions at all 12 locations, although the amount is quite variable—ranging from an improvement of a few percent in Beijing to almost 80% in Moscow for high imbalance settlement prices.

In Fig. 17 the relative performance of the ensemble forecasts to the best guess forecast is shown. The ensemble forecasts had the estimated ensemble members errors added and the best guess forecast had the historical errors added. The solid line is the average over all days while the dashed line is the average over only those days on which the ensemble distribution could be distinguished from the best guess with the historical error at the 3σ (99%) level using a chi-squared test. In London, Moscow, Washington, Athens and Hamburg the average performance of the ensemble forecasts improved for those days when it was distinguishable from the singleton ensemble. For the other cities there was little difference in performance.

This result suggests that to improve the relative value of the ensemble forecasts further the fraction

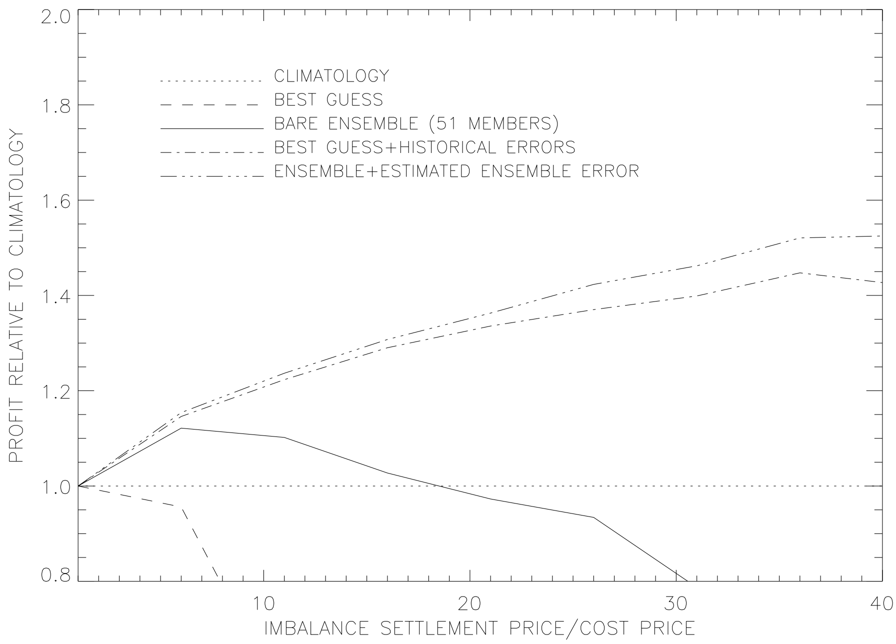

Figure 15: Evaluation of power demand forecasts based on ECMWF 4-day forecasts for Washington D.C. for the period June 1999 to October 2000. Total profits of users using different forecasts relative to climatology. The temperature-demand relationship was the cubic-stochastic relationship given by Eq. 7. The true demand was calculated using observed temperatures.

of days on which they are distinguishable from the singleton ensemble should be increased. This might be accomplished by increasing the size of the ensemble. We estimated a *lower bound* on the ensemble size needed to make the ensemble distinguishable from the singleton ensemble. We assumed that the ensemble distribution for n possible outcomes, e_i was actually the true distribution from which the N ensemble members were picked. If h_i is the distribution of the singleton ensemble then the χ^2 statistic measuring the difference between the ensemble distribution and the singleton ensemble is

$$\chi^2 = N \sum_{i=1} n \frac{(e_i - hi)^2}{h_i} \tag{9}$$

If the distributions e_i and h_i were to be distinguishable at the 3σ level then the minimum value of χ^2 required was

$$\chi^2_{3\sigma} \approx n + 3\sqrt{2n} \tag{10}$$

Given distributions e_i and h_i the minimum value of N required to distinguish them at the 3σ level was estimated using Eqs. 9 and Eq. 10. This was only a lower bound because if they were not distinguishable for the actual value of N then it may be the case that the ensemble actually was picked from h_i in which case even an infinite ensemble would not allow them to be distinguished. The minimum ensemble size for 3σ distinguishability was estimated for two years (1999-2000) of 4-day forecasts for the 12 cities used above. Figure 18 shows the distribution of the estimates minimum ensemble sizes to distinguish the ensemble demand forecasts from the singleton ensembles for the linear-deterministic dependence of demand on temperature. Although it varies from city to city, typically the ensembles can be distinguished on around 65% of days. If the ensemble size were around 12 (the size of NCEP ensembles) the ensembles would have been distinguishable on only around 10% of days. If the ensemble size was increased to over

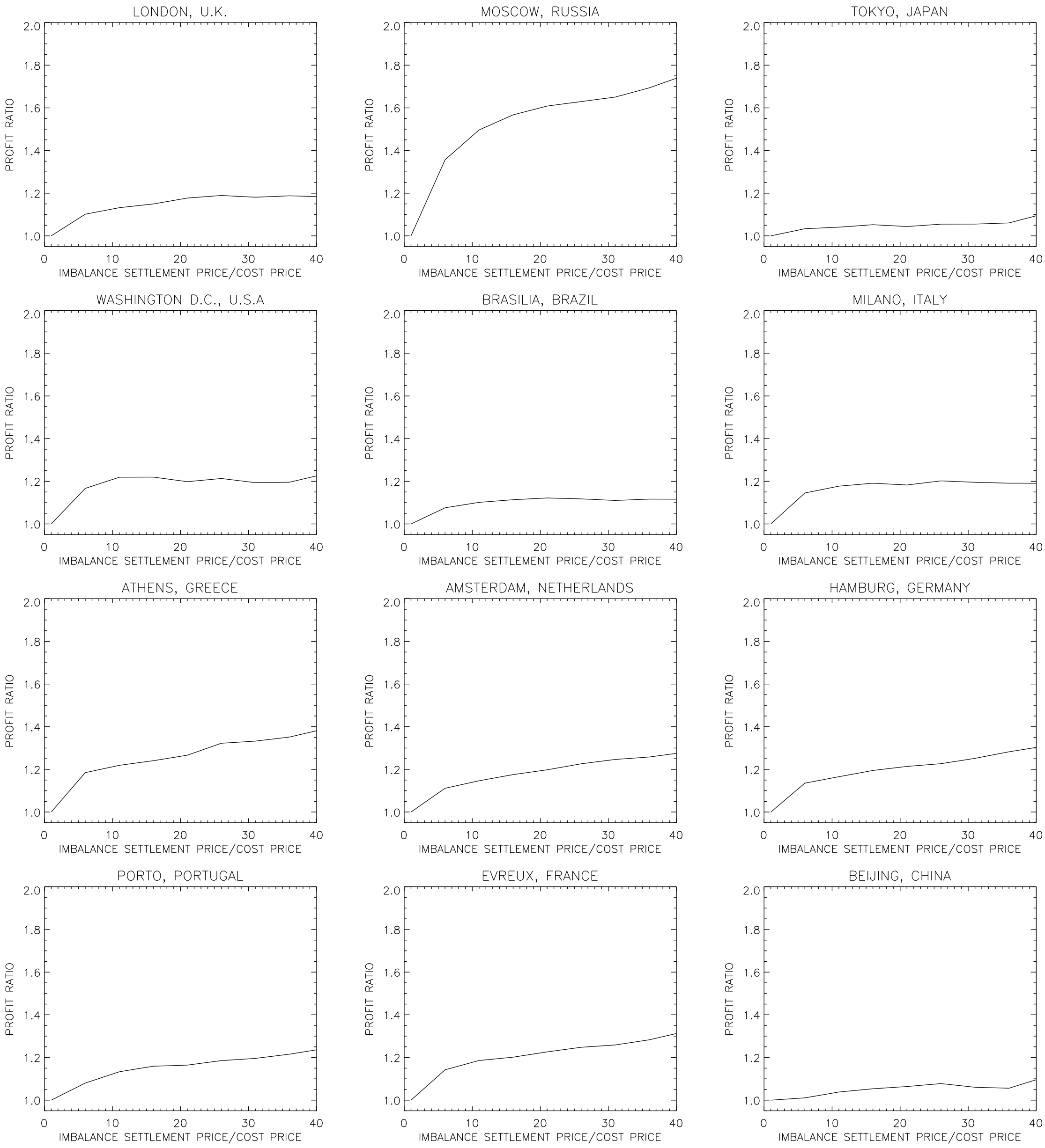

Figure 16: The profit of a power supplier basing demand forecasts on the ECMWF 4-day forecast with the historical error relative to to a supplier using a climatological demand forecast for 12 different locations. The results are averaged over 1999 and 2000. The temperature-demand relationship was the linear-deterministic relationship of Eq. 6.

100 it is possible that the ensembles would have been distinguishable on 95% of days, although it should be remembered that these estimates are *lower* bounds.

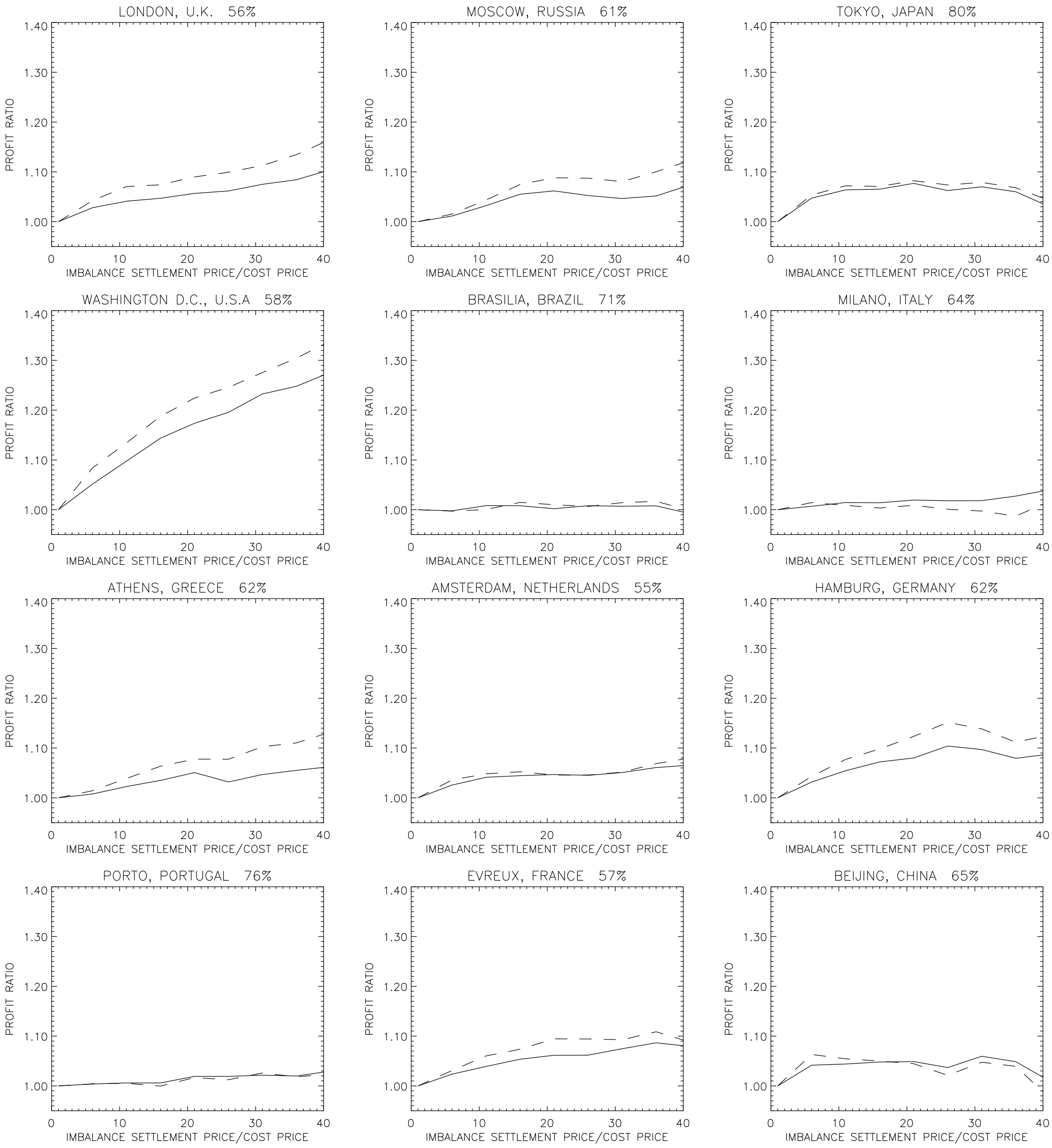

Figure 17: The profit of a power supplier basing demand forecasts on the ECMWF 4-day ensemble forecast with the estimated ensemble member errors relative to the best guess forecast with the historical error added. The solid line is the average over all days while the dashed line was averaged only over the days on which the ensemble forecast was distinguishable from the distribution of historical errors about the best guess forecast at the 3σ (99%) level. The percentage of such days is given at the top of each panel. The temperature-demand relationship used was Eq. 6.

The evaluation was then repeated for the cubic-stochastic temperature-demand relationship given by Eq. 7. Plots of this synthetic demand versus temperature are shown in Fig. 19.

The results of cubic-stochastic demand forecasting experiment are shown in Figs. 20 and 21. Figure 20 shows the profit using demand forecasts based on the ECMWF best guess forecast (with the historical error) relative to the profit obtained using climatological demand forecasts. From Fig. 20 it can be seen

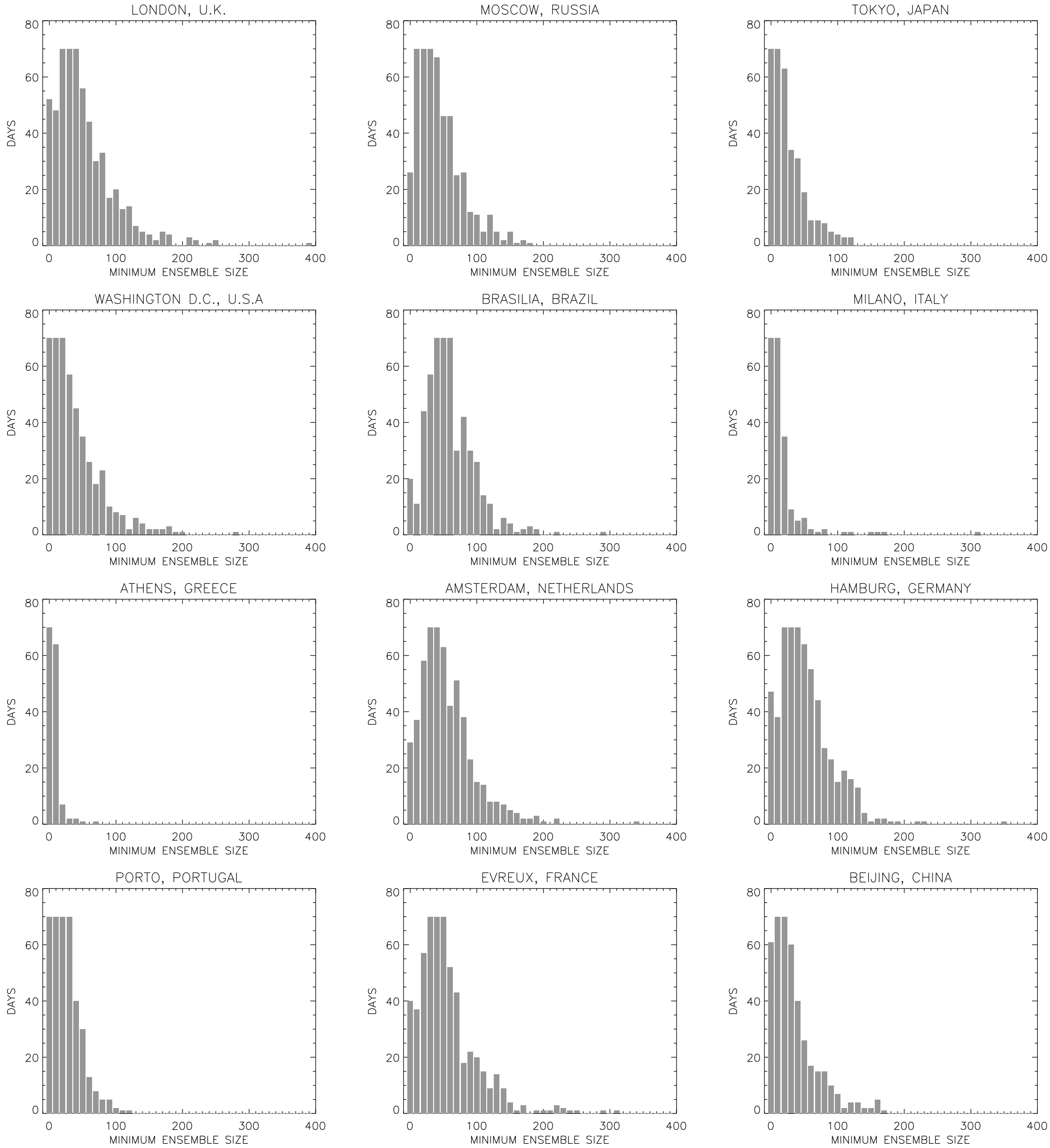

Figure 18: The distribution of estimated minimum ensemble sizes needed to distinguish the ensemble temperature forecasts from the singleton ensemble (best guess with historical errors) at a 3σ confidence level.

that again the value of the ECMWF forecasts to these particular users varied from location to location. Comparing Fig. 19 with Fig. 20 suggests that the relatively low value of the forecasts in Brasilia could be due to the fact that the temperature there had a small spread and tended to fall close the bottom of the U-shaped demand curve. Thus the synthetic electricity demand in Brasilia was not very temperature dependent. In Athens and Beijing the value of the forecasts was less than climatology. In these cities over the course of 1999-2000 the temperature varied enough for there to be strong dependence of demand on temperature. Figure 20, however, shows the value of the forecasts *relative to climatology*. The standard deviations of the departure of daily temperature from a seasonal cycle smoothed on a 100 day time scale are given in Table 1. Athens had a departure from climatology of 2.94° and Beijing one of 3.43°C. These were quite small but the demand forecasts were of value for locations with smaller departures such as Tokyo. In the end to end forecast formalism the value of the forecasts depends on the quality of all parts of the model, both the weather forecasting model and the model that predicts the economic variable using the weather as a predictor. Since the forecasts do have value relative to climatology at all 12 locations when demand is a linear-deterministic function of temperature their lack of value at some locations in Fig. 20 suggests that it is the nature of the temperature-demand relationship at those locations rather than a lack of forecast skill which reduced their value.

Table 1. The standard deviation of daily temperature from a seasonal cycle obtained by smoothing the daily record on a time scale of 100 days.

LOCATION	DEVIATION FROM CLIMATOLOGY °C
LONDON, U.K.	3.13
MOSCOW, RUSSIA	5.43
TOKYO, JAPAN	2.57
WASHINGTON D.C., U.S.A	3.96
BRASILIA, BRAZIL	2.29
MILANO, ITALY	3.35
ATHENS, GREECE	2.94
AMSTERDAM, NETHERLANDS	3.47
HAMBURG, GERMANY	3.97
PORTO, PORTUGAL	3.18
EVREUX, FRANCE	3.44
BEIJING, CHINA	3.43

Figure 21 shows the performance of the ensemble forecasts (with estimated gaussian errors) to the best guess forecast (with historical errors) when the demand was the cubic-stochastic function temperature. In this case the advantage of the ensemble forecasts is less than when demand was a linear-determinisitic function of temperature. In Athens the ensembles actually performed worse than the singleton ensemble. In this case a significant contribution to the error in the estimated demand came from the stochastic term in Eq. 7. This error was common to both forecasting strategies. Improved weather forecasts cannot reduce this source of uncertainty.

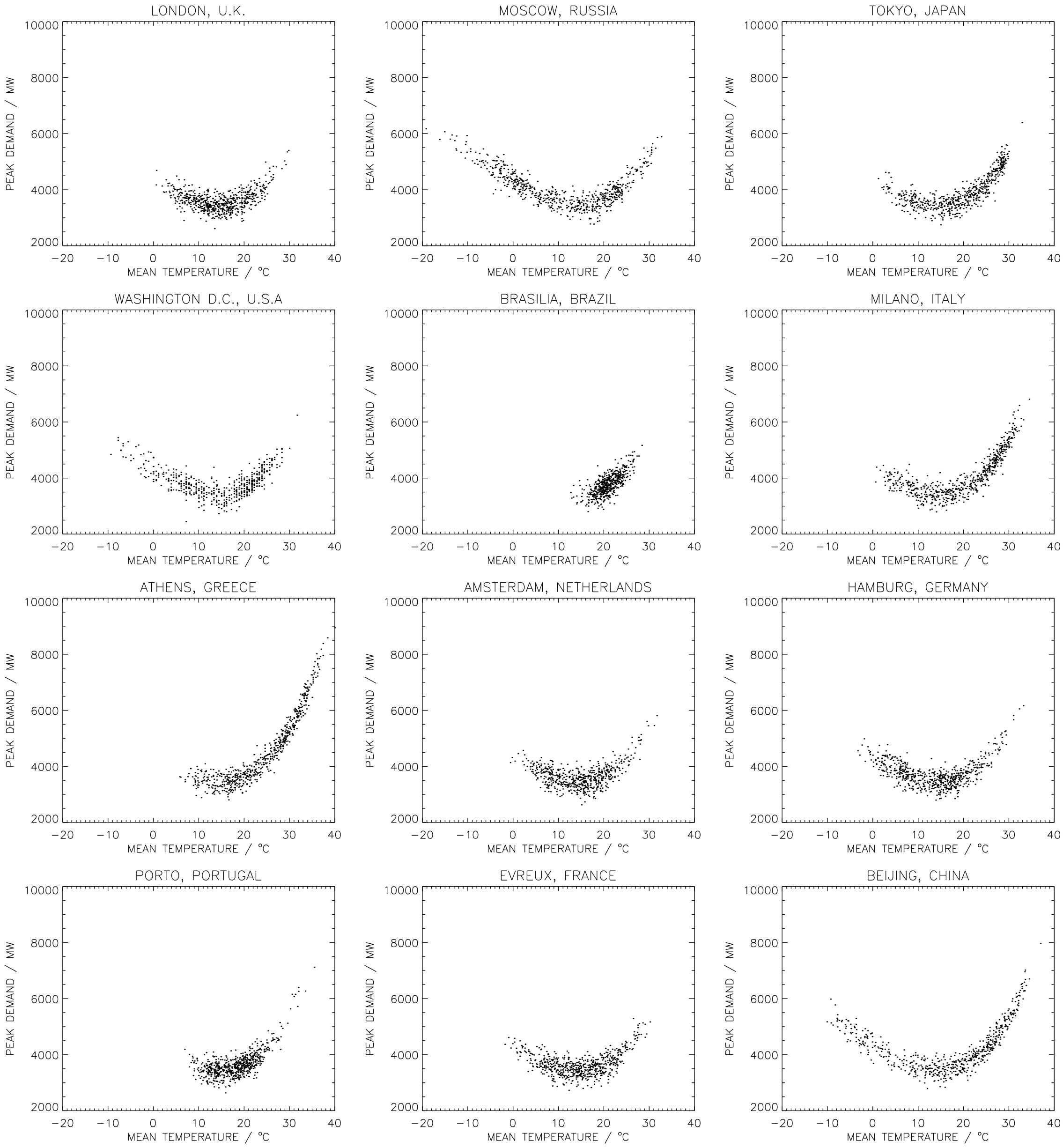

Figure 19: Synthetic demand plotted against temperature for the 12 locations used in the experiment. The temperature-demand relationship used is given by Eq. 7.

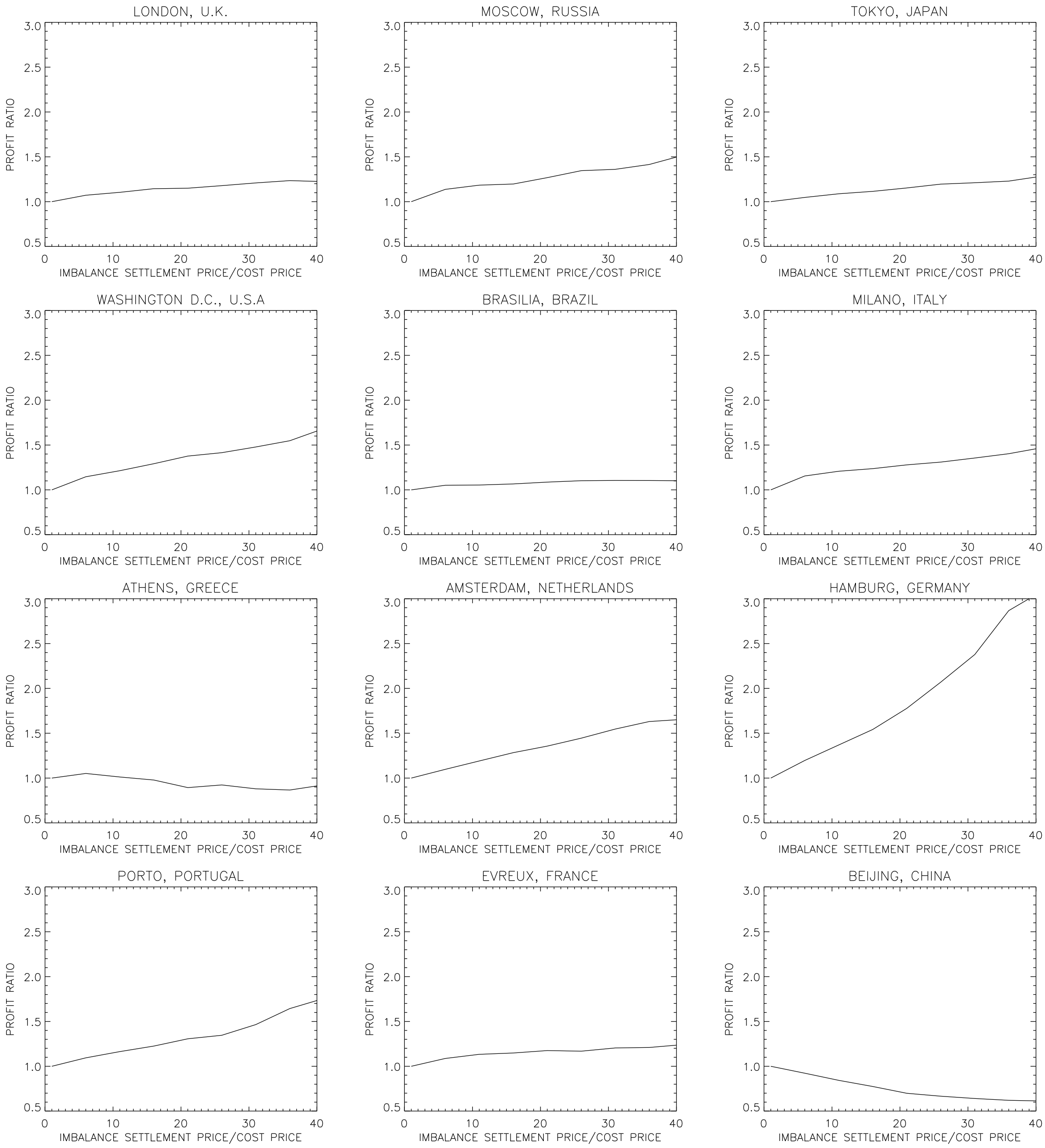

Figure 20: The profit of a power supplier basing demand forecasts on the ECMWF 4-day forecast with the historical error relative to to a supplier using a climatological demand forecast for 12 different locations. The results are averaged over 1999 and 2000. The temperature-demand relationship was that given by Eq. 7.

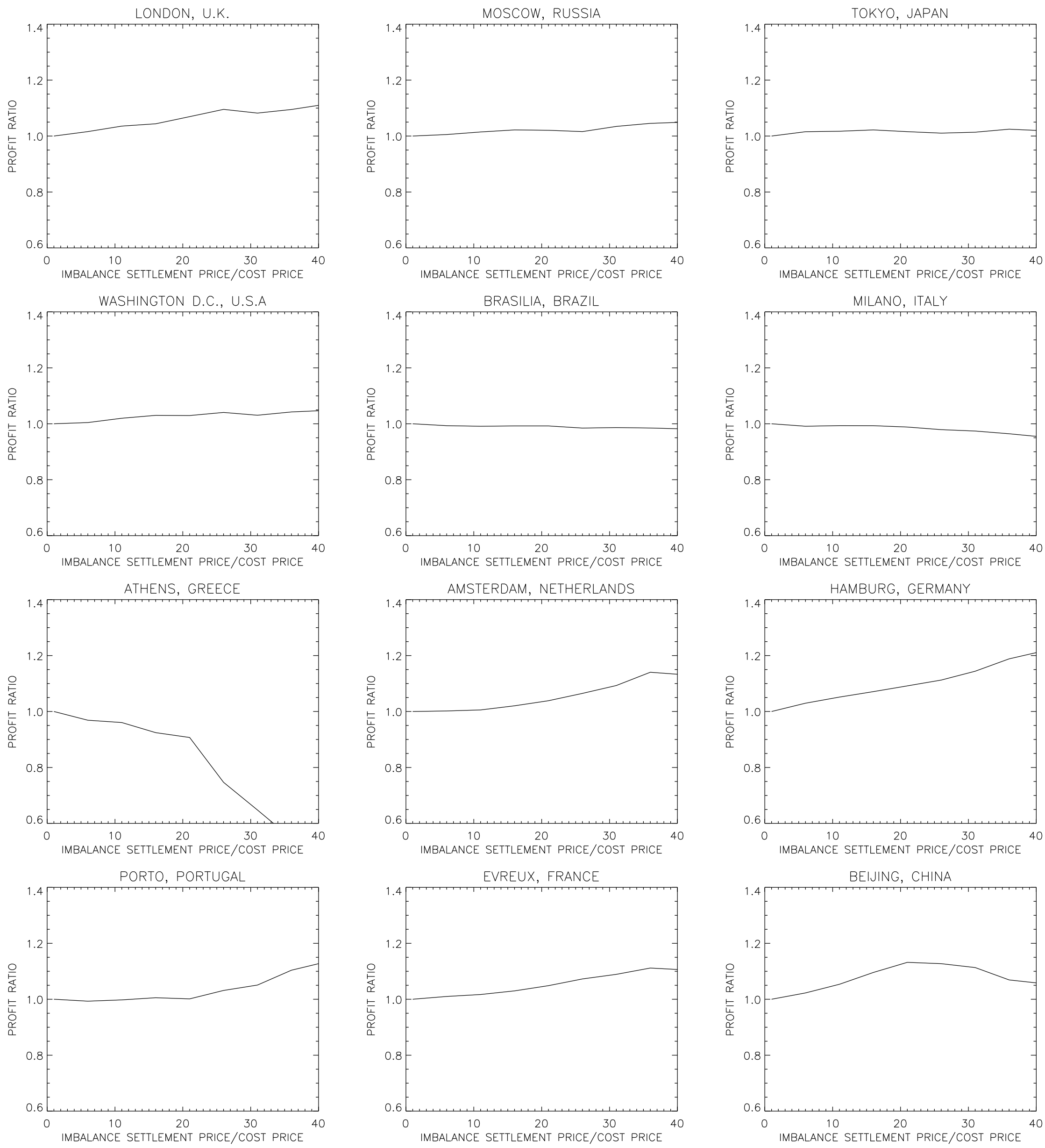

Figure 21: The profit of a power supplier basing demand forecasts on the ECMWF 4-day ensemble forecast with an estimated gaussian ensemble error to a supplier using the ECMWF best guess forecast with historical error for 12 different locations. The temperature-demand relationship was that given by Eq. 7.

4 Wind Energy Supply

In the previous section we described how weather forecasts can help electricity suppliers optimize decision making by providing better forecasts of demand. For electricity generators who produce electricity using renewable sources (wind, wave and solar) the weather is also a major factor in determining their capacity to export electricity to the grid. Under NETA it will be important that these generators be able to forecast their likely generation in advance since failure to provide power specified in a contract will lead to non-delivery charges. We will now consider the ability of ECMWF forecasts to forecast the potential generation of a wind farm.

The amount of power, P, generated by a wind turbine is given by

$$P = C_D \pi R^2 V^3 \tag{11}$$

where V is the windspeed, R is the radius of the turbine and C_D is an efficiency coefficient. Since C_D and R are properties of the turbine the key unknown is V^3. If a wind energy producer wishes to promise a certain power output over a certain period they would like to know what the average value of V^3 is likely to be over this period. Can forecasts of V at 12 hour intervals be used to forecast the mean V^3 that can be expected over a day? To answer this question we used windspeed data measured by the Energy Research Unit Test Site of the CLRC Rutherford Appleton Laboratory (RAL) in Oxfordshire. The data consisted of windspeed measurements at one minute intervals from January 1999 to July 2000. We compared the daily mean of the cube of the windspeed with the ECMWF best guess 4-day forecast for the windspeed for Heathrow, London. Figure 22 shows this comparison. Although there is a substantial amount of scatter the correlation coefficient between the daily mean cube of the windspeed at RAL with the cube of the 4-day windspeed forecast at Heathrow is 0.60.

To assess the value of the ECMWF windspeed forecasts might have for a company generating power using wind turbines we used a toy model for the wind generator's profits.

Suppose the generator must decide 4 days in advance how much power they will supply on a given day. Let this amount be, PP. They sell this power at a price of SP. If actual production, AP, exceeds this promised amount they cannot sell it. If actual production falls short of the promised amount the generator must replace the missing power at a replacement cost of RC. This replacement cost might be in the form of non-delivery charges or it might be the cost of running a conventional generator owned by the wind generating company. If we assume the wind generator is risk neutral their utility function is given by

$$\begin{aligned} U &= SP \cdot PS \qquad AP \geq PP \\ U &= SP \cdot AP - RC \cdot (PP - AP) \qquad AP < PP \end{aligned} \tag{12}$$

We also assumed that the actual production, AP, was proportional to the daily mean cubed windspeed at RAL. The relative profits of utility maximizing wind generators using different forecasts of mean cubed windspeed are shown in Fig. 23. In this case the climatology was taken to be nonseasonal. In this application, over this test period, the ECMWF ensembles had a clear advantage over climatology or the best guess ECWMF forecast, even when the historical errors were added. The best performance was

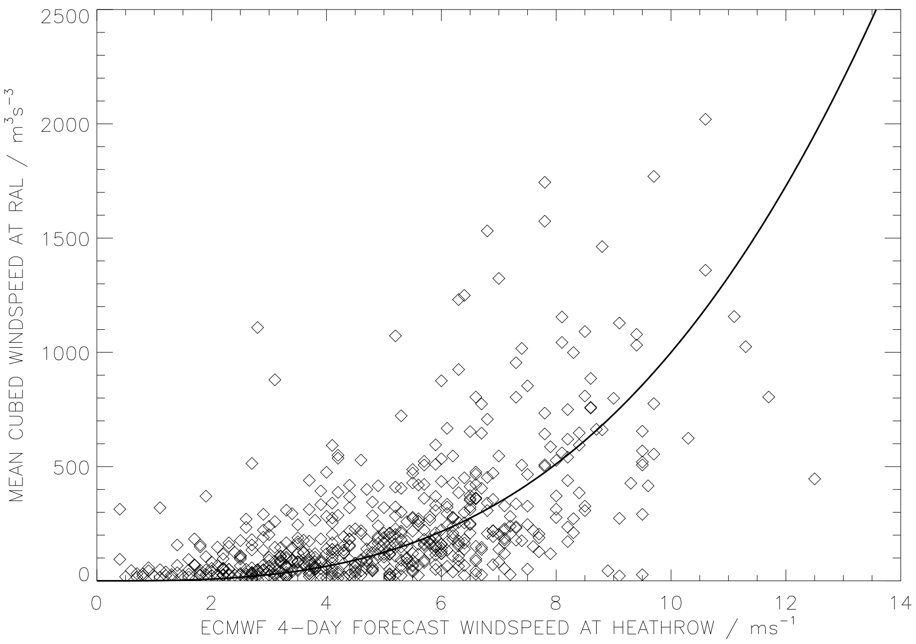

Figure 22: The daily mean cube of the windspeed sampled at one minute intervals at the Environmental Research Unit of the Rutherford Appleton Laboratory in Oxfordshire plotted against the 4-day ECMWF windspeed forecast for Heathrow corresponding to the same day. The line is the forecast windspped cubed.

achieved by the ensemble forecast with the estimated gaussian errors on each ensemble member. In the end to end ensemble approach these gaussian errors include the error in the model used to map the ECMWF ensembles onto daily mean cubed windspeed. From Fig. 22 it can be seen that these errors are quite large. For this example the method for constructing the forecast of mean cubed windspeed at RAL was very crude. Using ECWMF forecasts from several surrounding stations may have improved these forecasts. Nevertheless, the results indicate that a commericial wind generator operating in Oxfordshire would have substantially improved their profitability by using ECMWF ensemble forecasts to make decisions. The fact that potential wind generation goes as the cube of the windspeed means this particular economic quantity is a strong function of a physical weather variable so it is not suprising that weather forecasts should be so potentially valuable to a wind generating company.

A wind energy company that owns multiple windfarms would be interested in windspeeds at each of its sites. It is probable that the windspeed at different sites would be correlated. The company should take this correlation into account when making decisions. This correlation is automatically incorporated into the ensemble forecast. Calculating the correlation between errors associated with individual ensemble members is discussed by Roulston and Smith [2001].

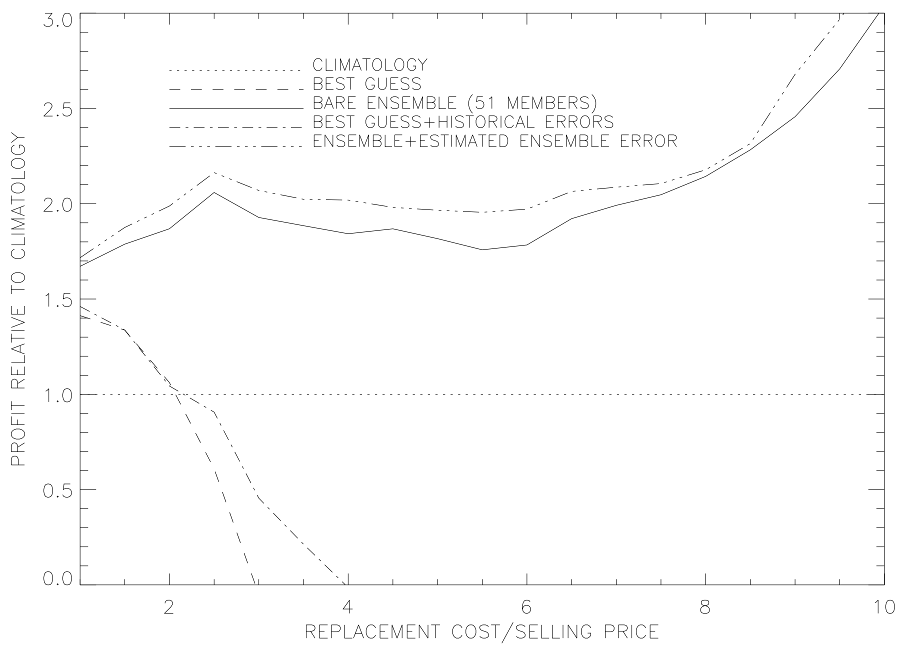

Figure 23: The relative profit of a wind power producer using different forecasts to predict average mean cubed windspeed at a lead time of 4-days. Relative profits are shown as a function of the ratio of the replacement cost of electricity promised by not provided to the selling price of the electricity. The results are averages for the period January 1999 to July 2000. All the ECMWF forecasts were for Heathrow while the actual production was taken to be daily mean cubed windspeed at RAL, Oxfordshire.

5 Summary and Conclusions

In this paper we have described some key concepts involved in decision making in the presence of uncertainty.

We have shown, using relatively simple examples, how ECMWF forecasts might be used by businesses that have a weather dependent component to the supply or demand of the good that they produce. In all the examples considered deterministic forecasts, whether it is a single best guess forecast or a member of an ensemble, must be converted into a probabilisitic forecast by attempting to estimate the error distribution on the forecast. Failure to do so would have caused very sub-optimal decision making due to the nonlinearity of the user's utility function. It has also been demonstrated that on average, the probabilisitic forecasts based on the ECMWF ensembles would have given better decision making than the probabilistic forecasts based on a single best guess forecast. The actual value of the ECMWF forecasts to a user depends on the quality of the weather forecasts, the quality of the model that relates weather to the relevant economic variable and also to the user's utility function. Since these last two factors vary from user to user making general statements about the value of weather forecasts can be difficult. The situation for each user will be different but the decision making framework outlined in this paper should apply to a large range of users.

In the case of electricity demand ECMWF forecasts can be used to improve forecasts of demand and so enable suppliers to decide how much electricity they should purchase in advance. It was found that the ensemble distributions were not always distinguishable from the distribution obtained by adding historical errors to the best guess forecast. With a 51 member ensemble they were distinguishable at a 3σ (99%) confidence level on around two thirds of days. It was also shown that on average, on days on which the ensembles are distinguishable the performance of the ensembles is, on average, better than on days when they are not distinguishable. Estimates of the minimum ensemble size required to make the ensembles distinguishable indicate that an ensemble of at least 100 members would be required to distinguish the ensembles on more than 95% of days.

In the case of wind generated electricity supply ECMWF forecasts could enable a generator to predict the likely production of a wind farm several days in advance. This could enable such a generator to compete more effectively in the electricity market where prices tend to be particularly volatile compared to other commodities. The dependence of potential generation on the cube of the windspeed clearly makes this particular economic variable very sensitive to weather. Because of the strong dependency of the wind generation on windspeed the ECMWF ensembles could yield substantial advantages to wind generators who make good use of them.

References

[*DTI* 2000] Department of Trade and Industry, An Overview of the New Electricity Trading Arrangements, 2000.

[*Johnson and Holt* 1997] Johnson, S.R. and Holt, M.T., The value of weather information, in *Economic Value of Weather and Climate Forecasts,* 75-107, (Eds. Katz, R.W. and Murphy, A.H.) Cambridge Univ. Press 1997.

[*Locke* 1999] Locke, J. US Power: Markets come of age, *Energy Power Risk Management,* Jul/Aug 1999.

[*Molteni et al.* 1996] Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T., The ECMWF ensemble prediction system: Methodology and validation, *Quart. J. Roy. Met. Soc., 122,* 73-119, 1996.

[*Richardson* 2000] Richardson, D.S., Skill and relative economic value of the ECMWF ensemble prediction system, *Quart. J. Roy. Met. Soc., 126,* 649-667, 2000.

[*Roulston and Smith* 2001] Roulston, M.S. and Smith, L.A., Evaluating probabilistic forecasts using information theory, submitted to *Mon. Wea. Rev.*, 2001.

[*Roulston and Smith* 2001] Roulston, M.S. and Smith, L.A., Information maximizing ensemble member error distributions, *in preparation*, 2001.

[*Smith* 1995] Smith, L.A., Accountability and error in nonlinear forecasting, in *ECMWF Seminar on Predictability, 1,* 351-368, ECMWF, Reading, U.K., 1995.

[*Smith* 2000] Smith, L.A., Distinguishing uncertainty and error: On the predictability of nonlinear systems, in *Nonlinear Dynamics and Statistics,* 31-64, Birkhäuser, Boston, 2000.

[*von Neumann and Morgenstern* 1944] von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behavior.* Princeton Univ. Press, 1944.