Reply

ROBERTO BUIZZA

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received and in final form 16 November 2010)

1. Introduction

Glahn (2011) discusses the results of Buizza (2010a), and concludes that ''Buizza (2010a) has not answered the key question'' that he posed in the introduction. This criticism is based on the fact that VAR3 (run with horizontal spectral truncation T399 up to forecast day 3 and T255 afterward) forecasts up to day 8 are 25% more costly than T319 forecasts, and thus the two configurations cannot be considered comparable. In his discussion, Glahn (2011) recognizes that Buizza (2010a) stated that a VAR3 forecast up to day 15 would be 2% cheaper than a T319 forecast, but states that since results were shown only up to forecast day 8 this does not really matter since VAR3 forecasts might deteriorate faster than T319 forecasts after day 8.

In this reply I will comment on some of the issues raised by Glahn (2011), and I will provide some new results that should address Glahn (2011)'s criticism.

2. CPU time estimates

Before discussing these new results, I recognize that I should have added a caveat such as ''where comparable meant that the two forecast systems required CPU times with differences smaller than 25%'' when I stated in Buizza (2010a) that the VAR3 and the T319 configurations have ''comparable'' cost. It is also worth clarifying that the CPU costs have been computed assuming that CPU is proportional to the squared spectral truncation and is inversely proportional to the time step. Thus, these estimates do not take into account any potential impact on the CPU cost of running the forecasts with different configurations (e.g., by using more or less processors, or by making the length of the control vector longer or shorter). This is why I have been using the term ''comparable cost'' and not ''identical cost,'' and why these CPU costs should be treated as estimates and not as precise numbers.

3. Results from two new configurations: VAR2 and VAR1

To address Glahn (2011) criticism, I have run the experiment for the same 3-month period analyzed in Buizza (2010a) and using the same model cycle and initial perturbations, two new sets of ensemble forecasts, VAR2 and VAR1, with horizontal spectral triangular truncation T399 up to, respectively, forecast days 2 and 1, and T255 thereafter. The 8-day forecasts run in these two configurations require, respectively, 7% more and 15% less CPU time than the T319 (Fig. 1). The VAR3 configuration requires about 25% (more exactly 28%) more CPU time than T319. If forecasts are run for 15 days, VAR3 requires 2% less CPU time than T319 and VAR2 requires 13% less CPU time (see section 5).

Figure 2 shows the predictability gains of the T319, VAR3, VAR2, and VAR1 configurations, measured using the ''gain'' index defined in Buizza (2010a), with forecasts verified in the ideal model error (IME) and realistic scenario. This figure corresponds to Figs. 15 and 16 of Buizza (2010a), but shows the gain index for the VAR1 and VAR2 instead of the VAR5, T399, and T799 configurations. Figure 2 shows that VAR2 performs only slightly worse than VAR3 and still better than T319, while VAR1 performs very similarly to T319, with larger differences when forecasts are verified in the IME configuration (Fig. 2, left panels).

Table 1 lists the values of the relative predictability gain (RPG) for the 500-hPa geopotential height field of configurations VAR3 and VAR2 with respect to T319, where RPG(VARx, T319) has been defined as

$$\mathrm{RPG(VARx,\ T319)} = \frac{I_2(\mathrm{VARx}) - I_2(\mathrm{T319})}{I_2(\mathrm{T319})}.$$

Corresponding author address: Dr. R. Buizza, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.
E-mail: buizza@ecmwf.int

DOI: 10.1175/2010MWR3657.1

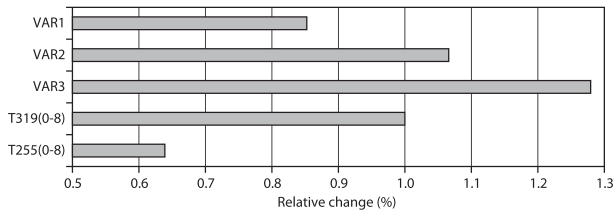

FIG. 1. The CPU time required to complete an 8-day integration in configurations VAR1, VAR2, VAR3 (i.e., with horizontal spectral triangular truncation T399 up to days 1, 2, and 3, and T255 thereafter) and T255, expressed in terms of the CPU time required by the T319 configuration.

Table 1 shows that for Z500 forecasts verified in the realistic scenario, 8-day forecasts run in configuration VAR3 (VAR2), which costs 28% (7%) more than T319, deliver predictability gains compared to the T255 configuration that are 43%–69% (32%–68%) larger than the predictability gains of configuration T319. In the realistic scenario, the relative predictability gains for Z1000 are very similar, while the gains for T850 are smaller, ranging for VAR3 (VAR2) between 20% and 43% (3% and 20%). Table 1 also shows that VAR1, a configuration that requires 15% less CPU time than T319, is performing similarly or slightly worse than T319.

These results indicate that it is beneficial to have a higher resolution in the early forecast range provided that the resolution is kept higher for enough time (e.g., around 2–3 days), and that the benefit decreases with the forecast range. This was already pointed out by Buizza (2010a), who showed (see Figs. 15 and 16) that VAR5 and T399 forecasts brought similar predictability gains. This conclusion is consistent with another set of very recent results

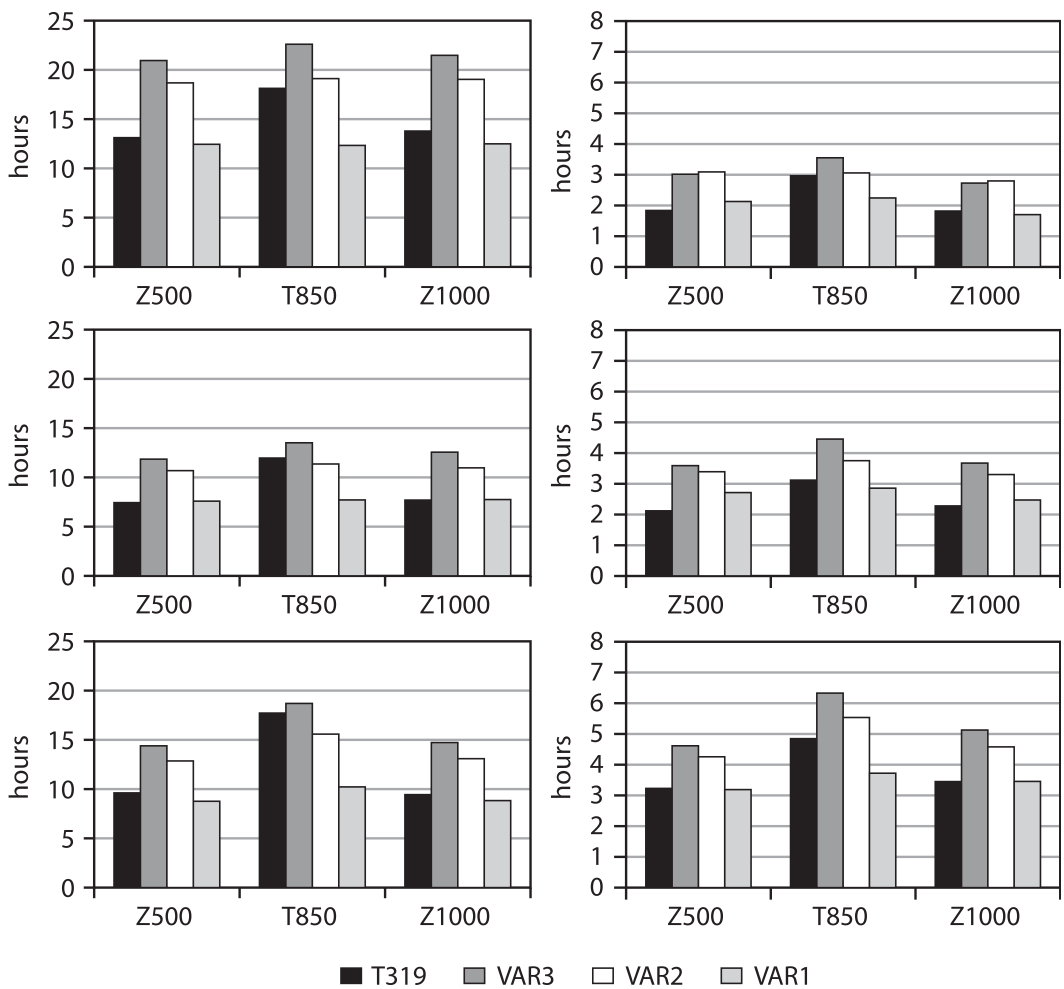

FIG. 2. (left) IME results and (right) realistic scenario results. Predictability gains, measured using the "gain" index I2, for configurations T319 (black), VAR3 (dark gray), VAR2 (white bars), and VAR1 (light gray) computed for the 500- and the 1000-hPa geopotential height (Z500, Z1000) and the 850-hPa temperature (T850) (top) control forecasts, (middle) ensemble-mean forecasts, and (bottom) probabilistic forecasts. A single forecasts' accuracy has been measured using root-mean-square error (rmse) over NH, and probabilistic forecast accuracy using RPSS over NH.

TABLE 1. Relative predictability gains (%) of VAR3, VAR2, and VAR1 configurations computed with respect to the T319 configuration, computed for Z500 forecasts verified in the IME and realistic scenario.

	IME scenario			Realistic scenario		
RPG(VARx, T319) for Z500 over NH	VAR3	VAR2	VAR1	VAR3	VAR2	VAR1
Rmse(CON)	60	42	−5	64	68	16
Rmse(EM)	59	44	2	69	60	28
RPSS	50	34	−9	43	32	−1
CPU extra cost compared to T319	28	7	−15			

obtained by Buizza (2010b), who compared the forecast skill of 15-day forecasts run with horizontal spectral triangular truncations T95, T159, T255, T319, T399, and T799. Buizza (2010b) concluded that "A strong sensitivity to model resolution of the skill of instantaneous forecasts has been found in the short forecast range (say up to about forecast day 3). But sensitivity has shown to become weaker in the medium-range (say around forecast day 7) and undetectable in the long forecast range."

Overall, the VAR2 results discussed in this communication further support Buizza (2010a)'s conclusion that a variable resolution configuration with higher resolution

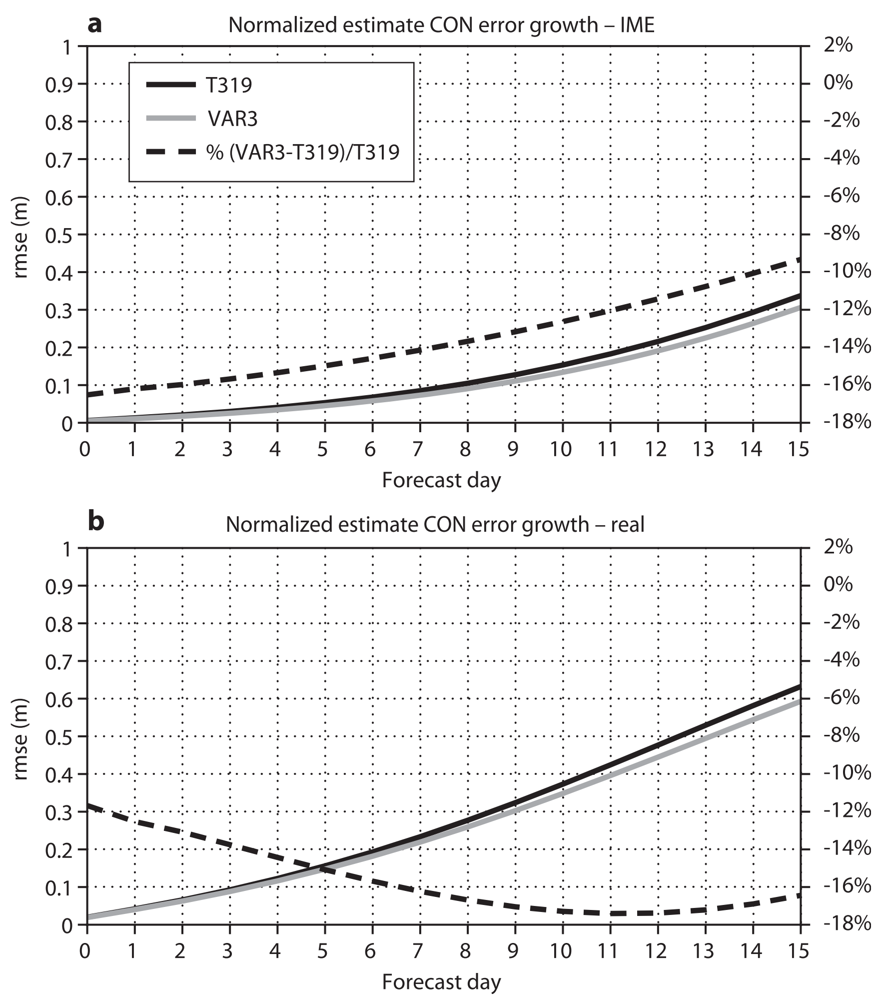

FIG. 3. (top) Total forecast error growth curve for the VAR3 (solid black) and the T319 (solid gray) control forecasts, and percentage difference (VAR3 − T319)/T319 (right axis), with errors computed in the IME scenario against T799 forecasts. (bottom) As in (top), but with errors computed in the realistic scenario against analyses.

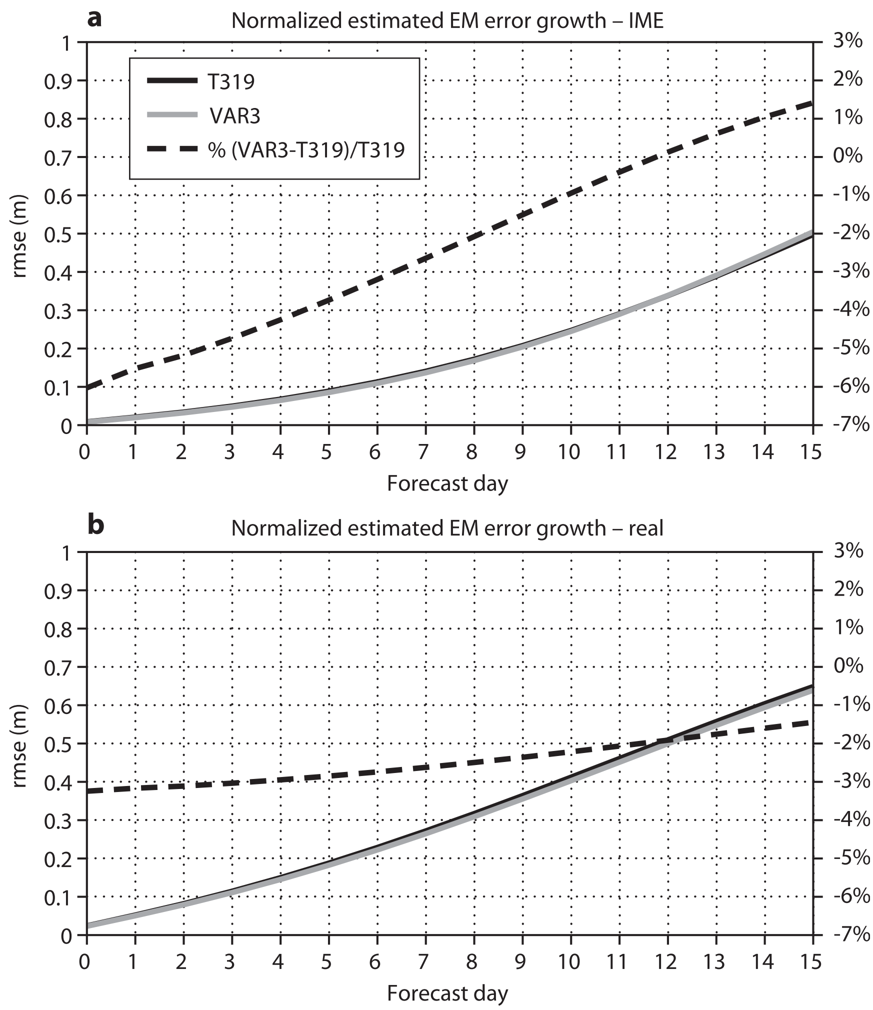

FIG. 4. As Fig. 3, but for the ensemble-mean forecast. (top) Total forecast error growth curve for the VAR3 (solid black) and the T319 (solid gray) ensemble-mean forecasts, and percentage difference (VAR3 − T319)/T319 (right axis), with errors computed in the IME scenario against T799 forecasts. (bottom) As in (top), but with errors computed in the realistic scenario against analyses.

in the earlier forecast range and a lower resolution afterward delivers higher predictability gains than a constant resolution with a comparable cost, where comparable means that the CPU time required to run the 2 configurations differ by less than 7%.

4. Three key messages from Buizza (2010a)

The fact that a variable resolution approach leads to a better system is just one of the conclusions that can be drawn from Buizza (2010a). As is stated in section 4, there are two other key conclusions that can be drawn from the results discussed in section 3:

> The results discussed in section 3 have indicated first that differences between ensemble performances that are large in the IME scenario, almost disappear completely in the realistic case. Second, they have shown that in the IME scenario differences between the ensemble configurations are larger if one considers single than probabilistic forecasts. Third, considering the two ensemble configurations with comparable CPU requirements, VAR3 and T319, they have shown that the VAR3 configuration performs better.

The first point highlights the role of model error not only due to horizontal resolution. As it is stated in section 5:

> In the IME scenario, the only source of forecast error is due to the fact that forecasts use a resolution lower than the one used in the verification… In the realistic scenario, forecasts errors are due not only to T799 versus T399/T319/T255 resolution differences, but also to the fact

that, independently of resolution, the "model" describes only the approximate reality (model "imperfection"). In this realistic scenario, the differences induced by using a T319 or a T399 resolution in the short forecast range are much smaller than in the idealized scenario because they are "masked" by the contribution to the forecast error due to model imperfection. In other words, the contribution to the forecast error due to the resolution difference is not any more dominant.

The second point raises the issue that pure future increases of horizontal resolution might have a smaller impact on the accuracy of probabilistic forecasts than on the accuracy of single forecasts.

5. Validity of Buizza (2010a) results beyond 8 forecast days

Glahn (2011) says that the fact that 15-day VAR3 forecasts are 2% cheaper is not relevant since Buizza (2010a) only shows results up to forecast day 8. To address this point one could use the three-parameter forecast error growth model introduced in section 5b to estimate the average forecast error growth beyond forecast day 8. Figures 3 and 4 show the forecast-error growth model curves for the VAR3 and T319 control and the ensemble mean, with the model coefficients estimated from the 3-month-average forecast errors and the curves normalized by the forecast error limit E_∞:

$$\eta(t) \equiv \frac{E(t)}{E_\infty} = 1 - \frac{1 + \dfrac{S}{aE_\infty}}{1 + C_1 e^{C_2 t}}.$$

As in Buizza (2010b), the forecast error curves have been normalized by the asymptotic limit to take into account the fact that different model resolutions (T255 after day 3 for the VAR3 configuration, and T319 for the T319 configuration) and forecast types (control and smoother ensemble mean) have different asymptotic values. Figures 3 and 4 also show the relative difference in the root-mean-square error between the VAR3 and T319 error curves (dashed line). For the control forecast (Fig. 3), the dashed curve remains below −4% in the 8–15-day forecast range in both the IME and the realistic scenario, with differences again larger in the IME scenario. For the ensemble mean (Fig. 4), the relative difference curve remains negative until forecast day 14. These two figures further confirm that the impact of a horizontal resolution increase is more detectable in the early forecast range, but also indicate that the fact that the VAR3 forecast resolution drops to T255 after day 3, on average, does not deteriorate the forecast in the 8–15-day forecast range, as Glahn (2011) suggested might occur.

6. Conclusions

I hope that this reply has clarified the meaning of "comparable" CPU time, and that the new results discussed above confirms Buizza's (2010a) conclusion that "a numerical weather prediction system with variable resolution, higher in the early forecast range and lower afterward, provides more skillful forecasts than a system with constant resolution."

REFERENCES

Buizza, R., 2010a: The value of a variable resolution approach to numerical weather prediction. *Mon. Wea. Rev.,* **138,** 1026–1042.

——, 2010b: Horizontal resolution impact on short- and long-range forecast error. *Quart. J. Roy. Meteor. Soc.,* **136,** 1020–1035.

Glahn, B., 2011: Comments on "The value of a variable resolution approach to numerical weather prediction." *Mon. Wea. Rev.,* **139,** 2300–2301.