On why dynamic subgrid–scale models work

By J. Jiménez[1]

1. Motivation

Dynamic subgrid models were introduced in (Germano *et al.* 1991) and have proved to be remarkably successful in predicting the behavior of turbulent flows. Part of the reasons for their success are well understood. It is known, for instance, that their behavior as the flow becomes smooth, such as near walls or during transition, is better than that of other "hand-tuned" models. Since they are constructed to generate an effective viscosity which is proportional to some measure of the turbulent energy at the high wavenumber end of the spectrum, their eddy viscosity vanishes as the flow becomes laminar. This alone would justify their use over simpler models.

But beyond this obvious advantage, which is confined to inhomogeneous and evolving flows, the reason why they also work better in simpler homogeneous cases, and how they do it without any obvious adjustable parameter, is not clear. The simplest case, and one of the first to be documented, is the decay of grid turbulence as measured in (Comte–Bellot & Corrsin 1971), which was shown to be well predicted by simple dynamic models in (Moin *et al.* 1991).

This lack of understanding of the internal mechanisms of a useful tool is disturbing, not only as an intellectual challenge, but because it raises the doubt of whether it will work in all cases. This note is an attempt to clarify those mechanisms. We will see why dynamic models are robust and how they can get away with even comparatively gross errors in their formulations. This will suggest that they are only particular cases of a larger family of robust models, all of which would be relatively insensitive to large simplifications in the physics of the flow. We will also construct some such models, although mostly as research tools.

It will turn out, however, that the standard dynamic formulation is not only robust to errors, but also behaves as if it were substantially well formulated. The details of why this is so will still not be clear at the end of this note, specially since it will be shown that the "a priori" testing of the stresses gives, as is usual in most subgrid models, very poor results. But it will be argued that the basic reason is that the dynamic formulation mimics the condition that the total dissipation is approximately equal to the production measured at the test filter level.

1 Center for Turbulence Research and School of Aeronautics, U. Politécnica, Madrid

2. Accomplishments

2.1 Numerical experiments

We will restrict ourselves to the simple case of the grid turbulence experiments in (Comte–Bellot & Corrsin 1971), reduced to a temporal decay through the usual Galilean transformation, and to the simplest formulation of the dynamic model (Lilly 1992). We establish the notation next.

Consider two filters with characteristic widths δ and $\Delta = 2\delta$. In all our experiments the filters are spectrally sharp, the code is spectral on a triply periodic cubic box (Rogallo 1981) with 32^2 Fourier modes before de-aliasing, and the narrower filter coincides with the grid.

The initial conditions are obtained by filtering a flow field which has been left to decay at a resolution of 64^3 to an energy and spectrum closely resembling those of Comte-Bellot and Corrsin at their first experimental section. The energy transfer, as measured by the skewness of the velocity gradients, is past its maximum value and has begun to decay. The initial skewness of the filtered field is about -0.27 and decays to about -0.21 at the end of the computation. Because the field is disturbed by the initial filtering operation, the cascade is initially perturbed, and it takes a few time steps to recover, but the recovery is fast and the decay proceeds thereafter in an approximately self-similar manner. Both the initial field and the original simulation code were kindly provided by T. Lund.

For the grid- and test-filtered velocity fields we compute Reynolds stresses and rate of strain tensors which we will call τ_{ij}, σ_{ij}, and T_{ij}, S_{ij}, respectively. The test-filtering operation will be denoted by $<\cdot>$, while an overbar will be reserved for averaging over the whole flow field. Because of our choice of the narrow filter, there is no explicit grid-filtering operation, although our numerical velocities should be interpreted as being related to the experimental ones by filtering at width δ

A tensor is denoted by the same letter as its components, and inner products and norms have their usual meaning. In a minor departure from usual LES practice, the symbol $|\cdot|$ is reserved for the L_2 norm, so that $|S|^2 = S_{ij}S_{ij}$, without the extra factor of two used by some authors.

We introduce the Smagorinsky weighted strains

$$M = 2\sqrt{2}\Delta^2|S|S, \qquad m = 2\sqrt{2}\delta^2|\sigma|\sigma, \tag{1}$$

and the differences

$$L = T - <\tau>, \qquad g = M - <m>. \tag{2}$$

The Smagorinsky assumption at both filter levels is that

$$T^* + cM = 0, \qquad \tau^* + cm = 0, \tag{3}$$

where the star stands for traceless projection, $T^* = T - \frac{1}{3}\mathrm{tr}(T)I$. Subtracting and neglecting the spatial variability of the proportionality constant c leads to the tensor equation

$$\lambda \equiv L^* + cg = 0, \tag{4}$$

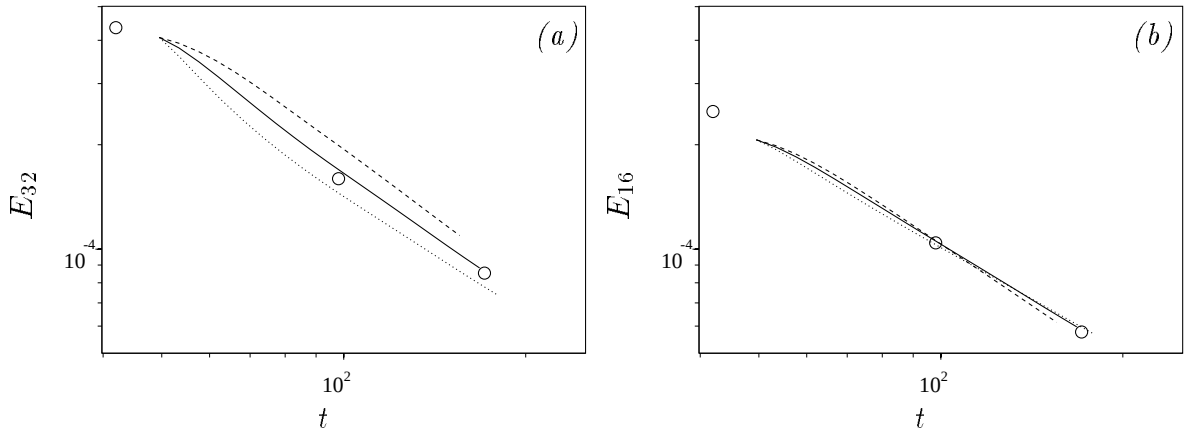

FIGURE 1. Decay of filtered energy for modified dynamic models. ——— : $f = 1$; – – – – : $f = 0.5$; ····· : $f = 2$. Symbols are experiment in (Comte–Bellot & Corrsin 1971). (a) Filtered at grid level. (b) Filtered at test level.

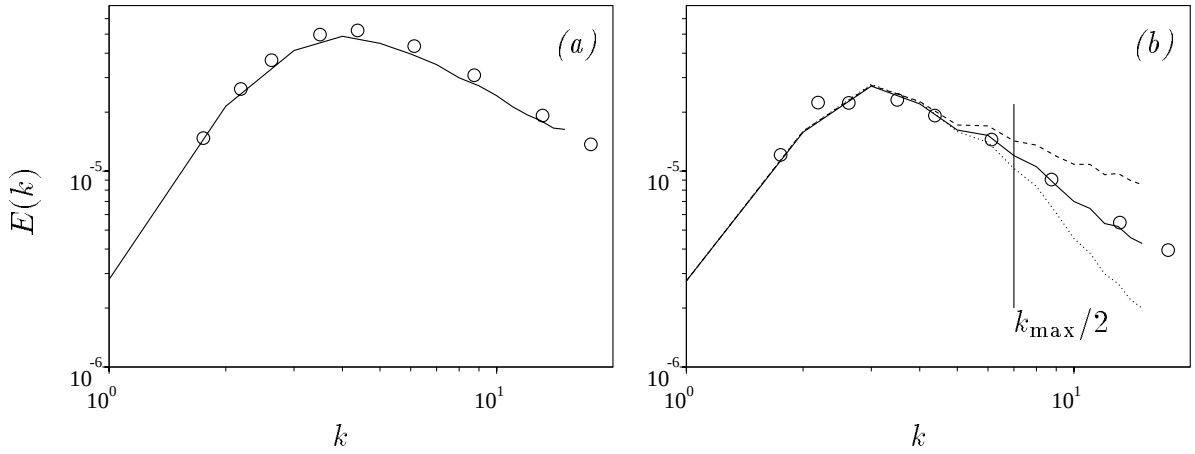

FIGURE 2. Energy spectra of modified dynamic LES runs. Symbols as in Fig. 1. (a) Initial numerical spectrum and $t = 42$ for the experiments. (b) $t \approx 98$.

The constant c is chosen so as to satisfy some contraction of (4), and it has become standard to use g as the contracting tensor (Lilly 1992), on the grounds that it minimizes the L_2 norm of (4). It is well known that when this is done locally numerical instabilities arise because of artificially high back-scatter in those points in which c becomes negative, but that this is cured by averaging over large volumes of the flow. In this note we always average over the whole flow field,

$$c = -f\,\frac{\overline{L^* \cdot g}}{\overline{|g|^2}}, \qquad f = 1, \tag{5}$$

where the unit factor f is introduced for later convenience. This choice minimizes the norm of (4) when its definition is taken to include integration over the whole volume. Other strategies have been proposed, and in particular the original formulation used S as the contracting tensor (Germano *et al.* 1991). We will not

present here results for that formulation, but experience, including ours during the preparation of this note, indicates that its performance is similar to that of (5).

The final step of the model is to apply (3) for the calculation of τ in the equations of motion.

2.2 Robustness

One way to understand a phenomenon is to observe its response to artificial perturbations, and to study (5) we undertook a series of numerical experiments in which errors were purposefully introduced into it by changing the arbitrary factor f.

As expected, the initial rates of decay are changed proportionally to the change of f, somewhat surprisingly, the effect is only temporary and the logarithmic rate of decay soon recovers the same value as in the undisturbed case, which is very close to the experimental one. The only lasting effect of the prefactor is an offset in the initial conditions (Fig. 1a).

The reason for this is clear once the spectra of the decaying turbulence are examined (Fig. 2b). The one computed with $f = 0.5$ has too much energy in the small scales, while the one computed with $f = 2$ is damped in that region. The large scales, on the other hand, are very similar in the three cases, even if the total energy in the flow has decayed from the initial condition (Fig. 2a) by more than a factor of two. The energy differences seen in Fig. 1a are almost totally due to the differences in the high wavenumbers of the spectra, while the large scales are unaffected by the change of the subgrid model.

In fact, if the energy of the flow is measured by filtering at the test level, which could be argued to be a more natural measure of performance, the three runs are indistinguishable (Fig. 1b), although they are separated by a factor of four in the definition of the model.

This is consistent with the classical idea that the rate of energy decay is fixed by the large scales of the flow (the production), while the small scales adjust themselves to dissipate whichever energy is fed to them by the cascade.

The way in which the adjustment occurs in this particular case is also clear. Consider first the classical Smagorinsky model in which c is a predetermined constant. The dissipation of the model is then $\tau \cdot \sigma \sim c|\sigma|^3$. If c is chosen too low, not enough energy is dissipated at the small scales to compensate for production at the large ones, and energy accumulates in the high wavenumbers. This in turn raises $|\sigma|$ and increases the dissipation, until both rates are again in equilibrium. For a $k^{-5/3}$ spectrum the strain depends mainly on the high wavenumbers, which contain little energy. As a result the adjustment can be accomplished with relatively little effect on the total energy of the flow, and the model is robust to mistuning of the constant c. The Smagorinsky model is in this sense slightly superior to regular viscosity because it makes the dissipation proportional to the cube of $|\sigma|$, rather than to the square, and it is therefore able to adjust itself with milder effects in the total energy.

If, in addition, we accept the last octave of the spectrum as a "sacrificial" range of scales available as a buffer for the model, the effect of the errors in c is minimal, as is the case in Fig. 1b.

2.3 Hyper-Smagorinsky models

This analysis suggests that subgrid-models could be made more robust than Smagorinsky by making their dissipation dependent on measures that are more concentrated towards the high wavenumber end of the spectrum, in such a way that they can adjust with still smaller effects on the total energy.

Consider for example, a "hyper-Smagorinsky" model,

$$\tau^* = -c_n|\sigma_n|\sigma, \tag{6}$$

based on a hyper-strain

$$|\sigma_n|^2 = \int k^{2n} E(k) dk, \tag{7}$$

Note that the case $n = 1$ is a "global" Smagorinsky, in which $|\sigma|$ is computed over the whole field rather than locally. Because of the higher powers of k inside the integral (7), the hyper-strain depends more locally on the tail of the spectrum when $n > 1$, and the models should be able to adjust the dissipation with less effects on the total energy. This is confirmed by the experiments in Fig. 3, where the prefactor technique is applied to the hyper-Smagorinsky models. For each value of n the optimal constant c_n is determined empirically to make the energy decay approximately as in the experiment, and is then modified by substituting it by fc_n.

There are three groups of curves in the figure. The central one corresponds to E_{32} with $f = 1$, while the upper group corresponds to $f = 0.5$ and the lower one to $f = 2$. It is clear that as n increases the sensitivity of the model to errors in the constant decreases, and this is confirmed in Fig. 4, in which the ratio between the energies computed with $f = 0.5$ and 2 is plotted as a function of decay time.

An ideal model would be completely insensitive to the prefactor and would maintain this ratio equal to one. The hyper-Smagorinsky models approach this behavior as n increases, but they never reach the optimum limit because they use an eddy viscosity, which cannot change the total dissipation without affecting broad ranges of the spectrum. A still better family of models would have a hyperviscosity component, but such models are numerically inconvenient and are not explored here. The dynamic model is also included in the figure and is shown to behave best of all, with a sensitivity that is roughly half that of Smagorinsky. This is easy to understand since the effect of large n's is to concentrate the model feedback "sensor" near the end of the spectrum, while the dynamic model computes its constant exclusively from the last octave through the effect of the two filters. Because of that, the dynamic formulation should be nearly optimal among eddy viscosity models with respect to robustness.

Note that in all these cases the initial jump of the energy ratio corresponds to a transient in which the spectrum has not had time to adjust to the incorrect dissipation and is accumulating or losing energy at the small scales.

2.4 Why does it work?

Even if we have shown above one of the reasons why a dynamic model should work reasonably well, even if its formulation is considerably in error with respect to

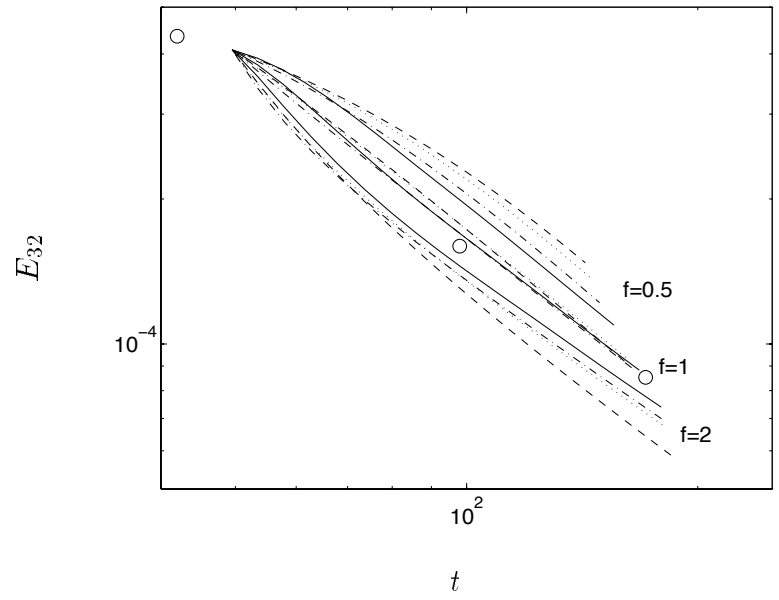

FIGURE 3. Sensitivity of energy decay to mistuning of the proportionality constant, for different "hyper-Smagorinsky" models. The central group of lines uses optimally tuned constants; the top group is modified by $f = 0.5$; the bottom one, by $f = 2$. ——— : dynamic model; - - - - : $n = 0$; ····· : $n = 1$; —·— : $n = 3$; Symbols are from the experiment of (Comte–Bellot & Corrsin 1971).

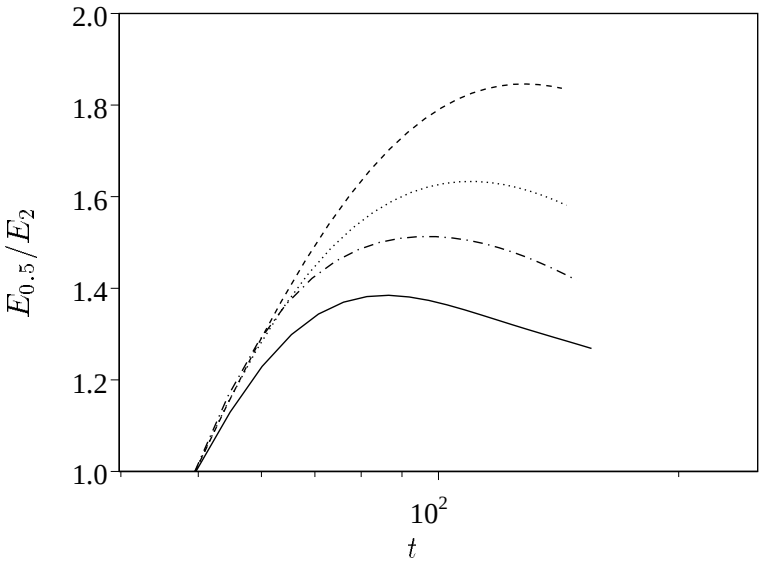

FIGURE 4. Ratio of energy obtained for different "hyper-Smagorinsky" models with $f = 0.5$ and $f = 2$. Symbols as in Fig. 3.

the true dynamics of turbulence, a simple inspection of the spectra in Fig. 2b shows that the standard formulation (5), with $f = 1$, must be very close to the "truth". The tail of its spectrum matches the experimental measurements much better than any of the modified models.

The classical explanation is, first, that the two Smagorinsky assumptions in (3) enforce a scale similarity between the two filter levels, which mimics the scale invariance in the inertial range (Germano *et al.* 1991) and, second, that the least squares approximation of (4–5) ensures that the original Smagorinsky assumptions are reasonably well satisfied (Lilly 1992).

We will argue now that this explanation is unlikely. In the first place, the Reynolds numbers in the (Comte–Bellot & Corrsin 1971) experiment are fairly low ($Re_\lambda \approx 70 - 60$), and the experimental spectra do not contain an inertial range. Their slopes are close to $k^{-4/3}$, and obtaining a computed $k^{-5/3}$ inertial range would require choosing a prefactor $f \approx 1.5$.

Next, the original stress similarity argument requires that the constant c obtained from (5) satisfies the tensor Eq. (4) in some approximate way. An approximation can be optimum and still be so bad that it makes no sense to consider that the model represents the data. This is unfortunately the case in (4). A good approximation would require that $|\lambda|^2/|L^*|^2 \ll 1$, which in turn would imply a high correlation between the tensors $-cg$ and L^*. This can be tested from the results of the calculation, and the correlation coefficient

$$\gamma = -\frac{\overline{L^* \cdot g}}{\left(\overline{|g|^2}\,\overline{|L^*|^2}\right)^{\frac{1}{2}}}, \tag{8}$$

is represented in Fig. 5. After an initial transient, it saturates around 20% and, since

$$\overline{|\lambda|^2}/\overline{|L^*|^2} = 1 - \gamma^2, \tag{9}$$

this implies that 95% of the magnitude of the stresses remain unexplained by their dynamic Smagorinsky approximation. That the optimal Smagorinsky approximation of the subgrid stresses only explains a small fraction of their magnitude was already noted by Bardina, Ferziger and Reynolds (1983).

This result shows that the Leonard stress L^* and the Germano strain g are far from being coaxial, and that there is little point in trying to model one as proportional to the other. On the other hand, the fact that the method works proves that something is being modeled. Bardina *et al.*, in the same work, noted that the correlation between the model prediction and the true dissipation is much higher than that for the stresses, and it is easy to see that (5) is actually a dissipation formula. The least square approximation results in an exact cancellation of the projection of the tensor over one of its summands, and the projection of the stress on the strain is the dissipation. In fact (5) can be rewritten as

$$\tau_g = -cg, \qquad L \cdot g = \tau_g \cdot g, \tag{10}$$

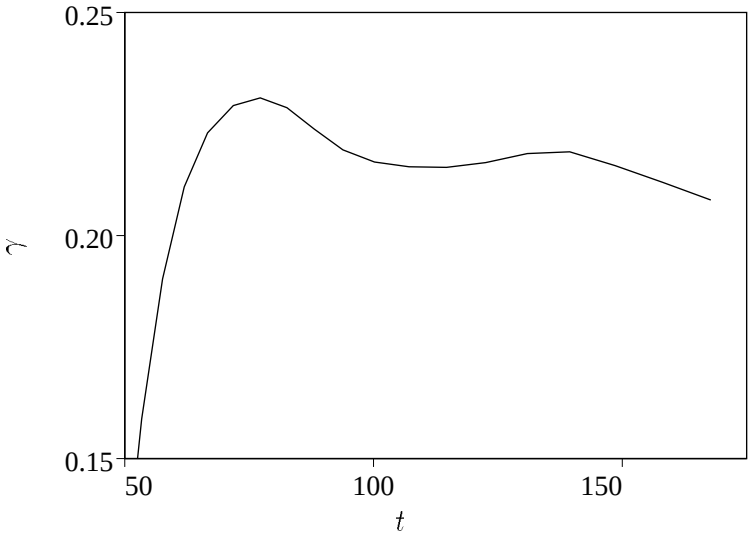

FIGURE 5. Correlation coefficient between the two tensors in (4), computed from a calculation using the dynamic model.

which says that the "dissipation" generated by the Smagorinsky stresses τ_g, is the same as the "production" of the Leonard stresses. Since, in any numerical flow without an explicit grid filter, the grid Reynolds stresses are identically zero and $L = T$, the Leonard production can be used as a surrogate for the production at the test level.

While this argument is suggestive, it is difficult to go much further. Direct computation shows that none of the actual productions and dissipations really match in the dynamic approximation. The numerical production $-T \cdot S$ remains about twice smaller than the dissipation of the Smagorinsky stresses, mainly because a substantial amount of energy is dissipated by the subgrid model on the flow scales between the test and grid filters. Other combinations can be tested with similar lack of success. While there is qualitative agreement in all the obvious balances, the quantitative details are always masked by the broad support of the second order dissipation. Equation (10), while indicative, does not seem to correspond directly to any physical property of the flow.

3. Conclusions and future work

We have shown that a large part of the good behavior of dynamical subgrid models is probably due to their robustness to approximations in the physics. This is shared by other models, with the main requirement being that the formula for the eddy viscosity contains a sensor which responds to the accumulation of energy in the high wavenumber part of the spectrum before it contaminates the energy containing range. The regular Smagorinsky model derives this property from the $|\sigma|$ factor in the eddy viscosity. The classical dynamic model is about twice less sensitive because its constant is computed exclusively from the part of the spectrum between the two filters. Any model with this feedback property, and which contains a reasonable approximation to the flow physics, is likely to represent the energy containing scales

essentially correctly. The quality of the modeling improves substantially if the last octave of the turbulent spectrum is filtered when evaluating the results, and only the large scales are kept.

All this is in addition to the main advantage of the dynamic models, which remains their ability to generate vanishing eddy viscosities in smooth flows, and their resulting good behavior near walls.

From this point of view, the use of the Smagorinsky model as the basis for the dynamic formulation is probably nonessential, and simpler formulations must exist in which the eddy viscosity is computed directly instead of through the Smagorinsky constant.

The classical justification of the dynamic model in terms of scale similarity and optimal approximation of the stresses has been examined and found weak. The approximation provided by the least square formula is so poor in practice as to make any argument based on the stresses meaningless. The least square formulation is a dissipation formula, and Lilly's formulation of the dynamic model can be understood as making the dissipation approximately equal to the measured production. The connection is, however, only approximate, and both quantities agree only qualitatively in computed flows (to within a factor of two). It should be noted that the poor prediction of the stresses, although worrying at first sight for the application to shear flows, in which the stresses are the main results of the computation, is probably not serious. The *mean* Reynolds stresses, in the same way as the total flow energy, are contained in the large flow scales and, if the latter are reasonably well predicted, the former should also be.

Further experiments are needed in cases different from the (Comte–Bellot & Corrsin 1971) decay to make sure that the specially good behavior of the spectrum for the standard model is not accidental. In the same way, tests should be undertaken with other model formulations. The main result of this note should be the realization that the present form of the dynamic model is not unique and probably not optimum, and that other formulations can be developed in terms of considerations such as numerical expedience, not necessarily fully based on strict inertial range physics.

REFERENCES

Bardina, J., Ferziger,J. H. & Reynolds, W. C. 1983 Improved subgrid-scale models based on large eddy simulation of homogeneous, incompressible, turbulent flows. *Rep.* TF-19*, Dept. Mech. Engng, Stanford, CA.*

Comte-Bellot, G. & Corrsin,S. 1971 Simple Eulerian time correlation of full and narrow-band velocity signals in grid-generated 'isotropic' turbulence. *J. Fluid Mech.* **48,** 273-337.

Germano, M., Piomelli, U., Moin, P. & Cabot, W. H. 1991 A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids.* **A 3,** 1760-1765.

Lilly, D. 1992 A proposed modification of the Germano subgrid-scale closure method. *Phys. Fluids.* **A 4,** 633-635.

Moin, P., Squires, K., Cabot, W. & Lee, S. 1991 A dynamic subgrid-scale model for compressible turbulence and scalar transport. *Phys.Fluids.* **A 3,** 2746-2757.

Rogallo, R. S. 1981 Numerical experiments in homogeneous turbulence. *NASA Tech. Mem..* **81315,** See also Canuto, C., Hussaini, M. Y., Quarteroni, A. & Zang, T. A., *Spectral methods in fluid dynamics.* Springer (1981), 85-86.