Bayesian assessment of horizontal resolution in a nested-domain WRF simulation

Michel d. S. Mesquita[1,2,*], Bjørn Ådlandsvik[3], Cindy Bruyère[4] and Anne Sandvik[2,3]
[1]Uni Bjerknes Centre, Bergen, Norway; [2]Bjerknes Centre for Climate Research, Bergen, Norway;
[3]Institute of Marine Research, Bergen, Norway; [4]National Center for Atmospheric Research, Boulder, USA; *contact info: michel.mesquita@uni.no

1. Motivation

- High-resolution data can provide added information to the study of complex topography regions such as the Norwegian fjords (Heikkilä et al. 2011; Myksvoll et al. 2012)
- To make statistical inference about a model simulation, one needs a large sample to produce robust statistics (Lopez et al. 2006). Producing large samples at high-resolution can become computationally expensive. This is especially the case when testing different combinations of parameterization schemes
- Here, we present an alternative approach to analyzing output from limited area models based on Bayesian probability. This approach allows for the use of small samples to make inferences about the statistical population

2. Data and Methods

- **Model:** Weather Research and Forecasting (WRF) model version 3.1
- **Resolution:** parent domain at 9 km resolution and two nested domains at 3 km and 1 km, respectively (with feedback=1, two-way nesting) - see Fig. 1
- **Vertical levels:** 31
- **Parameterization schemes:** WRF Single-Moment 3-class scheme (mp physics=3); cumulus parameterization was turned off (cu physics=0); Yonsei University longwave scheme (bl_pbl_physics=1); RRTM scheme (ra_lw_physics=1); Dudhia shortwave scheme (ra_sw_physics=1)
- **LBC:** ERA-interim Re-Analysis obtained from the ECMWF Data Server
- **Other details:** the simulation was run from 2007 to 2009. April, May and June of 2008 and 2009 were retained for the analysis. Results are shown for the three-hourly 2 m temperature in April. Box selected for spatial averaging: 59.32N, 60.75N and 5.05E, 7.90E (Hardanger fjord region)
- **Prior:** Kvamsøy weather station (60.358N and 6.275E). Data were obtained from the Norwegian Meteorological Institute data server at eklima.no. Average surface temperature for April (2003-2011): 7.48±1.27°C

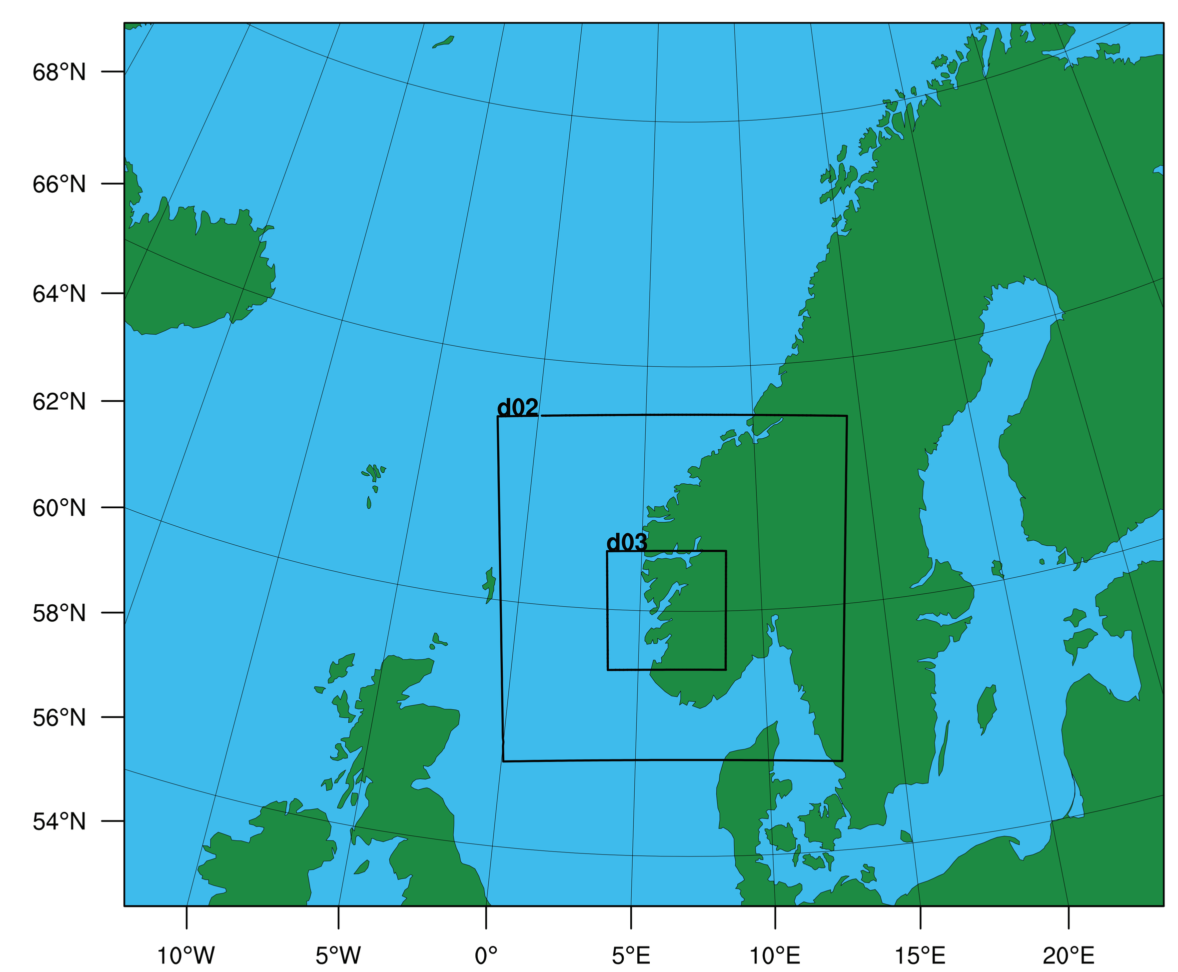

Figure 1 - WRF model domain setup: parent domain at 9 km (outer domain), nest at 3 km (d02) and nest at 1 km (d03).

3. Results

- Figure 2 shows the Monte Carlo samples from the joint distributions of the population mean and variance. The ERA Interim distribution (ERAi), on the top left, shows larger spread both for the mean and the variance as compared to the three domains. The 9 km domain seems to be off and does not match ERAi. The 3 km nest shows the closest approximation to ERAi, whereas the 1 km nest approximates the variance more closely.
- Figure 3 shows the marginal distribution of the mean, based on the Monte Carlo sampling. Red line indicates the mean value of the ERAi marginal distribution. Posterior bounds of the 9 km parent domain do not contain the ERAi mean. Table 1 shows that even though there is some overlap between the ERAi posterior bounds and the 9 km domain, this overlap is minimum. The 3 km and 1 km nests show a closer overlap with ERAi. The 3 km resolution domain is able to approximate the mean more realistically, also confirmed by the posterior bound overlap with ERAi (Table 1).
- The marginal distribution of the ERAi variance is approximated more closely by the 1 km resolution domain (Fig. 4). The mean value of the ERAi marginal distribution is within the posterior bounds for that resolution. In contrast, the 9 km and 3 km domains have posterior bounds outside of the ERAi mean value. There is, however, a better overlap between ERAi and the 3 km posterior distribution (Table 1).

Table 1. Posterior distribution summary for the mean (θ) and variance (σ^2) based on Monte Carlo sampling. The 95% posterior bound (PB) is also indicated for each variable. Temperature units given in degrees Celsius.

	θ	θ PB	σ^2	σ^2 PB
ERAi	4.26	(3.87, 4.66)	9.90	(8.30, 11.93)
d01	4.80	(4.57, 5.04)	7.08	(6.26, 8.07)
d02	4.19	(3.93, 4.44)	8.04	(7.12, 9.14)
d03	4.56	(4.29, 4.83)	9.19	(8.11, 10.46)

The Bayesian model

In this study, the Bayesian model is applied to the 2m temperature in the Hardanger fjord region. It considers the case in which the mean (θ) and variance (σ^2) are unknown (Hoff 2009; Gelman et al. 2004). For the joint prior distribution $p(\theta, \sigma^2)$ for θ and σ^2, the posterior inference will use Bayes' rule, as shown in Equation 1:

$$p(\theta, \sigma^2 \mid y_1, \dots, y_n) = \frac{p(y_1, \dots, y_n \mid \theta, \sigma^2) p(\theta, \sigma^2)}{p(y_1, \dots, y_n)} \quad (1)$$

where $y_1, \dots, y_n$, represent the data. Since the joint distribution for two quantities can be expressed as the product of a conditional probability and a marginal probability, the posterior distribution can likewise be decomposed (Eq. 2):

$$p(\theta, \sigma^2 \mid y1, \dots, y_n) = p(\theta \mid \sigma^2, y_1, \dots, y_n) p(\sigma^2 \mid y_1, \dots, y_n) \quad (2)$$

where the first part of the equation is the conditional probability of θ on the variance and the data; and the second part is the marginal distribution of σ^2. The conditional probability part of the equation can be determined as a normal distribution:

$$\{\theta \mid y_1, \dots, y_n, \sigma^2\} \sim normal(\mu_n, \sigma^2/\kappa_n) \quad (3)$$

Where $\kappa_n = \kappa_0 + n$ represents the degrees of freedom (df) as the sum of the prior df (κ_0) and that from the data (n). μ_n is given by: $\mu_n = \frac{(\kappa_0/\sigma^2)\mu_0 + (n/\sigma^2)\overline{y}}{\kappa_0/\sigma^2 + n/\sigma^2} = \frac{\kappa_0 \mu_0 + n\overline{y}}{\kappa_n}$, where $\overline{y}$ is the sample mean taken from the WRF simulation. The prior mean is given by μ_0. The calculation of σ^2 is explained next.

The second part of equation 2, the marginal distribution of σ^2, can be obtained by integrating over the unknown value of the mean, θ, as follows:

$$p(\sigma^2 \mid y_1, \dots, y_n) \propto p(\sigma^2) p(y_1, \dots, y_n \mid \sigma^2) \quad (4)$$

$$= p(\sigma^2) \int p(y_1, \dots, y_n \mid \theta, \sigma^2) p(\theta \mid \sigma^2) d\theta \quad (5)$$

Solving the integral, and considering the precision ($1/\sigma^2$) such that the distribution is conjugate, gives the following gamma distribution:

$$\{1/\sigma^2 \mid y_1, \dots, y_n\} \sim gamma(\nu_n/2, \nu_n \sigma_n^2/2) \quad (6)$$

Where $\nu_n = \nu_0 + n$ is the sum of degrees of freedom of the prior (ν_0) and of the data (n). σ_n^2 is given by $\sigma_n^2 = \frac{1}{\nu_n}[\nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_n}(\overline{y} - \mu_0)^2]$, where $\overline{y}$ is the sample mean and s^2 is the sample variance, both taken from the WRF simulation. σ_0^2 is the prior variance.

Samples of θ and σ^2 can be generated from their joint posterior distribution using the following Monte Carlo procedure (Hoff 2009):

$$\sigma^{2(1)} \sim inv\ gamma(\frac{\nu_n}{2}, \frac{\sigma_n^2 \nu_n}{2}), \quad \theta^{(1)} \sim normal(\mu_n, \frac{\sigma^{2(1)}}{\kappa_n})$$

$$\vdots$$

$$\sigma^{2(S)} \sim inv\ gamma(\frac{\nu_n}{2}, \frac{\sigma_n^2 \nu_n}{2}), \quad \theta^{(S)} \sim normal(\mu_n, \frac{\sigma^{2(S)}}{\kappa_n})$$

where σ^2 is estimated using an inverse-gamma distribution (*inv gamma*). Each $\theta^{(S)}$ is sampled from its conditional distribution given the data and $\sigma^2 = \sigma^{2(S)}$. The simulated pairs of $\{(\sigma^{2(1)}, \theta^{(1)}), \dots, (\sigma^{2(S)}, \theta^{(S)})\}$ are independent samples of the joint posterior distribution, i.e.: $p(\theta, \sigma^2 \mid y_1, \dots, y_n)$. The simulated sequence $\{\theta^{(1)}, \dots, \theta^{(S)}\}$ can be seen as independent samples from the marginal posterior distribution of $p(\theta \mid y_1, \dots, y_n)$, and so this sequence can be used to make Monte Carlo approximations to functions involving $p(\theta \mid y_1, \dots, y_n)$. While $\theta^{(1)}, \dots, \theta^{(S)}$ are each conditional samples, they are also each conditional on different values of σ^2. Together, they make up marginal samples of θ.

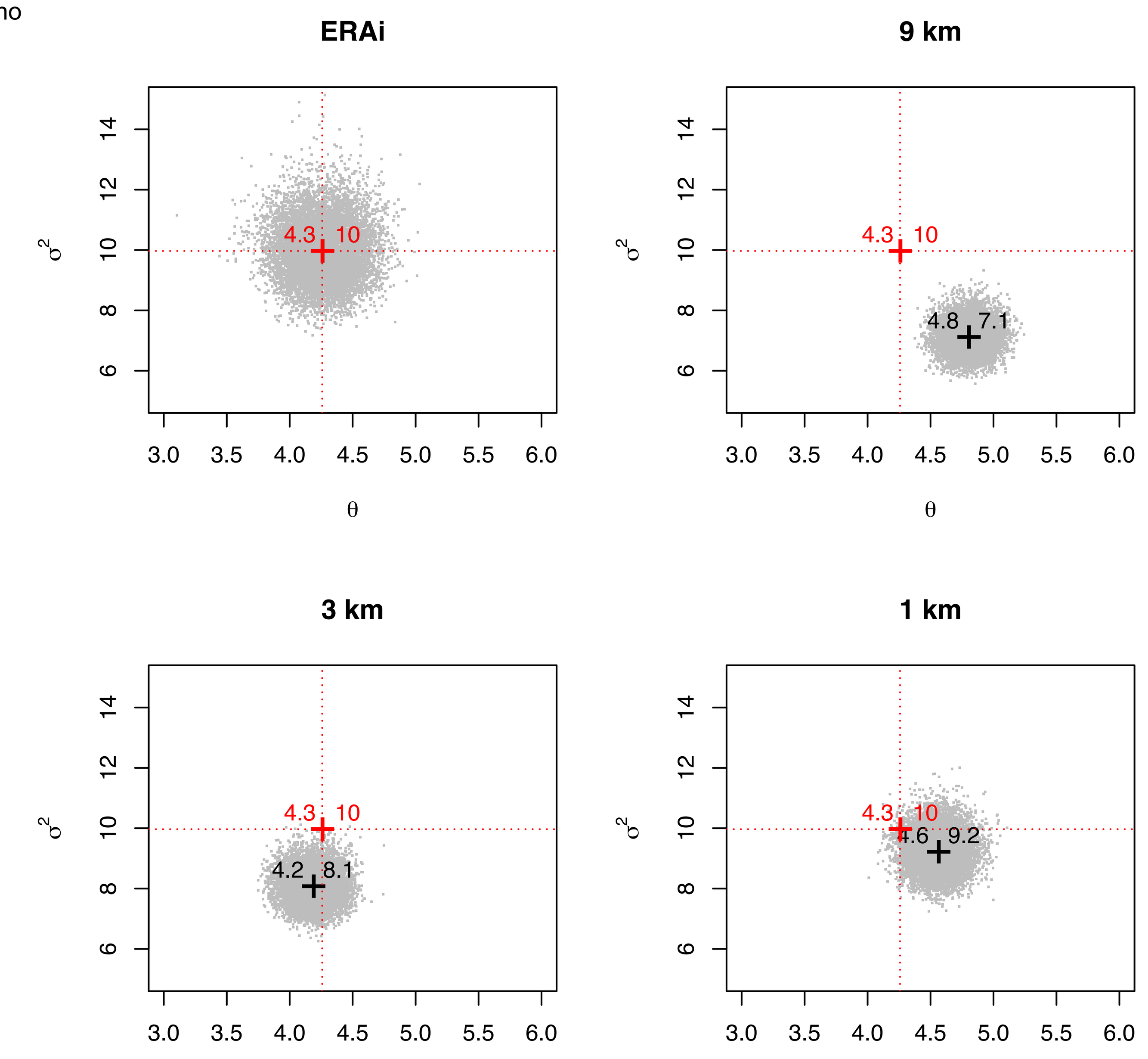

Figure 2 - Monte Carlo samples from the joint distributions of the population mean (θ) and variance (σ²) for ERA Interim (ERAi) and for the different domains. The values in black show the mean value of the population mean (right side) and of the population variance (left side). Mean values of θ and σ² for ERA Interim are indicated in red. Temperature given in degrees Celsius.

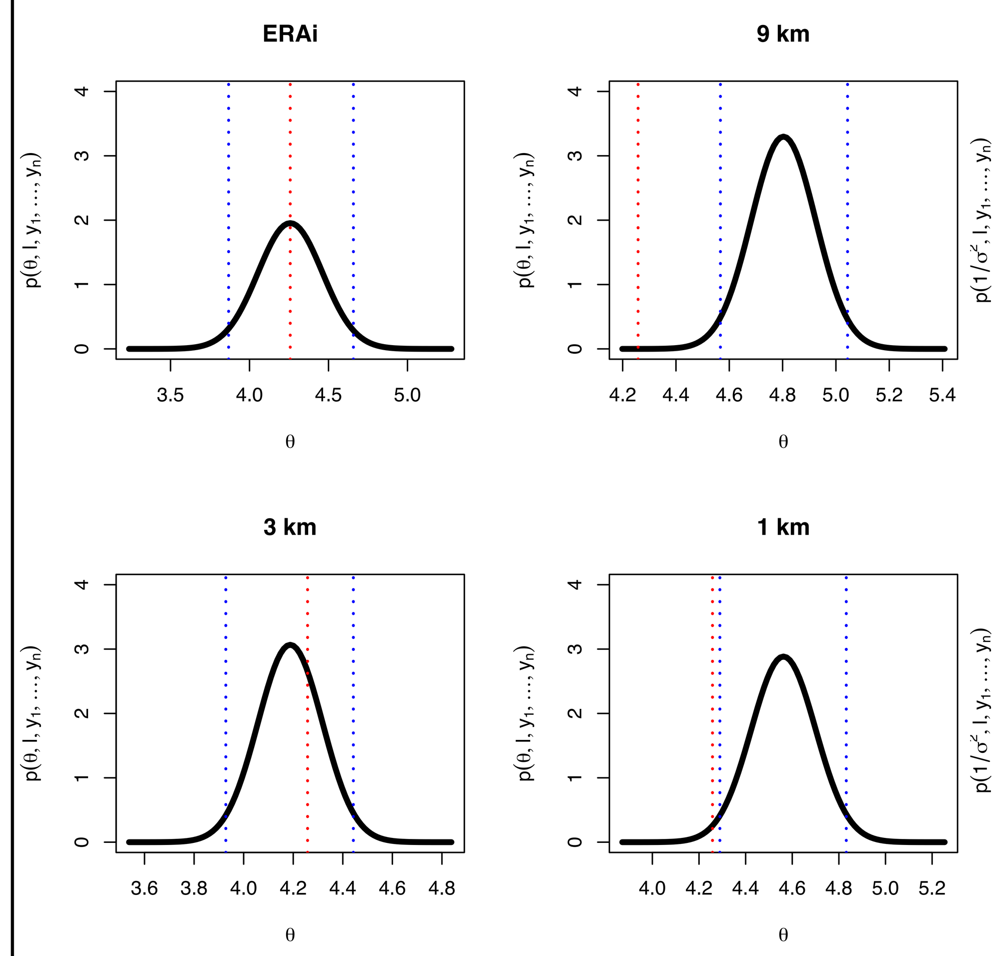

Figure 3 - Monte Carlo samples from the marginal distribution of θ for ERAi and for the different domains. The blue vertical lines give a 95% quantile-based posterior bound. The red line indicates the mean value of the ERAi posterior marginal distribution. Temperature given in degrees Celsius.

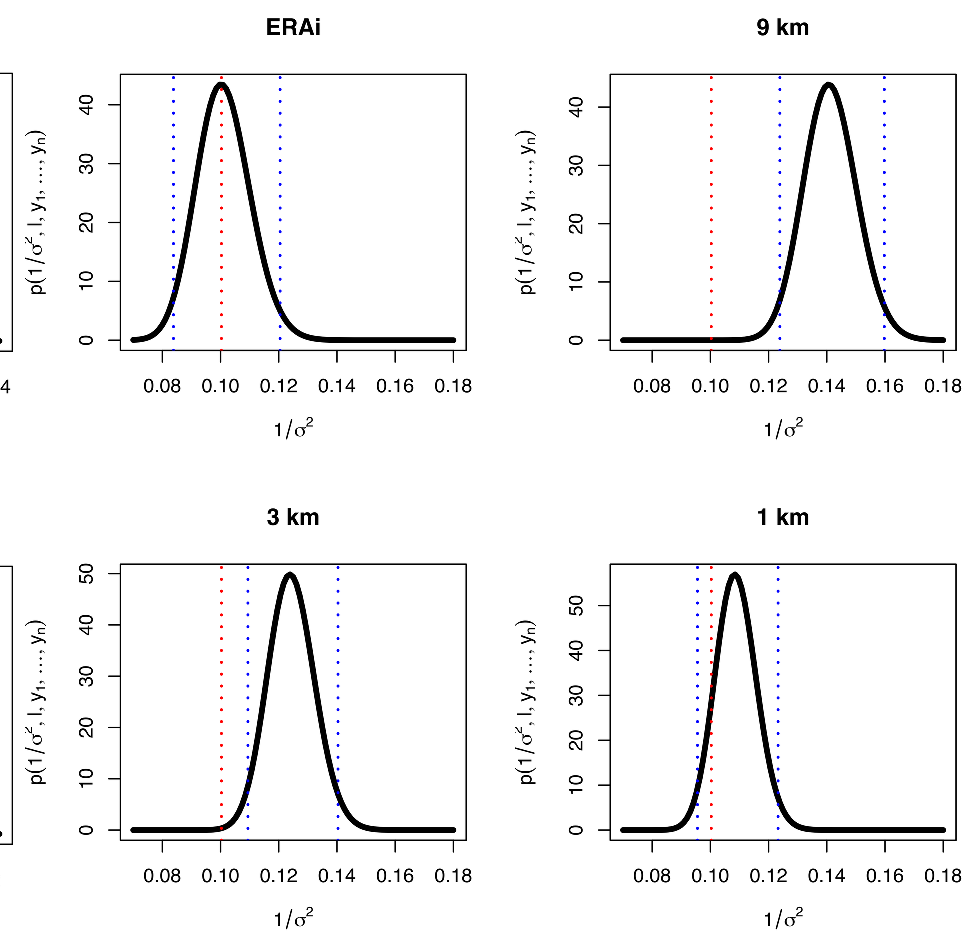

Figure 4 - The same as Figure 3, but for the precision, 1/σ².

4. Conclusion

The increased horizontal resolution is able to approximate the mean and the variance of the observations more closely. The Bayesian model provides a richer probabilistic view of the dataset and it obviates the use of long simulations for estimating the population mean or variance - thus saving computational resources. If one is to use standard statistics, a larger sample is needed to be able to make robust inferences. Hence, through the use of prior information, the Bayesian framework provides an alternative approach to estimating the statistical population, and in this case, for assessing the bias in the model simulation. It is also useful for sensitivity studies where one needs to compare not only resolution, but also the use of different parameterization schemes. This approach can also be applied to other variables by adapting it to their underlying distribution.

References

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004: Bayesian data analysis. 2d ed., Chapman & Hall/CRC, 668 pp.
Heikkilä, U., A. D. Sandvik, and A. Sorteberg, 2011: Dynamical downscaling of era-40 in complex terrain using the wrf regional climate model. Clim. Dyn., 37, 1551– 1564.
Hoff, P. D., 2009: A first course in Bayesian statistical methods. Springer, 270 pp.
Lopez, A., C. Tebaldi, M. New, D. Stainforth, M. Allen, and J. Kettleborough, 2006: Two approaches to quantifying uncertainty in global temperature changes. J. Cli- mate, 19, 4785–4796.
Myksvoll, M. S., A. D. Sandvik, J. Skardhamar, and S. Sundby, 2012: Importance of high resolution wind forcing on eddy activity and particle dispersion in a norwegian fjord. Estuar. Coast. Shelf Sci, submitted.

Acknowledgments

We would like to thank NCAR for making the WRF model publicly available. We also thank ECMWF and the Norwegian Meteorological Institute for the datasets provided. This study has been funded through the Downscaling Synthesis project at the Bjerknes Centre for Climate Research, Bergen, Norway.