Questions and Answers to Lecture 8

1.) Did I understand it correctly, that in the Ornstein-Uhlenbeck process, the parameter fluctuations are modelled with a normal (resp. lognormal) distribution?

Very strictly no, but still very close. The Ornstein-Uhlenbeck process is a continuous-time Gaussian process. A Gaussian process is a generalization of a normal distribution to a stochastic process. A stochastic process consists of random variables defined for any value of a continuous variable, in our case time (a normal distribution would not have such a continuous variable). A Gaussian process is defined by the property, that when evaluated at a finite set of values of the continuous variable, it is characterized by a multivariate random variable. The Ornstein-Uhlenbeck is also a Markov process which means that the conditional distribution of future states is unique given a state at a certain point in time (we do not need to know the past, just the present to predict the future).
For these two reasons, we can fully characterize the process by the equations (9.7) – (9.9). Given the state at time t0, the expectation exponentially approaches the mean. The variance increases from zero at time t0 (because we know the state) asymptotically to σ^2, the covariance is given by (9.9) and the distribution is normal.
For positive parameters we assume the log is an Ornstein-Uhlenbeck process.

2.) I am unsure about the aim of the MCMC for posterior sampling. To we want to get the distribution of the parameters that cover the output distribution?

The goal of Bayesian inference (in the sense we apply it here) is to learn about model parameter values from observed data. The posterior distribution thus combines our prior knowledge with information gained from new data. It is given by equation (10.8). Clearly, when simulating the model with parameters from the posterior distribution the model output will be close to the data (was this your question?).
The numerical problem is that the distribution (10.8) is hardly tractable. MCMC is an elegant way to sample from it. We can then represent the distribution approximately by this sample. We can then use the sample to derive means, variances, correlations, marginals, etc. of the parameters and to propagate it through the model to get a sample of results to be used to derive means, variances, correlations, marginals, etc. or the results

3.) Within the MCMC the word proposal distribution is used. Does this distribution concern the parameters or the output? And where does this distribution come from?

This is part of the specific Metropolis MCMC algorithm. It is used to propose a new point of the parameters sample in parameter space. Nearly any symmetrical distribution would work in principle. Conceptually, the choice is thus not relevant. However, the performance of the algorithm depends crucially on this choice. This is illustrated in the Figures 10.2 and 10.3.

4.) Within the MCMC we always talk about sampling from the posterior distribution to get a next value in the Markov chain. I don't really understand, what this means?

MCMC is a method to draw a correlated sample from a distribution. This would not have to be a posterior in Bayesian inference. Working with samples of a distribution is an often used numerical approach for distributions that are not easy to deal with analytically. As explained above, samples can then be used to calculate distributional properties and to be further propagated to derive samples e.g. of model outputs that can again be used to derive properties of the model output distribution. In MCMC the sample is derived point by point. Starting from an initial value, we sample the next point and so on. In the Metropolis technique, this requires a proposal distribution and an acceptance/rejection step as outlined by the steps 1 to 4 on pages 156-157. As we only need the ratio of the probability densities between the proposal point and the current point (equation 10.17) this technique is particularly interesting for sampling posterior distributions in Bayesian inference, as the unknown normalization constant (the integral in the denominator of eq. 10.8) cancels out.

5.) In case of modelling the uncertainty, we take parameters from a Gaussian distribution. Is this just the distribution we take to approximate the fact, that we don't exactly know, which parameter is the true one?

Yes. We assume that we can represent our uncertain knowledge about the true value by a normal resp. lognormal distribution.

6.) In Exercise 6 - Questions - question 4. Is this aiming at, that it is important to have the log so we don't receive negative values or that it is important to pay attention to the transformation as it is not u = exp(u) but u = exp(m+s^2/2).

This question is about making non-negative parameters stochastic. Without transformation, an Ornstein-Uhlenbeck process has always finite probability or reaching negative values which do not make sense for non-negative parameters (e.g. a negative death rate would produce algae from POM which does not make sense). Nevertheless it is also good if you recall that exp(m) will not be the mean of the back-transformed process. This applies also to uncertain, not just to stochastic parameters.