1. Line 476: I would cite the reference as (e.g., Oldenburg and Li, 2005) as there are many references describing the concept of inversion. You could actually add some more general references such as (Menke, 1984 – discrete inverse theory; Tarantola – inverse problem theory, 1987 …)
2. Line 496: after "from a sedimentary basin" add "of given geometry and/or density model"
3. Line 502: I would be more quantitative and instead of "equally well" put "with the same degree of accuracy"
4. Line 521: you are introducing the vector parameter $\vec{p}$. This assumes that you deal with a discrete problem, where the continuous parameter manifold is approximated by its discrete version. Because you describe the discretization of your forward model later in this section, it would be a good idea a this stage to explain the reader that your problem is a discrete one, where a finite set of parameters is used to approximate the real model in a discrete sense. Perhaps you should spend few words to explain the difference between discrete and continuous inverse problems, just for the unexperienced readers
5. Eq. 3.1: $f(\vec{p})$ is a vector, i.e., $\vec{f}(\vec{p})$, otherwise the equation is inconsistent. Furthermore, in eq.3.2 you introduce the index i which is a vector index for $\vec{f}(\vec{p})$. But again, there is an inconsistency in eq. 3.2 where you lose the vector notation for f. Define the dimension M (parameters) and N (data) at this stage, which makes perfect sense if you are describing a discrete problem.
6. Line 534: because $\vec{f}(\vec{p})$ is a vector, the use of the transpose of a scalar to demonstrate your equation is incorrect (even if the result is correct). Use Einstein notation with proper index i and j as follows
$$\frac{\partial\phi(\vec{p})}{\partial p_i} = \frac{\partial}{\partial p_i}\sum_j \left[d_j^{obs} - f_j(\vec{p})\right]\cdot\left[d_j^{obs} - f_j(\vec{p})\right] = \sum_j \frac{\partial}{\partial p_i}\left[d_j^{obs} - f_j(\vec{p})\right]\cdot\left[d_j^{obs} - f_j(\vec{p})\right]$$
$$= \sum_j -\frac{\partial f_j(\vec{p})}{\partial p_i}\cdot\left[d_j^{obs} - f_j(\vec{p})\right] - \left[d_j^{obs} - f_j(\vec{p})\right]\cdot\frac{\partial f_j(\vec{p})}{\partial p_i} = -2\sum_j \frac{\partial f_j(\vec{p})}{\partial p_i}\cdot\left[d_j^{obs} - f_j(\vec{p})\right]$$
which becomes a matrix equation
$$\vec{\nabla}\phi(\vec{p}) = -2J(\vec{p}) * \left[\vec{d}^{obs} - \vec{f}(\vec{p})\right]$$
where $\vec{p}$ has dimension M and $\vec{d}^{obs}$ has dimension M, and the MxN Jacobian matrix J is given by
$$J_{ij} = \frac{\partial f_j(\vec{p})}{\partial p_i} = \begin{pmatrix} \frac{\partial f_1}{\partial p_1} & \cdots & \frac{\partial f_N}{\partial p_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial p_M} & \cdots & \frac{\partial f_N}{\partial p_M} \end{pmatrix}$$
7. Eq. 3.9 is linear, i.e., a solution can be found by using the inverse or pseudo-inverse of matrix J. However, since the equation has a non-trivial kernel we use a minimum-norm regularized solution via lsqr (which is iterative)
8. Line 557: if the problem is linear the derivative is constant (not linear), i.e., the derivative does not depend from the parameters (density in your case)
9. Line 562: I disagree, non-linear problems are very different from linear problems. However, they can be "linearized" in proximity of the minimum of the likelihood function and can be solved by iteration.
10. Line 585: now you describe your discretization model. I would have introduced the concept of discrete inversion before (see point .4) and link to this discussion.
11. Line 641: not really a complete Bouguer correction. Topography, which is the main input of Bouguer correction, is your unknown. You subtract a series of effects from your measured gravity (including topography around the inversion area) to isolate the effect of topography from your data, and you correct long-wavelength effects from deep sources by subtracting the regional field.
12. Line 721: inherit is inherent?
13. Line 722: after "regularization" add "i.e., a unique solution is determined subject to minimizing an additional objective functional".

14. Line 727: an obvious question here is why you haven't used the analytical expression of the derivative of the prism gravity equation. You have to explain that this can be calculated but the final expression is very complicated and would slow down your inversion significantly.
15. Line 768: explain that LSQR gives the minimum-norm solution, i.e., amongst the set of solutions which fit the data within the same level of accuracy, the one with minimum $||p||^2$ is chosen.
16. Line 848: does it happen? Do you get "runaway" solutions?
17. Line 1017: I cannot understand from Fig.3.14 why gridding with equivalent sources produces a better result relative to gridding with the other technique.
18. Line 1908: this is the key. You calculate Bouguer corrections using data outside of the inversion domain, i.e., this is not really a Bouguer correction but you are rather subtracting e series of effects to isolate the contribution of topography!