

Avaliação e Correção de Dados do Geobank em Rondônia

Geólogo Leandro Guimarães da Silva
(leandro.guimaraes@cprm.gov.br)
Pesquisador em Geociências
Matr. 31.835.341

Residência de Porto Velho (REPO)

Revisão: Dalton Rosemberg Valentim da Silva

Com colaboração de:

Jordan Feijó
Ueliton Bento da Silva
Dalton Rosemberg Valentim da Silva

1. Introdução

Este relatório visa registrar a metodologia utilizada para corrigir o posicionamento de parte das amostras e pontos de observação (afloramentos) do Projeto Platina Rondônia, registrados no Geobank. A metodologia engloba princípios de arquivologia, geoprocessamento e banco de dados e merece extrema atenção na sua execução, pois percorre extenso caminho e com vários pontos de risco para ocorrência de erros. A motivação para a criação desta metodologia foi a identificação de deslocamentos significativos de pontos de coleta de amostras individuais, além de distorções aleatórias e deslocamentos de malhas de amostragem geoquímica, encontrados durante o resgate de dados dos projetos antigos arquivados no Geobank.

Tendo a CPRM centenas de projetos executados desde sua fundação, nos anos 1960, é inestimável o valor dos dados obtidos ao longo do tempo, considerando recursos financeiros, humanos e tecnológicos dispensados para o desenvolvimento do País. Adiciona-se a isto o eventual custo de revisita a áreas já estudadas e a definição crescente de Unidades de Conservação, que apresentam acessibilidade restrita, fazendo com que o valor dos dados históricos aumente significativamente.

Porém, nem sempre estes dados estão disponíveis para os usuários da melhor forma. Ainda, é comum a ocorrência de erros e inconsistências em bancos de dados, principalmente quando grandes quantidades de dados têm seu formato convertido de analógico para digital, suas bases são migradas entre plataformas e mesclas de bancos com estruturas e modelagens distintas são efetuadas. O Geobank, banco de dados da CPRM, armazena dados de projetos de várias localidades do país e com diversas origens são armazenados, sendo um dos maiores repositórios de dados geológicos do Brasil.

Espera-se com este relatório que as unidades regionais da CPRM tenham uma referência de procedimento para que possam aplicar sistematicamente as correções necessárias para ajuste dos dados e, com isso, melhorar o planejamento de campo no que se refere ao conhecimento prévio das áreas a serem investigadas, além de retornar os dados corrigidos para a DIGEOP proceder com a correção no Geobank. É recomendada a descentralização da correção destes dados pelas Unidades Regionais (UR), pois as cadernetas, mapas e até os participantes dos projetos históricos tendem a manter contato com a UR de origem. Assim, os dados corrigidos poderão ser disponibilizados à comunidade em geral, com confiabilidade incrementada e suportada por princípios de arquivologia, que requerem respaldo documental de dados eletrônicos.

2. Exposição do problema

Os dados digitais muitas vezes representam registros obtidos de documentos em papel, que registram o mundo real observado por alguém em um determinado momento. No caso dos levantamentos geológicos, o mundo real pode ser entendido a partir das observações feitas em campo, registradas em cadernetas e depois convertidas em mapas, que por sua vez se tornam generalizações do contexto geológico estudado em uma determinada área.

Ao iniciar os projetos do calendário de projetos 2015 da Residência de Porto Velho, foi feita a recuperação de parte dos dados históricos da CPRM para auxiliar no planejamento de etapas de campo. Porém, foram verificados deslocamentos de malhas e pontos do Projeto Platina, desenvolvido já na década de 1990, de forma que seu uso ficou limitado. Foram identificados deslocamentos de até 200 km em pontos esporádicos (Figura 2-1), além de distorções em malhas e também deslocamento das mesmas, que serão descritas a seguir.

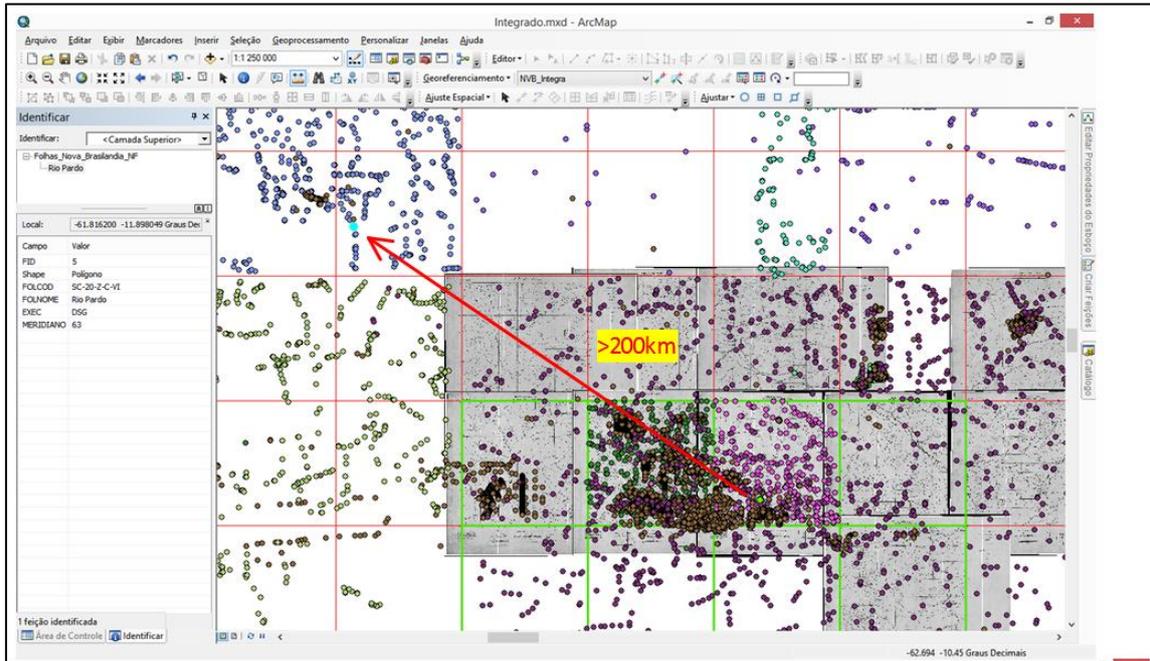


Figura 2-1: Ponto de mapeamento com deslocamento em torno de 200 km (Projeto Sudeste de Rondônia, 1977).

Deslocamentos de pontos em bancos de dados da dimensão do Geobank podem ser considerados aceitáveis, dada a complexidade das etapas envolvidas na conversão, transcrição e processamento dos dados, além dos métodos de posicionamento utilizados à época dos levantamentos. Porém, o esperado é que o banco tenha o mínimo de erros possíveis, garantindo a qualidade dos dados disponíveis. Grande parte dos pontos de observação teve seu posicionamento alterado em relação à documentação original dos projetos e apresentam diferenças aparentemente aleatórias em distância e azimuth (Figura 2-2) quando plotados em mapa.

No caso das malhas, as inconsistências se referem a distorções na geometria original da malha, além de deslocamento variável no conjunto de pontos (Figura 2-3 e Figura 2-4).

distribuição dos pontos na folha (em verde). Este problema foi identificado em todas as folhas que apresentam pontos de observação (afioramento) do referido projeto.

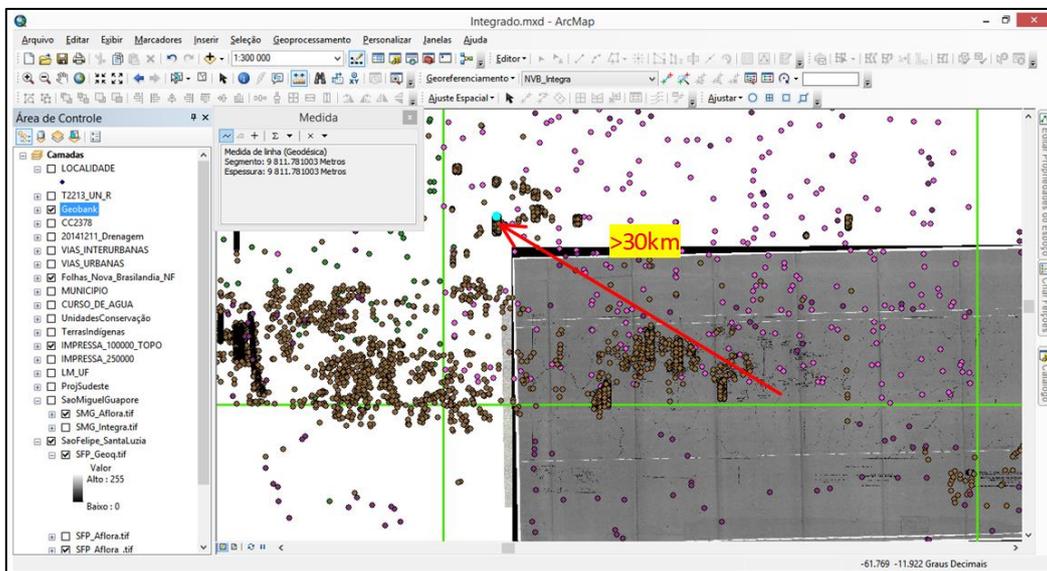


Figura 2-4: Deslocamento de malha na ordem de 30 km.

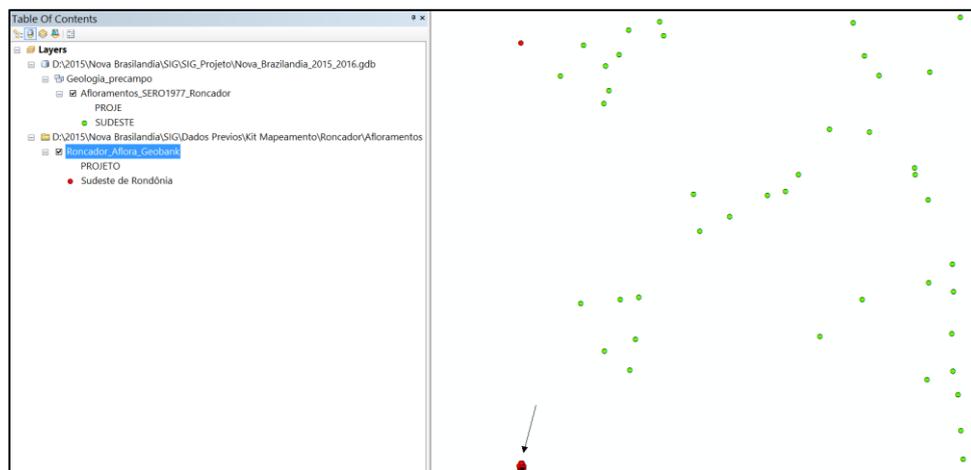


Figura 2-5: Deslocamento na localização de todos os pontos de afloramentos (seta) do Projeto Sudeste de Rondônia (1977) recuperados do Geobank, na área da Folha Roncador.

Para minimização dos erros encontrados, foi elaborada uma sequência de passos a serem seguidos, de forma que os dados possam ser aproveitados com minimização dos problemas de deslocamento encontrados. Também, com base neste estudo, sugestões são feitas para que os erros antigos sejam corrigidos no Geobank.

3. Metodologia

Esta metodologia não pretende identificar a causa dos deslocamentos, se limitando a uma sugestão de trabalhos sequenciais para resolução do problema, com base em material documental (mapas). Também, este trabalho não prevê correções dentro do Geobank, apesar de ser apresentada uma proposta de estratégia para validação sistemática desses dados nas

unidades regionais. Os dados utilizados durante a elaboração deste trabalho e desenvolvimento/aplicação da metodologia descrita em seguida são referentes ao Projeto Platina Rondônia, apesar da ocorrência de deslocamento não estar restrita a um único projeto, como já foi demonstrado acima.

3.1. Resgate dos Dados

A metodologia inicia com o inventário da documentação existente dos projetos a serem avaliados, incluindo cadernetas, relatórios e mapas, parciais e finais, além de arquivos digitais, tabelas, registros e anotações em papel. Adicionalmente, para melhor entendimento do andamento dos projetos à época de sua execução, entrevistas com os participantes dos projetos, mesmo que não façam mais parte do quadro ativo de funcionários da CPRM é recomendada. A memória viva dos projetos é ainda a melhor fonte de informações complementares que não fazem parte dos relatórios, por mais detalhados que sejam. Esta estratégia é essencial para um rápido entendimento do que o projeto gerou de dados, pois muitas vezes atividades paralelas são efetuadas, mas nem sempre registradas adequadamente.

A utilização de entrevistas foi utilizada com sucesso em julho de 2014 durante o resgate dos dados históricos para montagem do banco de dados do Projeto Carvão-Área Chico Lomã-Morungava, dentro do Projeto Patrimônio. Além de extensas consultas aos funcionários do quadro atual (bibliotecária, geólogos, sondadores, entre outros), dois dos coordenadores do projeto, atualmente fora do quadro, foram convidados à SUREG-PA para tirarem dúvidas dos pesquisadores envolvidos no resgate dos dados. A contribuição foi significativa, para entendimento das codificações utilizadas, interrupções, mudanças de escopo e estratégia ocorridos no projeto original, durante os anos 70 e 80.

O uso de mapas históricos na recuperação de dados de posicionamento também foi utilizado no Projeto Patrimônio, no alvo Miriri (no prelo). Os furos foram reposicionados de acordo com a malha indicada nos mapas, melhorando a distribuição dos teores e a coerência geológica, uma vez que as coordenadas indicadas nos logs de sondagem continham erros diversos em boa parcela dos furos.

3.2. Digitalização e Georreferenciamento dos Mapas

Inicialmente, os mapas foram escanerizados em resolução compatível com a qualidade do documento e georreferenciados no sistema de coordenadas original. Vale ressaltar que ocorreram pelo menos seis mudanças de *data* ao longo da história cartográfica brasileira, sendo os mais comuns o Córrego Alegre e o SAD69, ressaltando que estes passaram por mudanças em seus parâmetros durante sua vigência como *datum* oficial.

Os mapas foram georreferenciados no aplicativo RSI ENVI, com pontos de controle coletados a partir de cruzetas de coordenadas, além de confluência de drenagens importantes. A base utilizada como referência foi o conjunto de cartas 1:100.000 do DSG (Diretoria de Serviço Geográfico do Exército), baixadas diretamente da página da instituição (Figura 3.2-1). Adicionalmente, foram utilizados os vetores de drenagens gerados pela SEDAM (Secretaria de

Estado do Desenvolvimento Ambiental), a partir da mesma base do DSG, para o Estado de Rondônia. Ao fim do georreferenciamento, todos os mapas foram convertidos para o *datum* SIRGAS-2000.

O uso de imagens de satélite foi considerado ineficiente, pois diversos rios e igarapés ficam desfigurados por causa do desmatamento ocorrido nos últimos anos, além de outras intervenções antrópicas como construções de barragens e usinas de energia. Além deste fato, os mapas da CPRM foram feitos utilizando-se a base do DSG, o que reforça a necessidade de se utilizar a base original.

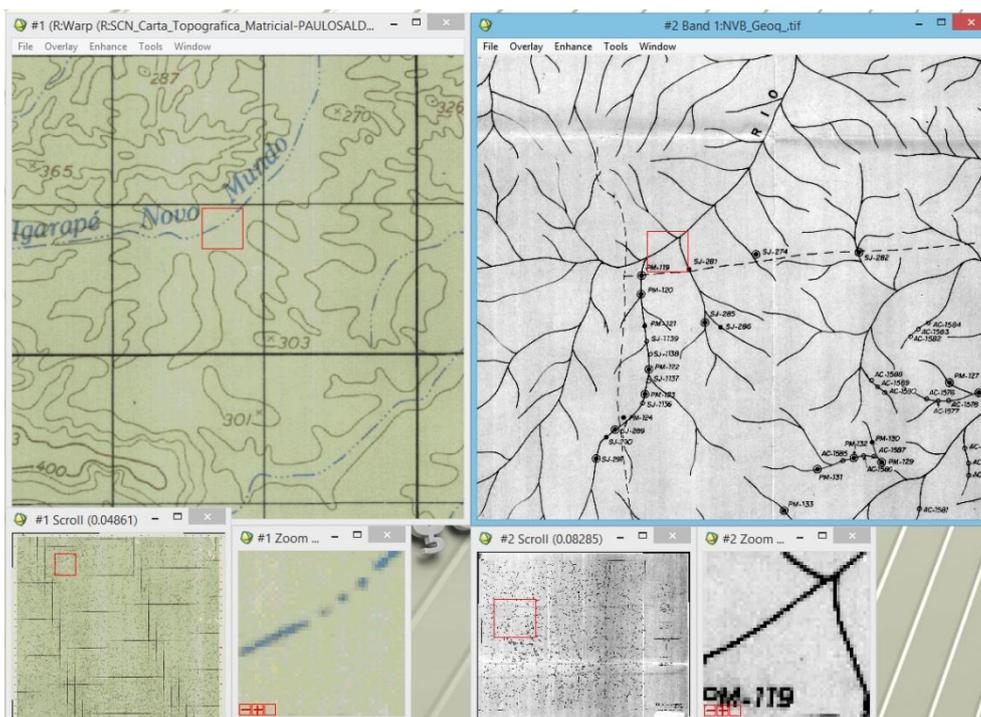


Figura 3.2-1: Telas de georreferenciamento do ENVI. Cartas topográficas do DSG e mapa de projeto da CPRM, Projeto Platina.

3.3. Vetorização dos Pontos

Após o georreferenciamento, foram criados no aplicativo ArcGIS 10.2 arquivos de vetor (*shapefiles*), cujos nomes foram compostos por: centro de custo do projeto, sigla do mapa ou área investigada, sigla e classe da amostra. Os pontos foram coletados de forma que na tabela de atributos apenas duas colunas (campos) fossem necessárias: número e observações. O primeiro indica o número sequencial para o coletor e o segundo alguma inconsistência ou informação adicional que tenha sido observada durante o processo. É importante ressaltar também que os mapas dos projetos às vezes apresentam erros de grafia, sendo mais um ponto de cuidado na hora de vetorizar os pontos de amostragem. Nestes casos, foi inserido algum comentário indicando que há erros no mapa original.

Agrupar os arquivos de vetor com base no coletor e classe de amostra foi importante para viabilizar a vetorização dos pontos por vários usuários, pois os mapas diferenciam símbolos para classes de amostras coletadas por um mesmo coletor (Figura 3.3-1). Ou seja, mais de um

usuário pode selecionar um símbolo para digitalizar em um mesmo mapa, sem risco de trabalho duplicado ou confusão na construção dos arquivos de vetor. Uma planilha de controle, porém foi utilizada para verificar o andamento das tarefas e dar celeridade ao processo sem que houvesse interrupção para saber o que deveria ser digitado ao fim da criação de um arquivo. Desta forma, um colaborador pode, por exemplo, vetorizar os pontos do coletor “SJ” enquanto outro pode digitalizar os pontos do coletor “PM”, atingindo de maneira mais rápida o final do processo.

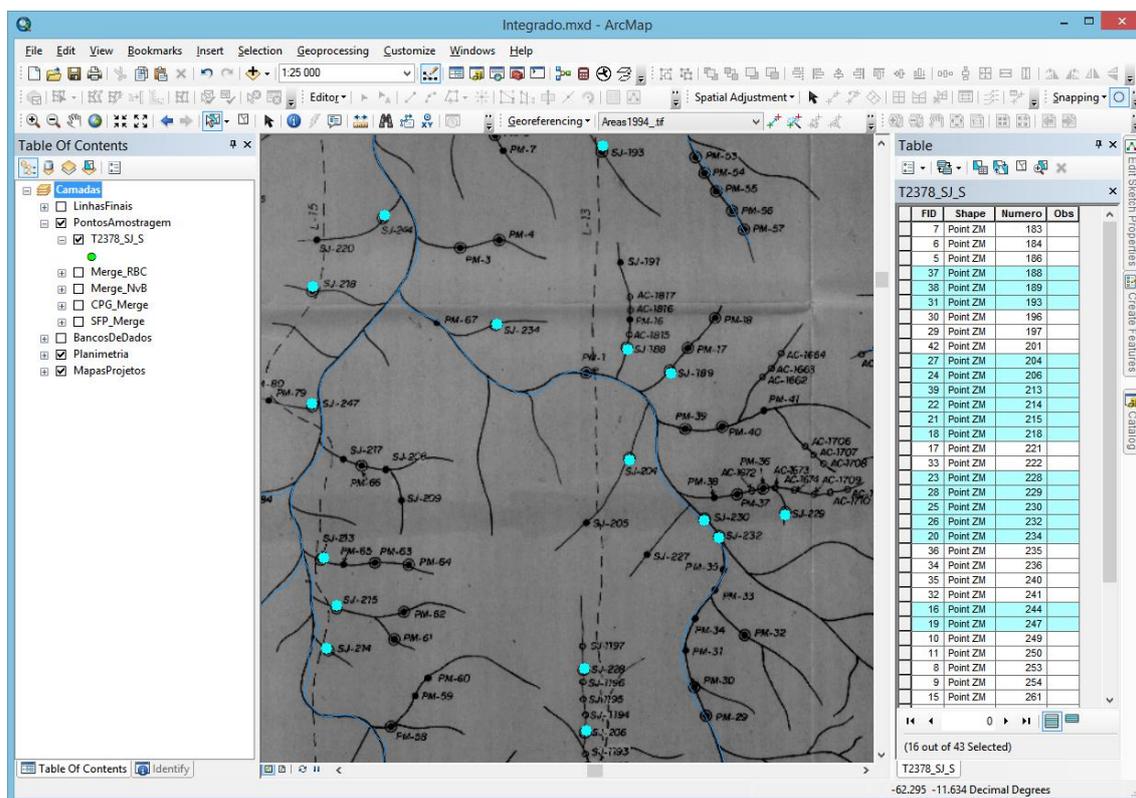


Figura 3.3-1: Vetorização de pontos. Nome do arquivo (shapefile) é identificado pelo centro de custo, sigla do coletor e classe da amostra (T2378_SJ_S.shp). Tabela de atributos somente com sequencial e observações. O exemplo mostra as amostras de sedimento de corrente do coletor SJ sendo vetorizadas.

Nos mapas de geoquímica, é muito comum encontrar símbolos mistos, indicando coleta de amostras de classes diferentes no mesmo ponto (e.g. concentrado de bateia e sedimento de corrente). Nestes casos, o nome do arquivo foi identificado como classe BS e depois o arquivo foi duplicado para contemplar as duas classes de amostra. Como consequência, os deslocamentos encontrados para estes pontos de observação foram também duplicados, pois o ponto é o mesmo, apesar de as amostras apresentarem características distintas. Esta é mais uma consequência indesejável causada pela confusão entre identificações do ponto de observação e da amostra.

3.4. Edição de Atributos e Junção de Arquivos

Finalizada a vetorização dos arquivos de pontos, foi feito o complemento dos atributos, colocando em colunas as informações relativas ao centro de custo, folha ou área mapeada, sigla do coletor e classe de amostra. Ou seja, as informações contidas no nome do arquivo

foram transferidas para a tabela de atributos dos vetores. Foi dada atenção à padronização dos nomes dos campos, para evitar conflitos e erros na junção dos arquivos. Para definição dos atributos padronizados, foi utilizado o *Field Calculator* do ArcGIS 10.x, garantindo um preenchimento uniforme dos campos da tabela de atributos (Figura 3.4-1). O processo de vetorizar todos os pontos, classificando primeiramente apenas os números sequenciais e observações e só depois adicionar as demais informações aos atributos, permite dar maior celeridade ao desenvolvimento do trabalho.

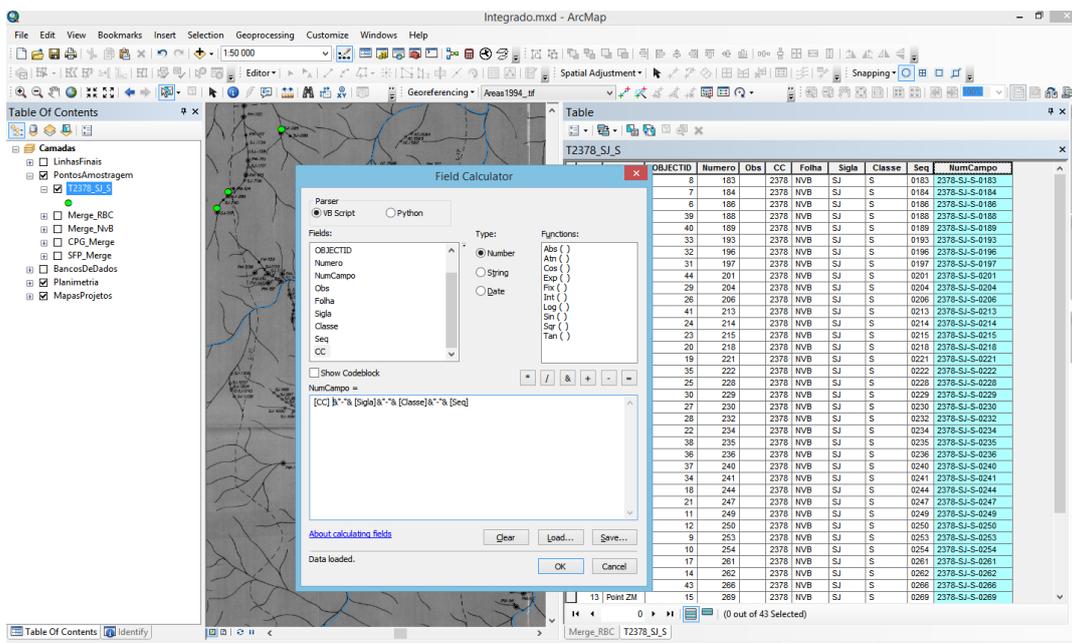


Figura 3.4-1: Transferência dos dados do nome do arquivo para a tabela de atributos. A identificação única (primary key ou chave primária) é definida pela concatenação do centro de custo, sigla do coletor, classe da amostra e sequencial.

A definição de uma chave primária para os pontos é fundamental para não haver duplicidade de pontos, então foi escolhida a codificação do número de campo, definida pelos manuais de geologia da CPRM. Apesar de atualmente ser discutida a sua funcionalidade, é o registro comum entre o Geobank e os mapas históricos, pois os mapas utilizados não apresentam número de laboratório. Assim a chave primária foi gerada por concatenação de campos das tabelas de atributos separados por hifens, conforme abaixo:

[Centro de custo]+[Sigla do coletor]+[Classe da rocha]+[Número sequencial].

Alguns cuidados foram tomados para que não houvesse repetição dos números de campo. O mais importante foi garantir que o número sequencial tivesse quatro dígitos. Isso é feito utilizando um campo do tipo texto (*Text* ou *String*) e adicionando zeros à esquerda, pois os campos numéricos não os preservam. Outro cuidado foi verificar se havia mais de um coletor com a mesma sigla ou mais de uma sigla para um mesmo coletor, ambos os casos não identificados na execução deste trabalho. Isso garantiu a utilização de um único código de centro de custo, CC2378, apesar de o Projeto Platina Rondônia ter recebido pelo menos três códigos (4270, 2381 e 2378) ao longo de sua execução. Com isso, fica garantida a unicidade dos números de amostras, que se confunde com o número de campo, no caso de levantamentos geoquímicos.

A junção dos vetores foi feita utilizando-se o comando “Merge” do ArcGIS 10.2, primeiramente para todos os arquivos de uma mesma folha/área (Figura 3.4-2). Tendo em mãos os arquivos de cada área do projeto já consolidados, foi feita a junção dos pontos de todo o projeto em um mesmo arquivo, utilizando-se o comando “Merge” ou “Append”.

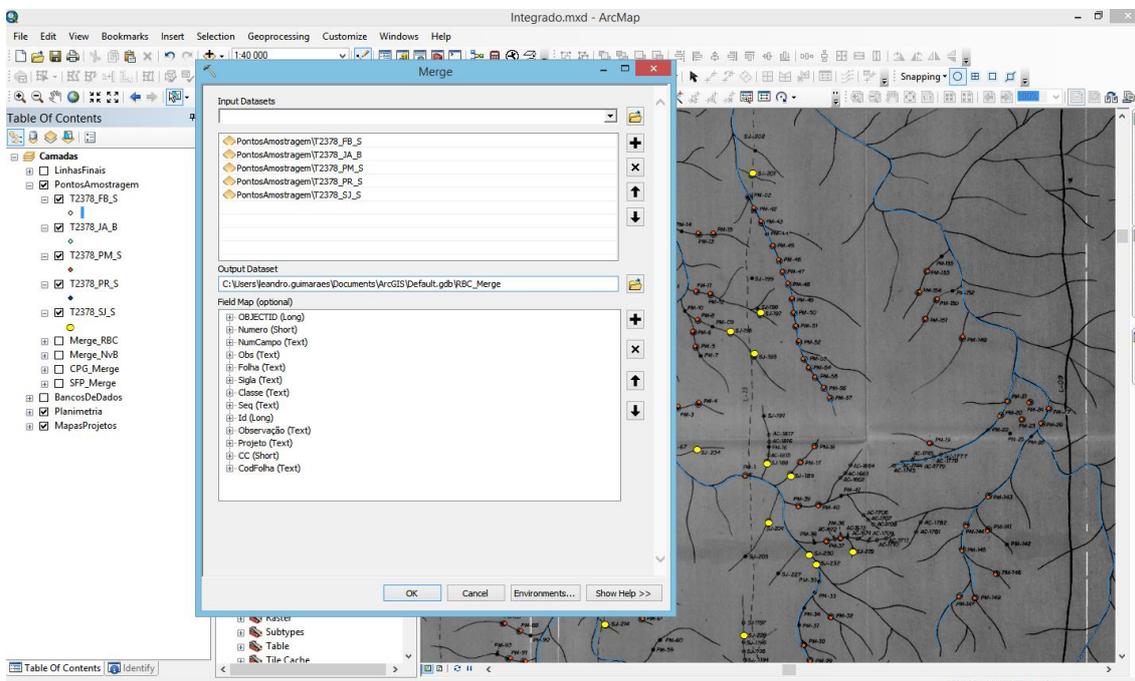


Figura 3.4-2: Junção de arquivos (merge), considerando padronização de informações na tabela de atributos.

3.5. Junção de dados vetorizados e Geobank

A etapa seguinte consistiu de trazer os dados do Geobank aos pontos vetorizados do projeto, atribuindo então, as informações analíticas aos pontos com posicionamento corrigido. Assim, todos os dados de geoquímica, química mineral, mineralometria e qualquer outro que esteja no Geobank puderam ser utilizados.

Depois de adicionada a tabela de interesse no projeto do ArcGIS, foi feita a padronização do nome de campo, conforme descrito no Item 3.4. O número de campo armazenado no Geobank apresenta apenas a sigla do coletor e sequencial, formatado em desacordo com as normas vigentes da CPRM. Assim, foi feita a padronização, adicionando o centro de custo e o código da classe, além de ser formatado de acordo com a norma. Por exemplo, a amostra de concentrado de bateia coletada no ponto **AC0150**, da tabela do Geobank, foi formatado para **2378-AC-B-0150**, possibilitando assim o estabelecimento de relacionamento com a tabela de atributos dos pontos vetorizados.

Utilizando-se o número de campo como coluna comum entre as tabelas de atributos dos vetores e dos dados de geoquímica e descrição de rochas do Geobank, foi possível ajustar as coordenadas para melhor adequação aos mapas originais de projetos (Figura 3.5-1).

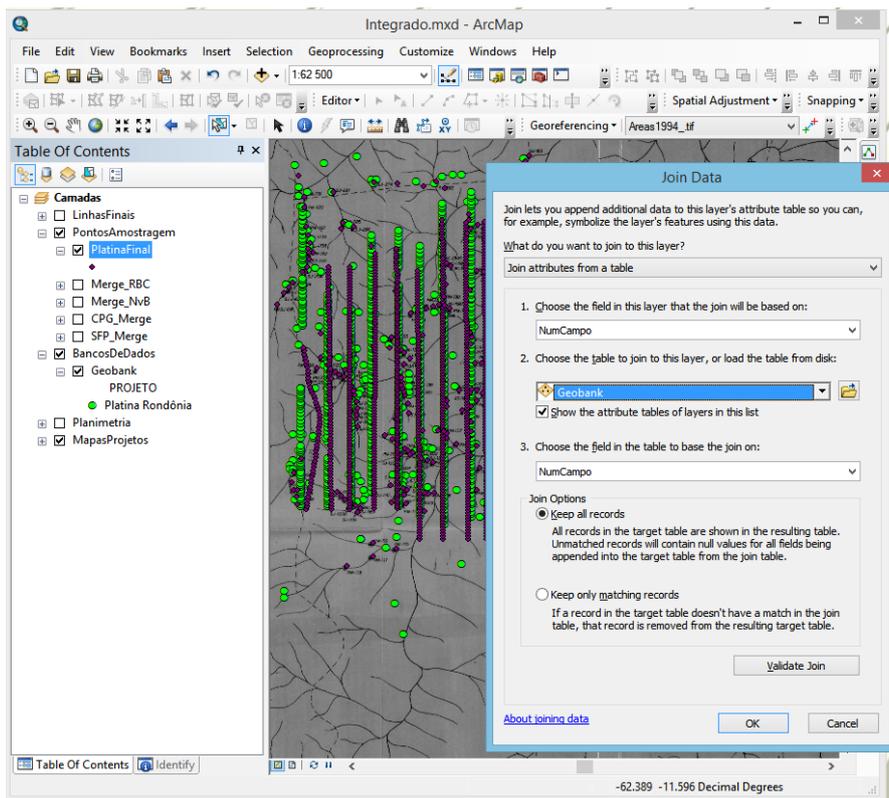


Figura 3.5-1: Junção de pontos vetorizados a partir dos mapas do projeto e dados do Geobank. Após a junção, as coordenadas obtidas pela vetorização são utilizadas para trazerem os dados do Geobank para a localização correta.

4. Análise dos Deslocamentos

Com a aplicação do método acima descrito, foi possível quantificar os desvios encontrados em dados do projeto Platina Rondônia armazenados no Geobank, em relação aos mapas originais do projeto, utilizando como referência o nome de campo das amostras. Tanto nas coordenadas leste (UTM-E) quanto nas coordenadas norte (UTM-N) foram observados desvios iguais ou superiores a 100m em aproximadamente 80% das amostras (**Erro! Fonte de referência não encontrada.**). Este fato demonstra que grande parte dos dados do Geobank para este projeto está com problema de posicionamento, colocando em questão a qualidade do banco de dados.

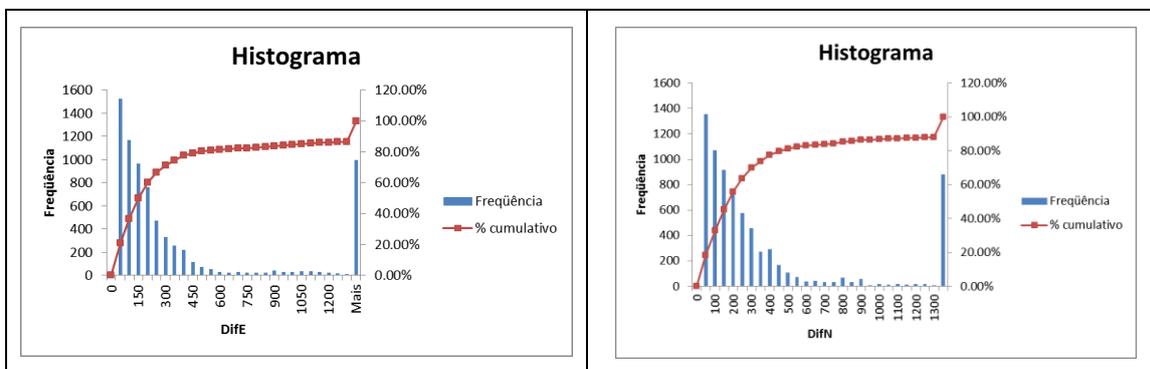


Figura 4-1: Histogramas da diferença absoluta em metros entre coordenadas UTM-E e UTM-N da metodologia apresentada e do Geobank.

Na Figura 4-1 observa-se que mais de 10% das amostras estão com mais de 1300m de erro no seu posicionamento, em cada um dos eixos de coordenadas. As causas desta diferença não são foco deste relatório, que tão somente se presta a demonstrar uma forma de solucionar o problema. Como os dados de posicionamento devem representar algum documento, foi admitido que os mapas dos projetos são os documentos mais confiáveis, se não forem os únicos, capazes de respaldar o posicionamento das amostras do banco de dados.

Tabela 1: Quantitativo e percentual de amostras com erros de posicionamento acima de limiares selecionados, para coordenadas leste e norte.

| | Limiar (m) | Amostras | % |
|---------------------|------------|----------|--------------|
| Desvio UTM-E | >100 | 5814 | 79.24 |
| | >500 | 1528 | 20.83 |
| | >1000 | 1177 | 16.04 |
| Desvio UTM-N | >100 | 5983 | 81.55 |
| | >500 | 1479 | 20.16 |
| | >1000 | 984 | 13.41 |
| Desvio total | >100 | 7044 | 96.01 |
| | >500 | 2161 | 29.45 |
| | >1000 | 1403 | 19.12 |

Total de Amostras 7337

Quando considerado o deslocamento total dos pontos (vetor deslocamento), verifica-se que 96% dos dados avaliados apresentam mais de 100m de diferença no posicionamento em relação aos mapas originais do projeto, 30% estão deslocados acima de 500m e 19% acima de 1000m. Estes fatos, apesar de representarem a realidade de parte amostras em relação ao conjunto de dados total do projeto, é um indicativo de que os dados carecem de maior confiabilidade e que é necessária a correção dos dados com base em critérios claros, objetivos e calcados em material documental.

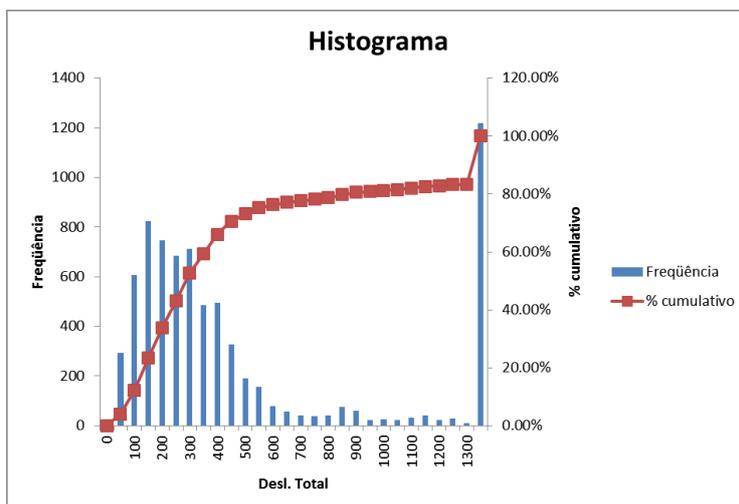


Figura 4-2: Histograma do comprimento do vetor deslocamento dos pontos avaliados.

Os azimutes dos vetores de deslocamentos foram calculados no ArcGIS 10.2 com a ferramenta Easy Calculate 10. Estes apresentaram distribuição aproximadamente aleatória, com tendência para o Norte, mais especificamente no primeiro quadrante, conforme indicado na Figura 4-3A. Um grupo de vetores ficou acumulado entre 295° e 300°, que estão espacialmente associados às malhas de solo da folha São Felipe-Santa Luzia, totalizando mais de mil pontos deslocados

em torno de 30km. Pela Figura 4-3B também se depreende que este conjunto representa os maiores vetores, ou seja, os pares de pontos que têm maiores discrepâncias de deslocamento.

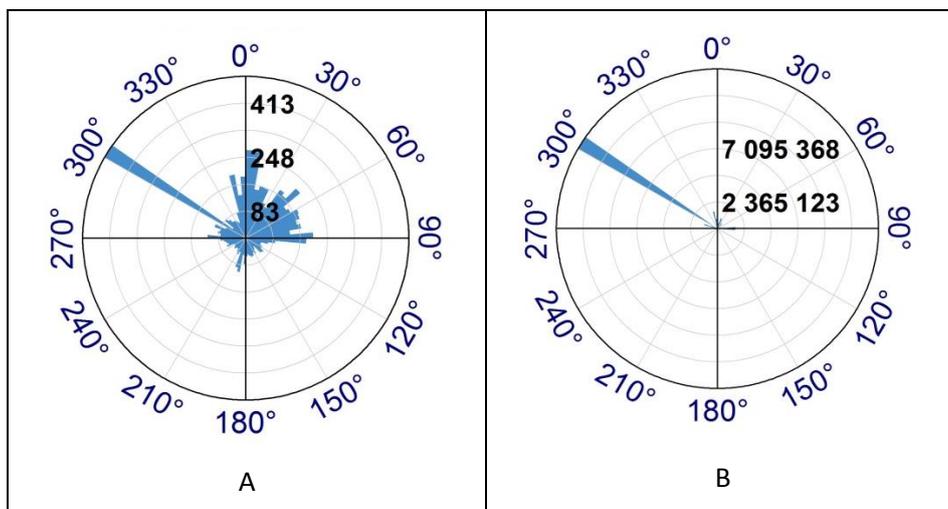


Figura 4-3: Diagrama de roseta e histograma indicando: (A) azimute dos vetores de deslocamento; (B) azimute normalizado pelo tamanho dos vetores de deslocamento.

Os vetores de deslocamento foram criados tomando-se como ponto inicial o par de coordenadas obtido na vetorização dos pontos e, como ponto final, o par de coordenadas contido no Geobank, para cada número de campo comum Primeiramente, foram atribuídas as coordenadas de cada ponto na tabela de atributos, utilizando-se a ferramenta “Add XY Coordinates”, do ArcToolbox (Figura 4-4).

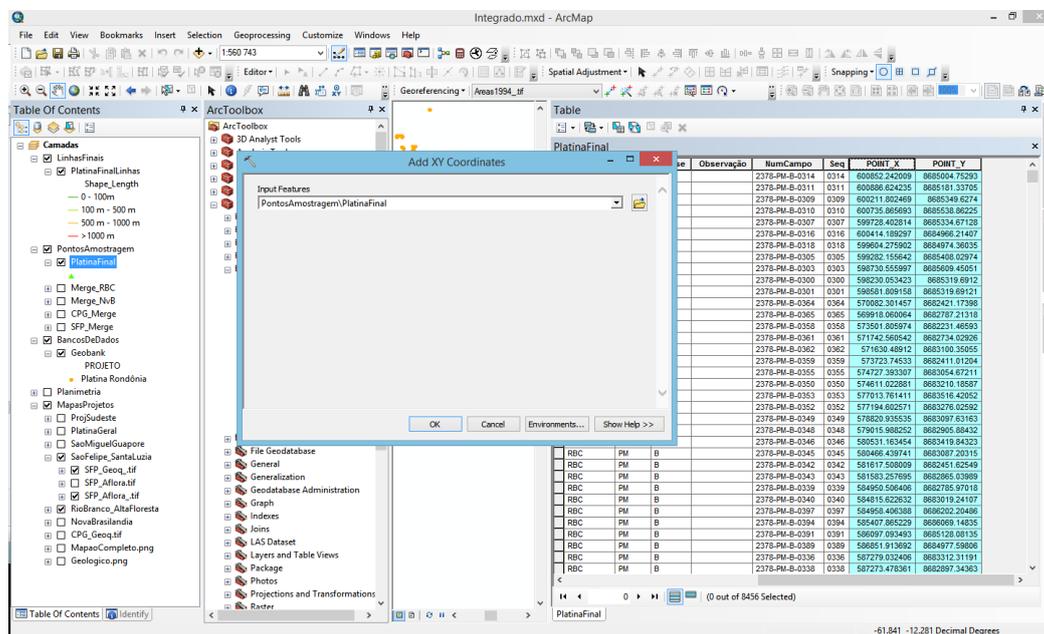


Figura 4-4: Adição de coordenadas aos pontos obtidos por vetorização dos mapas históricos.

Após a junção com a tabela do Geobank, foi possível criar os vetores de deslocamento, utilizando-se a ferramenta “XY To Line”, atribuindo-se o ponto inicial como sendo os pontos vetorizados e o ponto final os pontos do Geobank, para um mesmo número de campo (Figura 4-5).

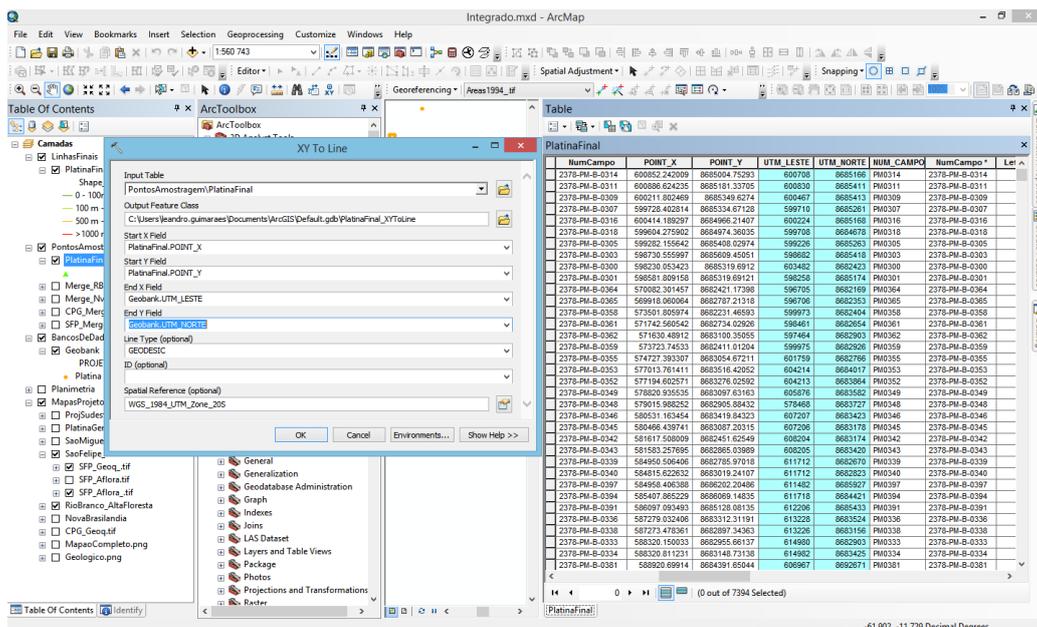


Figura 4-5: Definição dos pontos iniciais e finais de cada vetor de deslocamento.

Em mapa, apesar da visualização um tanto caótica, é possível perceber os problemas decorrentes da localização errônea dos dados, indicados conforme a legenda de cores (Figura 4-6). Não foi verificada a correlação dos outros dados do Geobank em relação a documentos originais dos projetos, tarefa muito mais complexa e que demandará muito mais recursos.

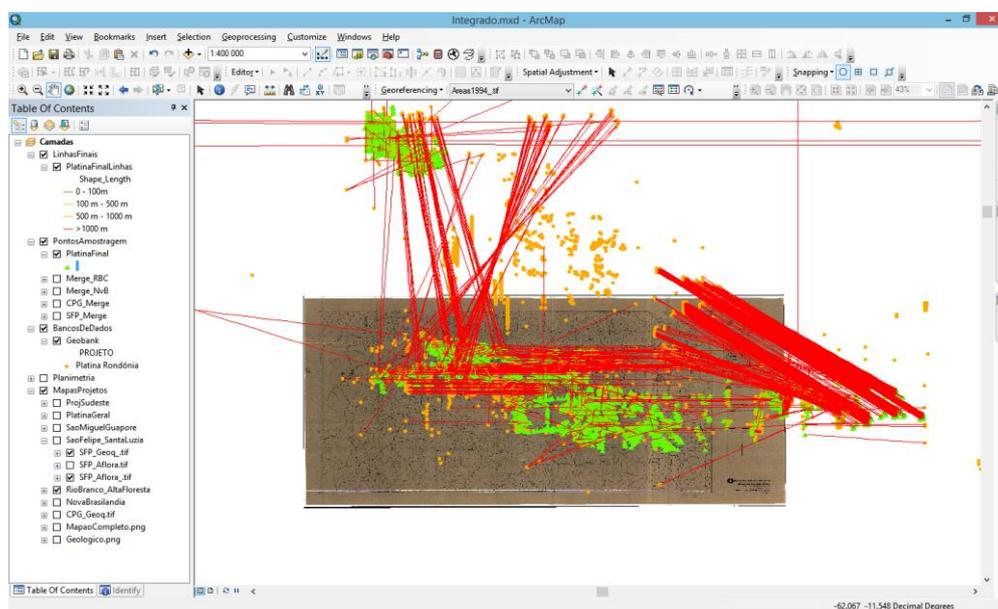


Figura 4-6: Vetores de deslocamento visualizados em mapa. Linhas vermelhas indicam deslocamentos superiores a 1000m.

A partir da correção dos dados do Geobank, se torna possível a aplicação de técnicas de processamento de dados, tais como análises de grupo, geoestatística, dentre outras, para melhor utilização dos produtos do banco de dados. Além disso, espera-se que a coerência espacial dos dados de geoquímica seja incrementada, quando relacionadas aos domínios geológicos definidos pelo mapeamento sistemático, o que claramente não seria possível com os dados originais atualmente armazenados no Geobank.

5. Conclusões e Recomendações

Dada a importância da utilização de dados históricos para o planejamento de futuros projetos da CPRM, se fazem necessárias a verificação e a validação sistemáticas dos dados armazenados no Geobank. Recomenda-se a criação de grupos de trabalho em cada Unidade Regional para a garantia de continuidade do processo de retificação de coordenadas, uma vez que o volume de dados é muito grande e as referências documentais dos dados armazenados se encontram nas UR's de origem do projeto. Neste caso ficaria a cargo de ser estruturado o procedimento de reenvio dos dados corrigidos à DIGEOP, responsável pela manutenção do Geobank.

Recomenda-se o agrupamento e verificação de dados utilizando-se o código de Centro de Custo como referência, pois estes podem variar dentro de um mesmo projeto. Assim, fica garantida a unicidade das amostras coletadas e a referência aos documentos originais se preserva. Não é recomendada a utilização do nome do projeto como referência para os dados, a menos que os dados de centro de custo não estejam disponíveis.

A identificação dos coletores também se faz necessária, pois esta faz parte da nomenclatura dos pontos de observação e das amostras coletadas dentro de um projeto, juntamente com o centro de custo. É necessário também garantir que cada sigla se refira a apenas a um indivíduo, para garantir unicidade das amostras.

O georreferenciamento deve ser feito observando sempre a data de execução e *datum* originalmente utilizado nos mapas. No caso de mapas confeccionados a partir de fotos aéreas e sem ajuste em base cartográfica, é necessária a retificação com polinômio adequado, para contemplar as distorções inerentes ao método.

Os erros encontrados nos mapas dos projetos devem ser registrados e caso haja alguma correção, esta deve ser informada no campo "Observações" da tabela de vetorização dos pontos.

A melhor identificação para cruzamento dos arquivos vetoriais gerados para os pontos e os dados do Geobank é o número de campo, desde que devidamente formatado segundo procedimento vigente na CPRM. No entanto, o uso do centro de custo e sub-centro de custo deve ser feito com cautela, pois os mesmos tiveram seus significados alterados ao longo da história, além de se repetirem para projetos ou UR distintas. É recomendada a criação de novo código único, em nível nacional, para os projetos que forem recuperados nas UR.

Muitas das amostras coletadas não possuem registros no Geobank. Assim, o valor deste tipo de informação fica restrito ao conhecimento de que aquele ponto foi visitado. Por outro lado, informações adicionais podem ser inseridas no banco, caso exista alguma documentação, como a caderneta de campo do coletor.

Leandro Guimarães da Silva
Geólogo – CPRM/REPO
Matr.: 318.353-41