

INVESTIGAÇÃO DA EXATIDÃO DE COORDENADAS OBTIDAS COM O PPP/GNSS USANDO ÁRVORE DE DECISÃO

Mauro Menzori. Professor na Faculdade de Tecnologia da Unicamp.
E-mail: mauro@ft.unicamp.br

Resumo: O objetivo do trabalho é analisar o quanto os resultados de posicionamento de pontos obtidos com o método PPP-GNSS se aproximam das coordenadas da posição real desses pontos no terreno. O posicionamento de pontos na superfície da Terra tornou-se mais simples com o surgimento dos sistemas de posicionamento por satélites artificiais. Embora atualmente apenas as constelações do sistema norte-americano GPS e do sistema russo GLONASS sejam as únicas a transmitirem dados utilizáveis, existem outras constelações que se encontram em estágio inicial e deverão a médio prazo fazer parte do GNSS. Dentre os diferentes métodos de posicionamento preciso que usam as observáveis GNSS está o método de Posicionamento por Ponto Preciso (PPP). Assim como em qualquer outro método de posicionamento, ele usa os dados das observáveis para produzir as coordenadas que resultam sempre acompanhadas de indicadores estatísticos sobre a sua precisão. Como se sabe, a precisão de um resultado é algo diferente da sua exatidão. Enquanto a precisão informa a qualidade dos dados, a exatidão nos diz o quanto as coordenadas calculadas estão próximas da posição real no terreno. Embora o vínculo entre a precisão e a exatidão esteja presente nos dados, o processamento não consegue alcançar essa correlação. Neste trabalho usou-se a ferramenta de mineração de dados conhecida como Árvore de Decisão para prever a exatidão de coordenadas das quais se conhece apenas a precisão. Para isso, constituiu-se um banco de dados povoado com coordenadas cuja precisão e exatidão eram conhecidas. Esse banco serviu como referência para o treinamento recursivo de uma Árvore de Decisão que a tornou capaz de prever a exatidão de coordenadas das quais se conhece apenas a precisão.

Palavras-chave: GNSS, PPP, Árvore de Decisão, Precisão, Exatidão.

PPP/GNSS OBTAINED COORDINATES ACCURACY ANALYSIS USING DECISION TREE TECHNIQUE

Abstract: The aim of this paper is to perform the analysis over the resulting coordinates obtained with the PPP-GNSS method to classify their accuracy. The point positioning over the Earth's surface has become simpler after the building of positioning systems that uses artificial satellites, especially nowadays, because of the existence of several systems that form the GNSS. Although actually the satellites on the GNSS constellations belong to the North American positioning system GPS or to the Russian positioning system GLONASS, the most structured systems, there are other systems integrating the GNSS and they will be fully functional in a few years. Among different methods that can be used to perform precise positioning using transmitted data by GNSS satellites we can find the PPP method to obtain geodetic precise coordinates. Similarly to other methods, PPP uses the observables to produce coordinates and their precision. As we know precision is something different from accuracy. While precision informs the data set quality, accuracy tells us how much the calculated coordinate is close to its position on the ground. Any positioning method only informs the precision but the accuracy. Although the correlation between precision and accuracy has been registered implicit in the observables, the processing methods can't achieve it. This study used the data mining tool known as Decision Tree to identify the rules that connect precision and accuracy, to predict the accuracy of coordinates from its known precision. A large set of coordinates with known precision and accuracy was used as reference for the recursive training of a Decision Tree that became able to predict the accuracy of coordinates just having its precision.

Keywords: GNSS, PPP, Decision Tree, Precision, Accuracy.

1. INTRODUÇÃO

A evolução dos sistemas globais de navegação por satélites orbitais que integram o Global Navigation Satellite Systems, ou simplesmente GNSS, tem acontecido de forma regular e constante ao longo das últimas décadas e nos leva a entender que em pouco tempo o mundo alcançará um novo patamar no posicionamento de pontos usando essa tecnologia. Dentre esses sistemas, o Sistema de Posicionamento Global (GPS) encontra-se em estágio mais

avançado, em plena etapa de modernização, com a previsão de lançamento de satélites do bloco III a partir de 2017. O sistema russo Global Navigation Satellite System (GLONASS) encontra-se nos instantes finais para a obtenção da certificação de plena operabilidade ou Full Operational Capability (FOC), enquanto que o sistema europeu GALILEO e o sistema chinês Compass Navigation Satellite Experimental System (BEIDOU) encontram-se em estágios intermediários de implantação.

Nesse contexto de novidades com a consequente ampliação de horizontes, alguns pontos ainda merecem ser pesquisados, uma vez que eles pertencem mais à técnica fundamental que é aplicada no posicionamento do que a um sistema em particular. A relação entre a exatidão e a precisão de um posicionamento é o tema abordado no presente trabalho, e foi investigada a partir de dados observados com receptores GNSS de dupla frequência.

O objetivo deste trabalho foi estudar a exatidão de coordenadas obtidas pelo método de Posicionamento por Ponto Preciso (PPP) e a viabilidade de usá-las em trabalhos de engenharia que requerem boa exatidão. Para entender o comportamento da exatidão desenvolveu-se a análise dos resultados PPP obtidos durante um período de seis meses, levando em conta as diferentes fontes de erro que agem sobre o sinal propagado e que causam desvios acima da tolerância aceitável para fins de engenharia.

No trabalho aplicou-se a técnica computacional de Aprendizado de Máquina que usa um banco de dados povoado com as precisões e as exatidões conhecidas de um conjunto de pontos previamente medidos para, através de um treinamento computacional, induzir (criar) uma Árvore de Decisão (AD), com capacidade para estimar a real exatidão de um novo posicionamento do qual se conhece apenas a precisão.

Diferentes são os métodos de observação que podem ser desenvolvidos empregando receptores de sinais transmitidos pelos satélites que compõem as constelações que integram o GNSS. Esses métodos produzem as coordenadas geodésicas de pontos, com diferentes precisões, praticamente em toda a superfície física da Terra. Dentre eles, o método absoluto conhecido como Posicionamento por Ponto Preciso (PPP) é um método que permite fazer o posicionamento de precisão empregando apenas um receptor, com o qual se adquire os dados das fases das ondas portadoras para depois processá-los de forma combinada com as efemérides precisas disponibilizadas pelo International GNSS Service (IGS). Trata-se de um método muito útil na determinação de coordenadas de pontos que se encontram afastados de uma rede de referência.

Por ser um método absoluto o PPP não se conecta às redes geodésicas existentes na região do trabalho e, portanto, as coordenadas determinadas com o seu uso não possuem os resíduos de ajustamento de uma rede geodésica existente. Pode-se dizer que usando o PPP, cada ponto é um ponto independente que possui a sua própria precisão. No entanto, os trabalhos que venham a usar as coordenadas desse ponto certamente farão a sua conexão com redes geodésicas existentes, o que pode ser um problema se a sua exatidão não for adequada. O Instituto Brasileiro de Geografia e Estatística (IBGE) disponibiliza online o serviço PPP (IBGE-PPP) que processa os dados GNSS e fornece as coordenadas do ponto que foi medido, vinculadas ao Sistema de Referência Geocêntrico para as Américas (SIRGAS2000) e ao International Terrestrial Reference Frame (ITRF). Segundo HOFMANN et al. (2008), a técnica empregada na determinação das coordenadas pelo método PPP usa um ajustamento matemático pelo critério dos mínimos quadrados (M.M.Q.) que fornece indicadores estatísticos sobre a precisão da solução encontrada no ajustamento. Como se sabe, precisão não é exatidão e por esse motivo existe algum risco em assumir como boas as coordenadas que resultam do processamento PPP tendo por base apenas a sua precisão. Em muitas situações as coordenadas determinadas com altíssima precisão não possuem boa exatidão e,

por esse motivo, não representam a verdadeira posição do ponto na superfície física da Terra. Isso acontece porque os dados iniciais adquiridos pelos receptores contêm perturbações de alguma espécie, como por exemplo, o multicaminhamento, que segundo MONICO (2000), é uma interferência local capaz de degradar as observáveis das fases e dos códigos e produzir as coordenadas do ponto certamente afastadas da sua real posição no terreno.

Dessa forma, o trabalho aqui apresentado foi desenvolvido no sentido de encontrar uma maneira para se estimar indiretamente o afastamento entre a precisão e a exatidão de uma solução de posicionamento PPP-GNSS. A hipótese principal neste trabalho é que uma vez conhecida a correlação entre a precisão e a exatidão de um conjunto significativo de dados GNSS, torna-se possível fazer a previsão da exatidão de uma nova medição empregando a técnica computacional de Aprendizado de Máquina conhecida como Árvore de Decisão (AD).

1.1 Método De Posicionamento Por Ponto Preciso (PPP)

O PPP é um método onde as coordenadas de posição do receptor são calculadas diretamente em função das coordenadas de posição dos satélites. Trata-se de um método de posicionamento absoluto, e por esse motivo o PPP também é conhecido como um método de Posicionamento Absoluto Preciso. As coordenadas georreferenciadas obtidas com esse método não estão associadas a qualquer rede planimétrica, ou a qualquer rede altimétrica existente, e, exatamente por esse motivo, segundo o IBGE (2013), podem apresentar diferenças significativas em relação aos pontos dessas redes. Em outras palavras, coordenadas determinadas com o PPP podem apresentar exatidões inaceitáveis.

O PPP é um método similar ao método de posicionamento absoluto simples, mas não é igual. Uma diferença notável é que o PPP usa as efemérides precisas dos satélites e não as efemérides transmitidas por eles, como acontece no posicionamento absoluto simples. De forma decisiva entram no cálculo PPP o movimento das placas tectônicas, as marés terrestres, os erros dos relógios do satélite e do receptor, os deslocamentos (offsets) do centro de fase da antena do satélite e da antena do receptor para assegurar que as coordenadas resultem com boa precisão. Outra diferença é que esse método foi desenvolvido para ser executado com um receptor de dupla frequência porque utiliza além dos dados do código C/A, também os dados das fases das portadoras L1 e L2. Com esse tipo de receptor se adquire os dados necessários para modelar a ionosfera e desenvolver o modelo conhecido como ionosphere-free, ou ionofree, que segundo XU (2003), elimina os efeitos da ionosfera através de combinações das equações dos códigos e das fases das portadoras. Trata-se de uma combinação linear dos dados, extremamente útil para eliminar os erros que os sinais absorvem quando atravessam a atmosfera terrestre.

Mais recentemente o método PPP passou a ser executado a partir de dados coletados com receptores de simples frequência que não têm como desenvolver o modelo ionofree. Isso se tornou possível porque os parâmetros de modelagem da ionosfera passaram a ser transmitidos pelos satélites junto com as observáveis.

A precisão do método PPP foi estimada por SILVA e SEGANTINE (2015), na ordem de 5 a 10 cm, embora alguns ensaios demonstrem que ela possa alcançar de 2 a 5 cm, principalmente quando o tempo de coleta é superior a duas horas e passa a existir uma convergência dos resultados. Ainda assim, trata-se de precisão e não de exatidão.

O método PPP passou a ser oferecido gratuitamente no Brasil pelo Instituto Brasileiro de Geografia e Estatística (IBGE) por volta do ano 2000, através do link <http://www.ppp.ibge.gov.br/ppp.htm> e usa o programa CSRS-PPP (GPS Precise Point Positioning) desenvolvido pelo Geodetic Survey Division of Natural Resources of Canada (NRCan). Todos os dados do presente trabalho foram processados com esse recurso.

1.2 Árvore de Decisão

A Aprendizagem de Máquina (AM) é uma característica de um sistema computacional que consegue, a partir de grandes quantidades de dados, aprender automaticamente sobre certo evento, permitindo que da próxima vez uma tarefa retirada de um mesmo grupo de tarefas possa ser feita com maior eficácia, WITTEN e FRANK, (2005).

Trata-se de um processo que identifica de forma automática os padrões implícitos em grandes quantidades de dados e devido a essa capacidade as técnicas de Aprendizado de Máquina (AM) são cada vez mais empregadas para tratar de problemas de grande complexidade e de difícil conceituação. De acordo com ZHAN-LI et al. (2015), o processo de AM é capaz de identificar e recuperar sinteticamente pontos tridimensionais que tenham se perdido durante a captura de uma sequencia de imagens de vídeo, num processo conceitualmente muito próximo ao da classificação de pontos determinados pelo PPP. Dentre as atuais técnicas de Aprendizado de Máquina encontram-se:

- 1) A Rede Neural, ou rede Multi Layer Perceptron (MLP), indicada para a classificação múltipla de eventos, onde a quantidade de exemplos de aprendizado é tipicamente de grande monta.
- 2) O algoritmo de Support Vector Machines (SVM), extremamente rápido, mas, com a desvantagem de resolver somente problemas binários, envolvendo apenas duas classes.
- 3) A Árvore de Decisão (AD) que trabalha com um número ilimitado de dados multivariados que servem de exemplos de teste na etapa de treinamento. Ela também possui a capacidade de interpretar e entender as regras implícitas nesse conjunto de dados, e, posteriormente consegue usar essas regras num processo de predição capaz de criar infinitas classes que servirão para classificar um evento novo através da similaridade das suas características.

A exatidão das coordenadas obtidas com o método de PPP pode ser entendida como um problema complexo pelo fato do PPP ser desvinculado de redes geodésicas existentes e por esse motivo, a Árvore de Decisão (AD) é uma ferramenta adequada para explicar de forma clara a exatidão de posicionamento implícita nos dados observados.

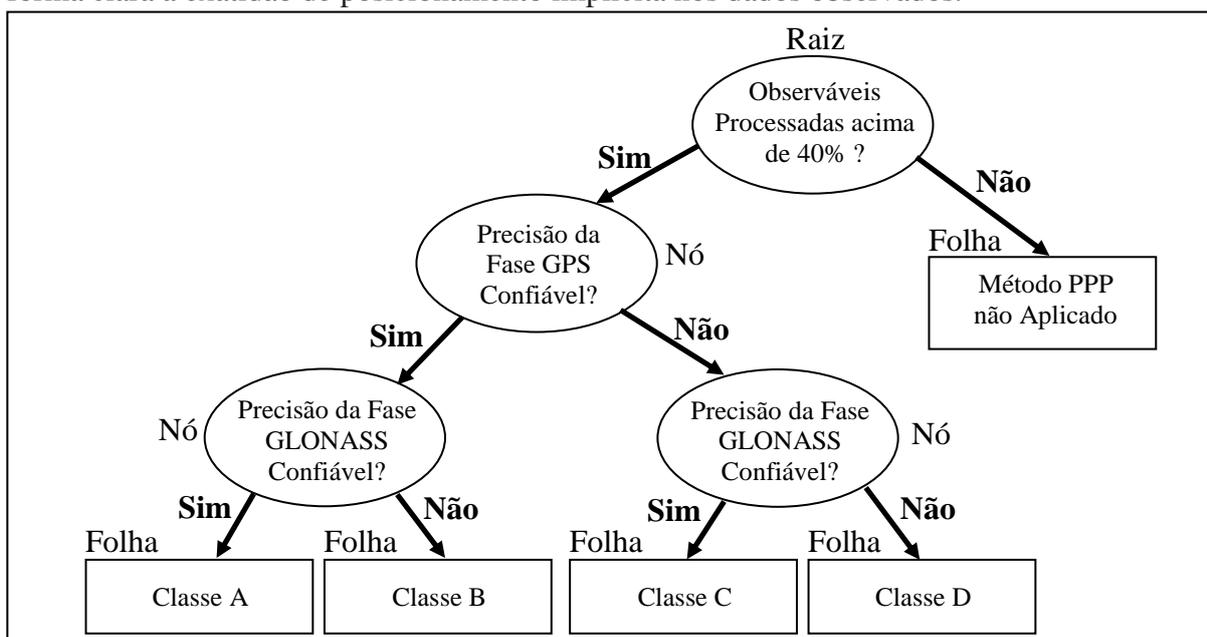


Figura 1: Estrutura da Árvore de Decisão.

Toda medição feita através do GNSS é constituída de diferentes variáveis que podem ser modeladas, tais como: as observáveis, a taxa de gravação, o tempo de coleta, a perturbação ionosférica e a perturbação troposférica, e portanto, elas são passíveis de serem analisadas por uma Árvore de Decisão (AD). Segundo LEVINE et al. (1988), uma AD é induzida (criada) a partir de um banco de dados confiável, sendo uma estrutura de dados constituída recursivamente por: nós de decisão e folhas. Os nós de decisão correspondem a um teste sobre alguma variável e as folhas correspondem às classes resultantes, conforme exemplifica a Figura 1.

Para classificar uma medição obtida a partir de observáveis GNSS, o processo começa pela raiz, seguindo cada teste até que a folha seja alcançada, momento em que acontece a classificação. Em termos computacionais, o conhecimento contido nesta árvore é sintetizado a seguir:

```
if observáveis processadas acima de 40% then
/  if precisão da fase GPS confiável then
/  /  if precisão da fase GLONASS confiável then
/  /  classe = Classe A
/  /  else
/  /  classe = Classe B
/  /  end
/  else
/  /  if precisão da fase GLONASS confiável then
/  /  classe = Classe C
/  /  else
/  /  classe = Classe D
/  end
else
/  classe = Método PPP não Aplicado
end
```

Assim, cada árvore pode ser representada por um conjunto de regras, onde cada regra tem o seu início na raiz da árvore e caminha até uma de suas folhas. Como qualquer outro procedimento automatizado e repetitivo a Árvore de Decisão apresenta vantagens e desvantagens, sendo que dentre as vantagens podem ser destacadas as seguintes:

- 1) A Árvore de Decisão é facilmente induzida (criada) e inteligível.
- 2) Não exige a definição “a priori” para qualquer parâmetro dos dados em análise.
- 3) O número de exemplos utilizados, a qualidade do banco de dados e a intensidade do treinamento controlam os algoritmos geradores da Árvore de Decisão, que são considerados instáveis e sensíveis variações nos dados de treinamento. Isso minimiza o surgimento de resultados fracos nos pontos de decisão da árvore (nós) e impede que um erro de inferência se propague para todas as ramificações subsequentes.
- 4) A Árvore permite a classificação simultânea de dados alfa, de dados numéricos e de dados alfanuméricos, com a condição de que o atributo de saída seja sempre uma classe alfa.

Após ser submetida a um treinamento recursivo, uma Árvore de Decisão (AD) produz como resultado a estratificação dos dados na forma de classes. Segundo RICH & KNIGHT (1991), a classificação é um componente importante para a solução de muitos problemas, sendo em sua forma mais simples considerada como uma tarefa direta de reconhecimento. Do ponto de vista do Aprendizado de Máquina o ato de classificar é o processo de atribuir a um determinado dado que foi recebido o nome e a classe a qual ele pertence. Para que a Árvore

de Decisão (AD) induzida no presente trabalho pudesse classificar a exatidão das soluções de novos posicionamentos, ela teve que ser treinada com um conjunto de dados de exatidão conhecida, a ponto de estabelecer as regras de inferência intrínsecas nesses dados e tornar-se capaz de classificar.

2. MATERIAIS E MÉTODOS

A indução de uma Árvore de Decisão (AD) depende de um banco de dados confiável para treiná-la. Somente aprendendo com um conjunto de dados conhecidos a AD consegue entender as relações intrínsecas existentes nos dados e preparar-se para inferir sobre novos eventos. A etapa de constituição de um banco de dados confiável é, portanto, uma parte delicada e fundamental na geração da AD.

2.1 Banco de Dados Referencial

O banco de dados referencial criado neste trabalho é composto por um conjunto de dados multivariados que foram usados para criar a Árvore de Decisão (AD) e torná-la apta para fazer previsões sobre a exatidão de resultados. Os dados do banco referencial foram adquiridos de três estações geodésicas da Rede Brasileira de Monitoramento Contínuo dos Sistemas GNSS (RBMC), gerenciada pelo IBGE, localizadas no estado de São Paulo, conforme Figura 2.

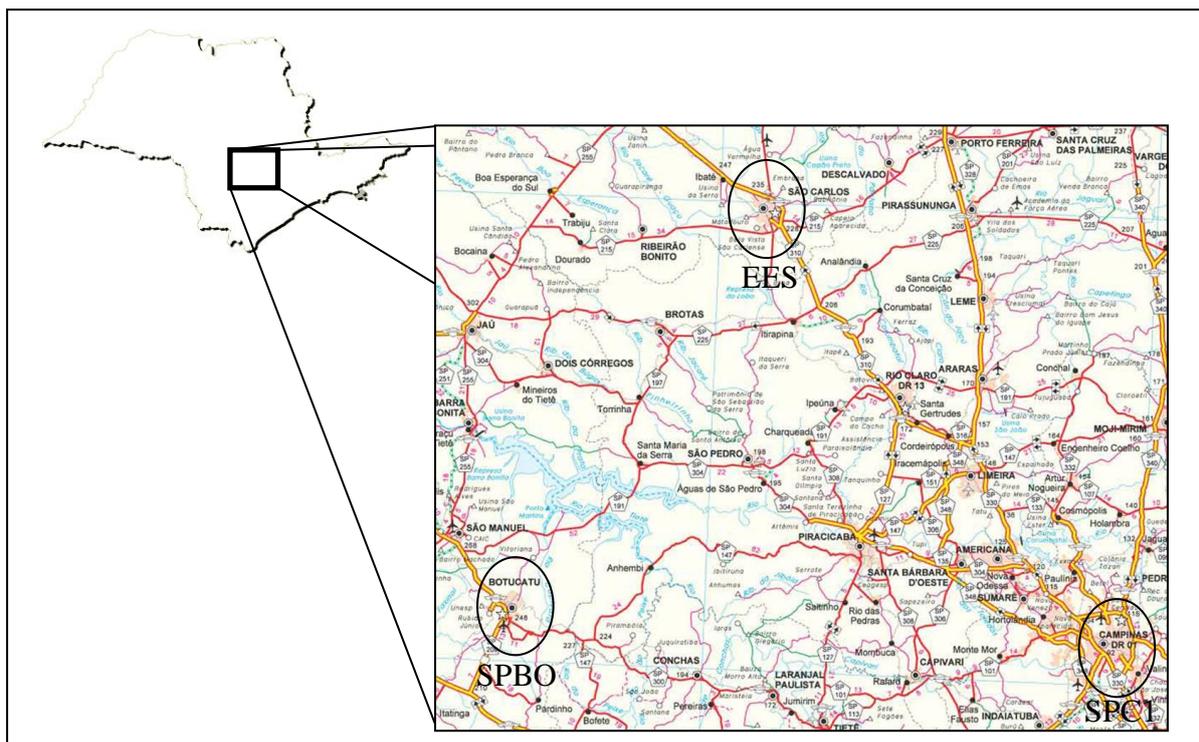


Figura 2: Estações geodésicas da RBMC utilizadas no trabalho.

Mais precisamente essas estações geodésicas da RBMC possuem as seguintes características:

Estação EESC, código 99560, com coordenadas oficiais publicadas pelo IBGE como sendo: latitude (ϕ) = 22° 00' 17,8160" S, longitude (λ) = 47° 53' 57,0497" W, e altura geométrica (h) = 824,587 m, fixa em uma torre metálica no teto da Escola da Engenharia de São Carlos, onde funciona um receptor GNSS Leica GR10, de dupla frequência, na cidade de São Carlos - SP.

Estação SPBO, código 99537, com coordenadas oficiais publicadas pelo IBGE como sendo: latitude (ϕ) = 22° 51' 08,8825" S, longitude (λ) = 48° 25' 56,282" W, e altura geométrica (h) = 803,122 m, fixa em um pilar cilíndrico na laje ao lado do Laboratório Didático de Topografia e Sensoriamento Remoto, do Departamento de Engenharia Rural da Faculdade de Ciências Agrônômicas da UNESP, onde funciona um receptor GNSS Leica GRX 1200 + GNSS, de dupla frequência, na cidade de Botucatu – SP.

Estação SPC1, código 96181, com coordenadas oficiais publicadas pelo IBGE como sendo: latitude (ϕ) = 22° 48' 58,6305" S, longitude (λ) = 47° 03' 45,6958" W, e altura geométrica (h) = 622,980 m, fixa em um cilindro de concreto no topo do edifício do Departamento de Geotecnia e Transporte, Faculdade de Engenharia Civil da Unicamp, onde funciona um receptor GNSS Trimble NETR9, de dupla frequência, na cidade de Campinas - SP.

Nessas estações foram adquiridos os dados GNSS relativos ao período de janeiro a maio de 2016, com intervalos espaçados de quinze em quinze dias, que depois de serem pós-processados pelo método PPP serviram para compor o banco de dados de referência. Como os dados observados em todas as estações geodésicas da RBMC são armazenados em arquivos de 24 horas contínuas, houve a necessidade de extrair, em cada dia da pesquisa, vários arquivos com 3 horas de dados, com avanço de uma hora cada um em relação ao arquivo anterior, ou seja, o primeiro arquivo do dia contém dados das 4:00 horas às 7:00 horas, o segundo arquivo das 5:00 h às 8:00 h e o último arquivo do dia contém dados das 15:00 horas às 18 horas. Foram preparados, portanto, 12 arquivos a cada dia cobrindo o período diurno, das 4:00 h às 18:00 h, período esse considerado como um horário comercial, quando a maior parte das empresas de georreferenciamento adquirem os seus dados que serão utilizados em algum serviço de engenharia. Além disso, os dados foram organizados em arquivos com 3 horas de observáveis GNSS, porque segundo o IBGE (2013) o resultado de um posicionamento pelo PPP passa a convergir após duas horas, e, no trabalho interessava analisar uma hora de dados com o mesmo padrão de convergência.

Essa forma de organização permitiu preparar 132 sessões de observação em cada estação geodésica, totalizando 396 sessões em estudo, que foram usadas na indução e treinamento da Árvore de Decisão. Conforme as instruções do manual PPP do IBGE, cada arquivo de três horas foi submetido online ao site <http://www.ibge.gov.br/home/geociencias/geodesia/ppp/default.shtm>, onde os dados das fases das portadoras L1 e L2 transmitidas pelos satélites das constelações GPS e GLONASS foram processados, com máscara de elevação superior a 10° através do método de Posicionamento por Ponto Preciso (PPP). Cada arquivo processado produziu várias informações relevantes à interpretação e análise dos resultados do posicionamento, dentre as quais se encontram as coordenadas do ponto e as dezessete variáveis apresentadas a seguir:

1. A precisão da solução PPP na latitude,
2. A precisão da solução PPP na longitude,
3. A precisão da solução PPP na altura geométrica,
4. O número de épocas GPS processadas,
5. O número de épocas GPS rejeitadas,
6. Os resíduos das Pseudodistâncias GPS,
7. Os resíduos das fases das portadoras GPS,

8. O número de épocas GLONASS processadas,
9. O número de épocas GLONASS rejeitadas
10. Os resíduos das Pseudodistâncias GLONASS
11. Os resíduos das fases das portadoras GLONASS
12. O percentual de épocas GPS rejeitadas,
13. O percentual de épocas GLONASS rejeitadas,
14. A exatidão da solução segundo a latitude,
15. A exatidão da solução segundo a longitude,
16. A exatidão da solução segundo a altura geométrica,
17. Classes de Exatidão.

A rigor, o processamento PPP-IBGE fornece as onze primeiras variáveis e as seis variáveis finais foram obtidas através de cruzamento dos dados, no sentido de realçar os pontos importantes no treinamento da Árvore de Decisão, que foram o percentual das épocas GNSS rejeitadas, tanto GPS quanto GLONASS, cuja proporção tem um direto relacionamento com a precisão do resultado de posicionamento.

2.2 Programa Empregado na Indução da Árvore de Decisão

Para interpretar as dezessete variáveis produzidas no processamento PPP e identificar como elas se relacionam foi utilizado o programa aberto, desenvolvido pelos professores Ian H. Witten e Eibe Frank, do Departamento de Ciências Computacionais da Universidade de Waikato, Nova Zelândia, conhecido como WEKA (Waikato Environment Knowledge Analysis), versão 3.8, que facilitou a construção da Árvore de Decisão pretendida neste trabalho pelo fato desse programa trabalhar com grandes volumes de dados e oferecer diferentes técnicas de Aprendizado de Máquina, incluindo as Árvores de Decisão. A Figura 3 apresenta o programa e a Árvore de Decisão induzida com o seu uso.

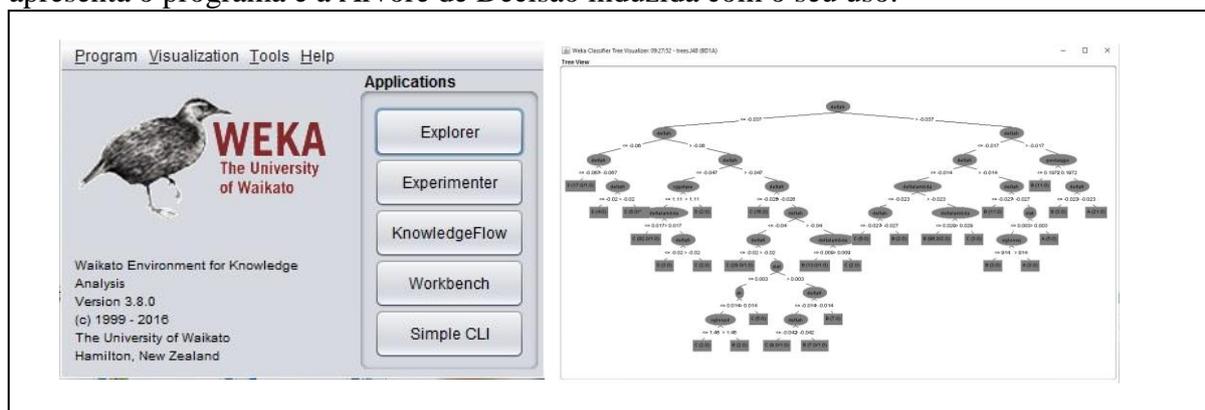


Figura 3: Programa Empregado na Indução da Árvore de Decisão

Esse programa foi adquirido, de maneira gratuita, no endereço: <http://www.cs.waikato.ac.nz>, e o seu uso produziu mais que os resultados esperados, superando as expectativas iniciais do trabalho.

2.3 Classes de Exatidão

Durante o treinamento computacional de uma Árvore de Decisão (AD) o sistema computacional cria regras de classificação a partir da situação conhecida para prever sobre

novos eventos, e por esse motivo a AD precisa ser instruída sobre as abrangências das classes a serem consideradas.

Trabalhando com estações geodésicas que possuem coordenadas conhecidas é sempre possível identificar-se a qualidade do posicionamento do método PPP, fazendo-se a comparação entre as coordenadas determinadas no PPP e as coordenadas conhecidas, o que torna possível estabelecer as classes e suas abrangências que devem ser respeitadas na predição dos resultados. No presente trabalho foram definidas as seguintes classes de exatidão para o treinamento da árvore:

CLASSE	EXATIDÃO
A	0,0 a 2,0 cm
B	2,1 a 4,0 cm
C	4,1 a 6,0 cm
D	6,1 a 8,0 cm
Z	> 8,0 cm

O banco de referência que foi usado para fazer o treinamento da Árvore de Decisão foi povoado de 396 instâncias, que representam as sessões de medição, contendo cada uma, os 17 atributos mencionados. A classe de exatidão conhecida em cada caso foi definida pelo pesquisador e passou a ser o 18º atributo no banco. A Figura 4 mostra as exatidões implícitas nos 396 dados de treinamento da AD.

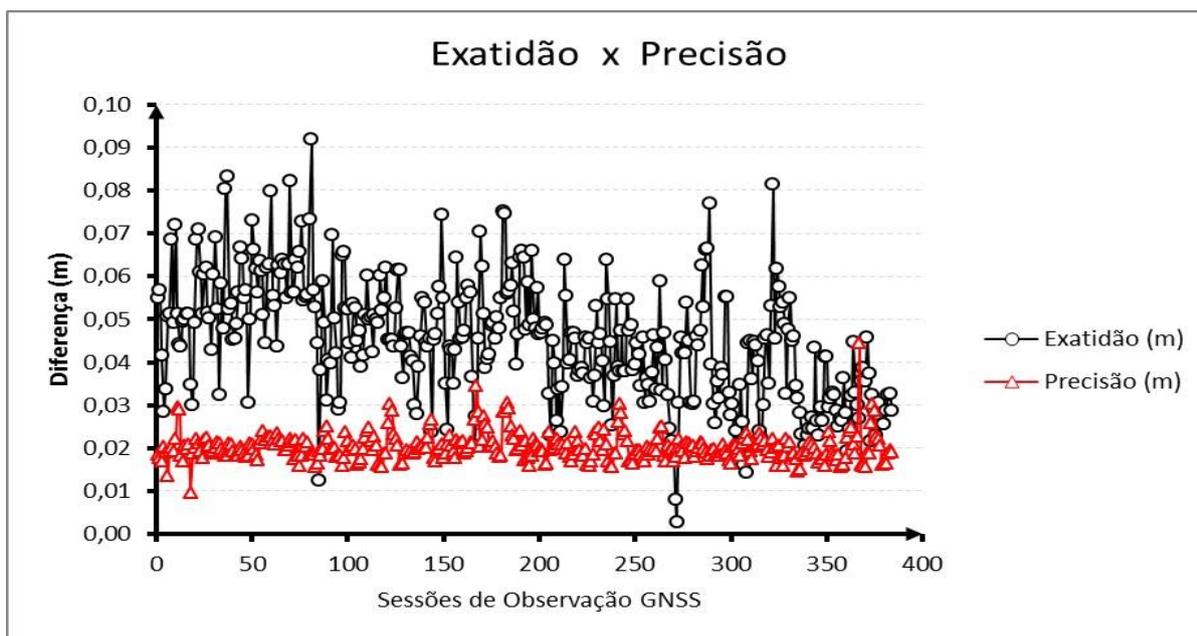


Figura 4: Precisões e Exatidões implícitas nos dados GNSS

Após o treinamento computacional foi induzida a Árvore de Decisão (AD) que alcançou um grau 86% de confiança estatística. Ela usada no trabalho em condições de prever sobre a exatidão de novas soluções PPP tendo por base apenas a precisão da solução.

2.4 Validação do Experimento

Toda vez que a Árvore de Decisão (AD) é acionada para classificar uma nova solução de posicionamento PPP da qual se conhece apenas a precisão, segundo WITTEN e FRANK (2005), ela segue as regras implícitas no banco de referência, identifica os vínculos que

possam existir entre as dezessete variáveis dessa nova medição com as respectivas variáveis existentes no banco, e faz a sua predição sobre a décima oitava variável que é a exatidão da nova solução, algo ainda desconhecido. Como se trata de um método inferencial sempre existe a probabilidade de acerto e a probabilidade de erro que são diretamente associadas com a qualidade do banco de referência. No sentido de verificar a qualidade das predições feitas pela AD foi desenvolvida uma etapa própria destinada à validação das suas predições.

Para a etapa de validação poderiam ter sido empregados quaisquer dados GNSS de dupla frequência, adquiridos em qualquer novo local, escolhido aleatoriamente, desde que fosse um local diferente das estações geodésicas EESC, SPBO e SPC1 que já haviam sido usadas na etapa de treinamento. Porém, se o local escolhido fosse um local qualquer, teríamos unicamente que aceitar a predição feita pela Árvore de Decisão (AD), sem meios de verificar a qualidade da predição. Para saber exatamente como a AD estava classificando os novos dados, optou-se por usar uma quarta estação geodésica da RBMC, localizada no interior do triângulo formado pelas três estações usadas na etapa de treinamento. Essa quarta estação geodésica forneceu os dados que foram usados na etapa de validação das predições. A estação escolhida foi:

Estação SPPI, código 99588, com coordenadas oficiais publicadas pelo IBGE como sendo: latitude (ϕ) = 22° 42' 10,9769" S, longitude (λ) = 47° 37' 25,0333" W e altura geométrica (h) = 561,88 m, fixa em um cilindro de concreto construído na Estação Meteorológica da USP/ESALQ, onde funciona um receptor GNSS Trimble NETR8, de dupla frequência, na cidade de Piracicaba – SP.

Os dados adquiridos nessa estação serviram para organizar trinta e três sessões de observação espalhadas de janeiro a maio de 2016, porém, em dias diferentes daqueles que foram usados na composição do banco de referência. Os arquivos de dados de cada sessão foram organizados da mesma forma que os arquivos usados no treinamento, ou seja, com três horas cada um, respeitando o período de convergência dos dados no processamento PPP. Esses arquivos foram enviadas online para o processamento PPP no site do IBGE.

A Tabela 1 mostra as diferenças encontradas nos valores de latitude (ϕ), longitude (λ) e altura geométrica (m), após a comparação entre as coordenadas conhecidas da estação SPPI e as coordenadas calculadas em cada sessão. Essas diferenças informam a exatidão real de cada sessão. A última coluna à direita contém a previsão feita pela Árvore de Decisão, através dos caracteres alfabéticos: A, B, C, D e Z, que representam a classe da previsão estimada para cada sessão, conforme item 2.3.

Tabela 1: Comparação entre as exatidões reais e as exatidões previstas pela Árvore de Decisão.

Sessão	Diferenças			Exatidão		Sessão	Diferenças			Exatidão	
	Latitude $\Delta\phi$ (m)	Longitude $\Delta\lambda$ (m)	Altura Δh (m)	Real (m)	Prevista (Classe)		Latitude $\Delta\phi$ (m)	Longitude $\Delta\lambda$ (m)	Altura Δh (m)	Real (m)	Prevista (Classe)
1	-0,03	-0,01	-0,03	0,04	B	18	-0,01	0,00	-0,06	0,06	C
2	-0,02	0,01	-0,04	0,05	C	19	-0,02	0,02	-0,06	0,06	C
3	-0,02	0,02	-0,04	0,05	C	20	-0,02	0,01	-0,05	0,05	C
4	-0,01	0,03	-0,06	0,07	Z	21	-0,02	0,01	-0,04	0,04	B
5	-0,01	-0,01	-0,03	0,03	B	22	-0,01	0,01	-0,02	0,02	A
6	-0,01	0,00	-0,06	0,06	C	23	-0,01	0,02	-0,05	0,05	C
7	-0,02	0,01	-0,08	0,08	Z	24	-0,02	0,01	-0,01	0,02	A
8	-0,02	0,01	-0,05	0,05	C	25	-0,02	0,02	-0,05	0,06	C
9	-0,01	0,01	-0,09	0,09	Z	26	-0,02	0,02	-0,03	0,04	B

10	-0,03	0,00	-0,09	0,09	Z	27	-0,03	0,01	-0,03	0,04	B
11	-0,02	0,00	-0,04	0,04	B	28	-0,01	0,01	-0,04	0,04	B
12	-0,02	0,01	-0,06	0,06	C	29	-0,02	0,01	-0,03	0,04	B
13	-0,03	-0,01	0,01	0,04	B	30	-0,02	0,01	-0,05	0,05	C
14	-0,01	-0,01	-0,04	0,04	B	31	-0,02	0,00	-0,04	0,04	B
15	-0,01	-0,01	-0,06	0,06	C	32	-0,02	0,01	-0,04	0,05	C
16	-0,02	0,02	-0,05	0,06	C	33	-0,02	0,01	-0,04	0,05	C
17	-0,02	-0,02	-0,07	0,07	Z	-	-	-	-	-	

Nesse ponto, cabe lembrar que na etapa de treinamento foram usados 18 atributos em cada instância do banco de referência, sendo o último, justamente a classificação da exatidão das coordenadas. Destaca-se esse ponto porque na etapa validação, cada instância representando uma sessão de observação, foi organizada com apenas 17 atributos, deixando-se o 18º atributo, relativo à exatidão, para que a AD fizesse a predição.

Todas as trinta e três predições feitas pela Árvore de Decisão (AD) puderam ser validadas porque a estação SPPI possui coordenadas conhecidas e a etapa de validação atingiu um resultado com vinte e nove predições corretas, o que coloca o grau de acerto nesse trabalho em 88%, um pouco acima da expectativa inicial de 86%.

3. RESULTADOS E DISCUSSÃO

Os resultados obtidos na indução da Árvore de Decisão (AD) se mostraram melhores à medida que foram introduzidos dados cruzados e não somente os dados iniciais, conforme previam WITTEN e FRANK, (2005). As variáveis de número 12 e 13 foram introduzidas para explicitar, respectivamente, a proporção de épocas GPS rejeitadas e a proporção de épocas GLONASS rejeitadas, além de deixar evidente o grau de participação de cada sistema de posicionamento para o resultado final. Além disso, essas variáveis mostram o grau de aproveitamento dos dados de cada sistema orbital, individualmente, o que ajudou a AD a se condicionar melhor para futuras interpretações.

Confirmou-se que, de fato, a precisão das medições feitas com o GNSS é algo muito diferente da exatidão. A Figura 4 deixa bem clara essa diferença. Além disso, ela mostra que a relação entre a precisão e a exatidão não é determinística e por esse motivo cada resultado de posicionamento tem que ser monitorado individualmente sob pena de se aceitar como bom um resultado que não é tão bom assim.

Ao trabalhar somente com a precisão o usuário acaba por desconhecer a exatidão do resultado, ficando exposto ao risco de adotar como confiáveis coordenadas que muito se afastam da posição real do ponto no terreno.

A Árvore de Decisão é uma ferramenta que ajuda o usuário antever a correlação entre a precisão e a exatidão de uma solução dando-lhe meios de melhor decidir sobre o resultado.

O processamento de dados pelo método PPP-GNSS alcançou, nesse estudo, uma precisão média de três centímetros, conforme já estimavam SILVA e SEGANTINE (2015). Com a Árvore consegue-se atestar que o método PPP-GNSS produz resultados com exatidão de até nove centímetros. Esse nível de qualidade coloca o método PPP em igualdade aos demais métodos de posicionamento preciso e numa condição muito melhor do que se supunha.

Os resultados obtidos são satisfatórios e totalmente dentro do esperado uma vez que eles mostraram um comportamento muito parecido entre si, tanto para o conjunto das

precisões quanto para o conjunto das exatidões. Embora quatro valores de exatidão não tenham acompanhado o comportamento do grupo das exatidões nas 396 sessões de medição, eles não devem ser ignorados porque podem de fato ocorrer.

O trabalho evidenciou que a exatidão das coordenadas como um parâmetro de importância superior à precisão, apesar dela ainda não ser tratada no meio técnico com essa relevância.

4. CONCLUSÕES

O objetivo do trabalho foi analisar o quanto os resultados de posicionamento de pontos obtidos com o método PPP-GNSS se aproximam das coordenadas da posição real desses pontos no terreno, classificando a magnitude do afastamento (exatidão) com o auxílio da Árvore de Decisão, um poderoso recurso de interpretação de dados que possuem múltiplas variáveis. Neste contexto, considera-se que o objetivo do trabalho foi alcançado e diante dos resultados obtidos conclui-se que:

O uso da Árvore de Decisão para classificar a exatidão de uma determinada solução de posicionamento é um procedimento viável devido à facilidade de treiná-la a partir de um grande número de medições compostas por múltiplas variáveis.

O estudo demonstrou que uma Árvore de Decisão (AD), quando corretamente induzida tendo por referência um banco de dados consistente, se torna uma ferramenta confiável para a estimativa da exatidão de soluções PPP-GNSS que possuam apenas a precisão. O grau de confiança da AD que foi induzida no presente estudo, alcançou o nível de 86%, fato que foi comprovado na etapa de validação.

As trezentas e noventa e seis sessões de medição que foram usadas para induzir e treinar a Árvore de Decisão mostram que a exatidão é algo diferente da precisão, além de uma clara correlação entre elas, sugerindo que possa existir uma ou mais regras de conexão, o que precisa ser melhor investigado.

A exatidão real das soluções de posicionamento usando o método PPP-GNSS apresentou um afastamento máximo de 0,09 m, o que pode ser visto na Tabela 1, colocando o método PPP-GNSS num patamar de exatidão decimétrica.

A Árvore de Decisão pode ser aplicada na investigação de resultados obtidos com outros métodos de posicionamento que usam dados GNSS, porque independente do método o resultado da solução de posicionamento é sempre um conjunto de coordenadas que vêm sempre acompanhadas dos mesmos indicadores estatísticos de precisão que serviram como variáveis no presente trabalho.

5. AGRADECIMENTOS

Ao Instituto Brasileiro de Geografia e Estatística (IBGE), que através da sua Coordenação de Geodésia organiza e disponibiliza os dados GNSS adquiridos pelos receptores da Rede Brasileira de Monitoramento Contínuo dos Sistemas GNSS (RBMC), assim como disponibiliza gratuitamente o processamento das observáveis pelo método PPP, ajuda imprescindível, sem a qual esse trabalho não seria possível.

6. REFERÊNCIAS

HOFMANN-WELLENHOF, B.; LICHTNEGGER, H.; WASLE E. *GNSS - Global Navigation Satellite Systems, GPS, GLONASS, Galileo & more.* Editora Springer-Verlag, New York, (2008).

MONICO, J.F.G. *Posicionamento pelo GNSS. 2ª. Edição.* Editora Unesp, São Paulo, (2008).

IBGE. *Manual do Usuário Aplicativo Online IBGE-PPP. Coordenação de Geodésia Diretoria de Geociências, Rio de Janeiro, (2013).*

LEVINE, R.I.; DRANG, D.E.; EDELSON B. *Inteligência Artificial e Sistemas Especialistas.* Editora Mc-Graw-Hill, São Paulo, (1988).

SILVA I.; SEGANTINE, P.C.L. *Topografia para Engenharia.* Elsevier Editora Ltda, Rio de Janeiro, RJ, (2015)

RICH, E. & KNIGHT, K. *Artificial Intelligence. Second Edition.* Editora MacGraw Hill – U.S.A., (1991)

WITTEN. I. H.; FRANK, E. *Data Mining Practical Machine Learning Tools and Techniques. Second Edition.* Morgan Kaufmann Publications, U.S.A., (2005).

XU, G. *GPS: Theory, Algorithms and Applications,* Editora Springer-Verlag, Berlin Heidelberg, Alemanha, (2003).

ZHAN-LI S., KIN-MAN L., QING-WAY G. *An Effective Missing-Data Estimation Approach for Small Size Images Sequences. IEEE-Computational Inteligence, V.10, N. 3, pp. 10-18, (2015).*